Molien generating functions, invariants and covariants of magnetic point groups[a)]

Rhoda Berenson
Physics Department, Nassau Community College, Garden City, New York 11530

(Received 30 March 1981; accepted for publication 26 June 1981)

The general properties of the Molien generating function and invariant and covariant tensors for the corepresentations of nonunitary point groups are presented. The generating functions and $(\mathscr{D}_1,\mathscr{D}_m)$ (invariant) and $(\mathscr{D}_r,\mathscr{D}_m)$ (covariant) tensors are obtained for the 32 grey point groups and the 58 black–white point groups.

PACS numbers: 02.20. + b

I. INTRODUCTION

There are various problems in physics such as the renormalization and Landau theories of phase transitions and the calculation of selection rules for high order optical processes which require the knowledge of polynomial invariants for the relevant representations of the symmetry group of the problem. The first step in determining these invariants is the calculation of the Molien function[1,2] for the particular representation of interest. Molien functions and polynomial invariants and covariants have been calculated for the 32 point groups[3–5] and a systematic procedure and some applications have been given for calculating Molien functions for space groups.[6] It is the purpose of this work to use corepresentation theory[7] in order to investigate properties of Molien functions, invariants and covariants of magnetic point groups. Recently, Saint-Aubin[8] has presented Molien functions and integrity bases for unitary representations of the finite subgroups of the Lorentz group O(3,1). These groups are isomorphic to the magnetic point groups. In addition, some discussion of integrity bases for unitary representations of magnetic point groups has also been given by Kopsky.[9] However, neither Saint-Aubin nor Kopsky consider the corepresentations of the nonunitary groups.

Section II will summarize the pertinent results of Refs. 4 and 5 for Molien functions and integrity bases for unitary point groups. Sections III and IV will consider Molien functions and invariants and covariants for nonunitary point groups. Finally, Sec. V will present the Molien functions, invariants and covariants for the 32 grey groups and the 58 black–white groups.

II. MOLIEN FUNCTIONS AND INVARIANTS OF UNITARY GROUPS

Let G be a finite unitary group having $|G|$ elements g and irreducible representations $\Gamma_j(g)$ of dimension l_j and characters $\chi_j(g)$. If one forms the n^{th} symmetrized product representation of the m^{th} irreducible representation, $[\Gamma_m]^{(n)}$, having characters $\chi_m{}^{(n)}$, then $[\Gamma_m]^{(n)}$ will contain the r^{th} representation Γ_r, $C^r_{m,(n)}$ times where

$$C^r_{m,(n)} = \frac{1}{|G|}\sum_{g\in G}\chi_r(g)^*\chi_m{}^{(n)}(g). \tag{2.1}$$

In particular, $C^1_{m,(n)}$ indicates the number of times Γ_1, the identity representation, appears in $[\Gamma_m]^{(n)}$ and is equal to the number of independent invariants of degree n transforming as Γ_m. That is, if Γ_m has bases $(\psi^m_1\cdots\psi^m_{l_m})$, then there are $C^1_{m,(n)}$ independent homogeneous polynomials $\mathscr{P}^{(n)}$ of degree n in the $\{\psi^m_i\}$ such that

$$\begin{aligned}P_g\mathscr{P}^{(n)} &= \Gamma_1\mathscr{P}^{(n)}\\ &= \mathscr{P}^{(n)}\ \text{for all } g\in G,\end{aligned} \tag{2.2}$$

where

$$P_g\psi^m_\mu = \sum_\nu \Gamma_m(g)_{\nu\mu}\psi^m_\nu. \tag{2.3}$$

These invariant polynomials will be called (Γ_1,Γ_m) tensors.

We can also construct (Γ_r,Γ_m) covariant tensors of degree n such that the tensor components $(f^r_1,\ldots,f^r_{l_r})$ are each homogeneous polynomials of degree n in the bases $\{\psi^m_i\}$ and the $\{f^r_i\}$ transform by Γ_r when the $\{\psi^m_i\}$ transform by Γ_m. The number of independent (Γ_r,Γ_m) tensors of degree n is given by $C^r_{m,(n)}$.

Following the procedure of Molien[1] as elaborated by Burnside,[2] $C^r_{m,(n)}$ can be found most easily from

$$\begin{aligned}\sum_n C^r_{m,(n)}\lambda^n &= \frac{1}{|G|}\sum_{g\in G}\frac{\chi_r(g)^*}{\det|E-\lambda\Gamma_m(g)|}\\ &\equiv B(\Gamma_r,\Gamma_m;\lambda),\end{aligned} \tag{2.4}$$

where E is the identity matrix.

As is pointed out in Refs. 4 and 5, the Molien function $B(\Gamma_r,\Gamma_m;\lambda)$ can be written as

$$B(\Gamma_r,\Gamma_m;\lambda) = \sum_p k_p\lambda^p\Big/\prod_q(1-\lambda^q), \tag{2.5}$$

where the k_p are positive integers. For each factor $(1-\lambda^q)$ in the denominator of $B(\Gamma_1,\Gamma_m;\lambda)$, there corresponds an algebraically independent polynomial of degree q, written $I^q(\Gamma_1,\Gamma_m)$. For each term $k_p\lambda^p$ in the numerator of $B(\Gamma_r,\Gamma_m;\lambda)$ there are k_p linearly independent tensors $E^p(\Gamma_r,\Gamma_m)$ of degree p. All $I^n(\Gamma_1,\Gamma_m)$ are linearly independent of each other and algebraically independent of any lower degree (Γ_1,Γ_m) invariant, but powers of the $E^n(\Gamma_1,\Gamma_m)$ may be expressable as polynomials in lower degree invariants.

III. MOLIEN FUNCTIONS OF NONUNITARY POINT GROUPS

A nonunitary group M can be written as

$$M = G + A_0G, \tag{3.1}$$

[a)] Supported in part by a grant from the National Science Foundation DMR-78-12399 and Faculty Research Award, PSC-BHE RF No. 13404.

TABLE I. Corepresentations of nonunitary groups :$M = G + A_0G$.

Type of representation Γ of G	Corepresentation $\mathscr{D}(u)$ of M	Corepresentation $\mathscr{D}(A_0u)$ of M
Type a: $\Gamma(u) = \beta\Gamma(A_0^{-1}uA_0)^*\beta^{-1}$ $\beta\beta^* = \Gamma(A_0^2)$	$\mathscr{D}(u) = \Gamma(u)$	$\mathscr{D}(A_0u) = \Gamma(A_0uA_0^{-1})\beta$
Type b: $\Gamma(u) = \beta\Gamma(A_0^{-1}uA_0)^*\beta^{-1}$ $\beta\beta^* = -\Gamma(A_0^2)$	$\mathscr{D}(u) = \begin{pmatrix}\Gamma(u) & 0 \\ 0 & \Gamma(u)\end{pmatrix}$	$\mathscr{D}(A_0u) = \begin{pmatrix}0 & -\Gamma(A_0uA_0^{-1})\beta \\ \Gamma(A_0uA_0^{-1})\beta & 0\end{pmatrix}$
Type c: $\Gamma(u)$ is not equivalent to $\overline{\Gamma}(u) = \Gamma(A_0^{-1}uA_0)^*$	$\mathscr{D}(u) = \begin{pmatrix}\Gamma(u) & 0 \\ 0 & \overline{\Gamma}(u)\end{pmatrix}$	$\mathscr{D}(A_0u) = \begin{pmatrix}0 & \Gamma(A_0uA_0) \\ \overline{\Gamma}(A_0uA_0^{-1}) & 0\end{pmatrix}$

where G is a unitary group, $A_0 = \theta u_0$ is an antiunitary element, θ is the time reversal element, and u_0 is a unitary element which may or may not belong to G. If u_0 belongs to G we have one of the 32 grey groups. If u_0 does not belong to G we have one of the 58 black–white groups. The corepresentations $\mathscr{D}$ of M may be of three types[7] depending on the relationship between $\Gamma(u)$ and $\overline{\Gamma}(u) = \Gamma^*(A_0^{-1}uA_0)$, where Γ is an irreducible representation of G. These types of corepresentations are presented in Table I.

$d^r_{m,(n)}$, the number of times the corepresentation $\mathscr{D}_r$ appears in the symmetrized nth product corepresentation $[\mathscr{D}_m]^{(n)}$, is given by[10]

$$d^r_{m,(n)} = \frac{1}{|G|}\sum_{u\in G} X_r(u)^* X^{(n)}_m(u) \Big/ \frac{1}{|G|}\sum_{u\in G} X_r(u) X_r(u)^*. \quad (3.2)$$

These sums are only over unitary elements. X_r is the character of corepresentation $\mathscr{D}_r$ and $X^{(n)}_m$ is the character of $[\mathscr{D}_m]^{(n)}$.

The $d^r_{m,(n)}$ can be determined more readily from a Molien function $B(\mathscr{D}_r,\mathscr{D}_m;\lambda)$, where

$$B(\mathscr{D}_r,\mathscr{D}_m;\lambda) = \sum_n d^r_{m,(n)}\lambda^n = \frac{1}{N}\frac{1}{|G|}\sum_{u\in G}\frac{X_r(u)^*}{\det|E - \lambda\mathscr{D}_m(u)|} \quad (3.3)$$

and

$$N = \begin{cases} 1, & \mathscr{D}_r \text{type a} \\ 4, & \mathscr{D}_r \text{type b.} \\ 2, & \mathscr{D}_r \text{type c} \end{cases} \quad (3.4)$$

The calculation of $B(\mathscr{D}_r,\mathscr{D}_m;\lambda)$ can be simplified by separately considering whether Γ_r and Γ_m are of type a, b, or c as described in Table I.

A. Γ_m is type a

If Γ_m is type a then from Eq. (3.3) and the structure of the corepresentations as given in Table I it is obvious that

$$B(\mathscr{D}_r,\mathscr{D}_m;\lambda) = \begin{cases} B(\Gamma_r,\Gamma_m;\lambda) & \Gamma_r \text{type a} \\ \tfrac{1}{4}B(\Gamma_r,\Gamma_m;\lambda) & \Gamma_r \text{type b} \\ \tfrac{1}{2}\{B(\Gamma_r,\Gamma_m;\lambda) + B(\overline{\Gamma}_r,\Gamma_m;\lambda)\} & \Gamma_r \text{type c.} \end{cases} \quad (3.5)$$

Since the $B(\Gamma_r,\Gamma_m;\lambda)$ are given in Refs. 4 and 5, the $B(\mathscr{D}_r,\mathscr{D}_m;\lambda)$ are readily obtained when Γ_m is a type a representation. Note that the case of Γ_r type b does not occur for point groups.

B. Γ_m is type b

All type b representations of point groups are real and one dimensional so that

$$B(\mathscr{D}_r,\mathscr{D}_m;\lambda) = \frac{1}{N}\frac{1}{|G|}\sum_{u\in G}\frac{X_r(u)^*}{[1-\lambda\chi_m(u)]^2} = \frac{1}{N}\frac{1}{|G|}\sum_{u\in G} X_r(u)^*\{1 + 2\lambda\chi_m(u) + 3\lambda^2\chi_m(u)^2 + \cdots\}. \quad (3.6)$$

But

$$\frac{1}{N}\frac{1}{|G|}\sum_{u\in G} X_r(u)^*\chi_m(u)^j = \begin{cases}\delta_{r,1} & (j \text{ even}), \\ \tfrac{1}{2}\delta_{r,m} & (j \text{ odd}).\end{cases} \quad (3.7)$$

Thus

$$B(\mathscr{D}_r,\mathscr{D}_m;\lambda) = \begin{cases}\dfrac{1+\lambda^2}{(1-\lambda^2)^2} & \text{for } \mathscr{D}_r = \mathscr{D}_1, \\ \dfrac{\lambda}{(1-\lambda^2)^2} & \text{for } \mathscr{D}_r = \mathscr{D}_m.\end{cases} \quad (3.8)$$

C. Γ_m is type c

If Γ_m is a type c representation, then

$$B(\mathscr{D}_r,\mathscr{D}_m;\lambda) = \frac{1}{N}\frac{1}{|G|}\sum_{u\in G}\frac{X_r(u)^*}{\det|E - \lambda(\Gamma_m(u)\oplus\overline{\Gamma}_m(u))|}. \quad (3.9)$$

B may be calculated directly from Eq. (3.9) or one can use the following[4]:

$$B(\Gamma_r,\Gamma_m\oplus\overline{\Gamma}_m;\lambda_1,\lambda_2) = \sum_{\Gamma_1\Gamma_2} B(\Gamma_{r_1},\Gamma_m;\lambda_1)B(\Gamma_{r_2},\overline{\Gamma}_m;\lambda_2)C^r_{r_1r_2}, \quad (3.10)$$

where $C^r_{r_1r_2}$ is the multiplicity of Γ_r in $\Gamma_{r_1}\otimes\Gamma_{r_2}$. Then

$$B(\mathscr{D}_r,\mathscr{D}_m;\lambda) = \begin{cases} B(\Gamma_r,\Gamma_m\oplus\overline{\Gamma}_m;\lambda) & \text{for } \Gamma_r \text{type a,} \\ \tfrac{1}{2}B(\Gamma_r,\Gamma_m\oplus\overline{\Gamma}_m;\lambda) & \text{for } \Gamma_r \text{type b,} \\ \tfrac{1}{2}\{B(\Gamma_r,\Gamma_m\oplus\overline{\Gamma}_m;\lambda) + B(\overline{\Gamma}_r,\Gamma_m\oplus\overline{\Gamma}_m;\lambda)\} & \text{for } \Gamma_r \text{type c.} \end{cases} \quad (3.11)$$

However, for crystallographic point groups, with the exception of representations Γ_6 and Γ_7 of T [for which Eq. (3.10) is the simplest method of calculation], all type c representations are one dimensional. In addition, except for type c representations of five black–white groups (see Table V), the remaining type c representations are such that $\overline{\Gamma}(u) = \Gamma^*(u)$ for all u. For these one dimensional complex representations the calculations of the Molien function can be simplified as follows:

$$B(\Gamma_r,\Gamma_m \oplus \Gamma_m^*;\lambda) = \frac{1}{|G|}\sum_{u\in G}\frac{\chi_r(u)^*}{(1-\lambda\chi_m(u))(1-\lambda\chi_m(u)^*)}$$
$$= \frac{1}{|G|}\sum_{u\in G}\chi_r(u)^*\sum_{j=0}^{\infty}(\lambda^j(\chi_m(u)^j))\left(\sum_{j'=0}^{\infty}\lambda^{j'}(\chi_m(u)^*)^{j'}\right). \tag{3.12}$$

Let z be the smallest nonzero integer such that

$$\chi_m(u)^z = (\chi_m(u)^*)^z = 1 \quad \text{for all } u \tag{3.13}$$

and define s and t such that

$$\chi_m(u)^s = (\chi_m(u)^*)^t = \chi_r(u) \quad \text{for all } u \tag{3.14}$$

where

$$0 < s \leqslant z, \quad 0 \leqslant t < z, \quad \text{and } s + t = z.$$

Then, since

$$\frac{1}{|G|}\sum_u \chi_r(u)^*\chi_m(u)^j(\chi_m(u))^{*j'}$$
$$= 1 \quad \text{for } [\chi_m(u)^j(\chi_m(u))^{*j'}] = \chi_r(u),$$
$$= 0 \quad \text{otherwise}, \tag{3.15}$$

the only nonzero contributions to (3.12) are for $j' = t + j + nz$, for each j and for $j = s + j' + n'z$ for each j' $(n,n' = 0,1,2,...)$. Thus

$$B(\Gamma_r,\Gamma_m \oplus \Gamma_m{}^*;\lambda)$$
$$= \sum_{j=0}^{\infty}\sum_{n=0}^{\infty}\lambda^j\lambda^{t+j+nz} + \sum_{j'=0}^{\infty}\sum_{n'=0}^{\infty}\lambda^{j'}\lambda^{s+j'+n'z}$$
$$= \frac{\lambda^t + \lambda^s}{(1-\lambda^2)(1-\lambda^z)}. \tag{3.16}$$

IV. INVARIANTS AND COVARIANTS OF NONUNITARY POINT GROUPS

If $\{\phi_1^m \cdots \phi_l^m\}$ are the bases for corepresentation $\mathscr{D}_m$, then an invariant polynomial $\mathscr{P}^{(n)}$ of degree n in these bases [i.e., a $(\mathscr{D}_1,\mathscr{D}_m)$ invariant] must satisfy

$$P_u\mathscr{P}^{(n)} = \mathscr{P}^{(n)} \quad \text{all } u\in G \tag{4.1}$$

and

$$P_A\mathscr{P}^{(n)} = \mathscr{P}^{(n)} \quad \text{all } A\in A_0G, \tag{4.2}$$

where

$$P_u\phi_u^m = \sum_v \mathscr{D}_m(u)_{v\mu}\phi_v^m \tag{4.3}$$

and

$$P_A\phi_\mu^m = \sum_v \mathscr{D}_m(A)_{v\mu}\phi_v^m \tag{4.4}$$

and

$$P_A\sum_j C_j\phi_j^m = \sum_j C_j{}^*P_A\phi_j^m. \tag{4.5}$$

Similarly, $(\mathscr{D}_r,\mathscr{D}_m)$ covariant tensors of degree n are such that the tensor components $\{f_1^r,...,f_{l_r}^r\}$ are each homogeneous polynomials of degree n in the bases $\{\phi_i^m\}$ and the $\{f_i^r\}$ transform by $\mathscr{D}_r(g)$ when the $\{\phi_i^m\}$ transform by $\mathscr{D}_m(g)$ for all $g\in M = G + A_0G$.

Two corepresentations $\mathscr{D}$ and $\mathscr{D}'$ are equivalent if there exists a unitary matrix V such that

$$\left.\begin{aligned}\mathscr{D}'(u) &= V^{-1}\mathscr{D}(u)V\\ \mathscr{D}'(A_0u) &= V^{-1}\mathscr{D}(A_0u)V^*\end{aligned}\right\}\text{for all } u\in G. \tag{4.6}$$

The matrix V transforms the bases $\{\phi_i\}$ of $\mathscr{D}$ to the bases $\{\phi_i'\}$ of $\mathscr{D}'$. That is,

$$\phi_\alpha' = \sum_\beta V_{\beta\alpha}\phi_\beta. \tag{4.7}$$

If $V = \omega E$ where ω is an arbitrary phase and E is the identity matrix, then $\mathscr{D}'(u) = \mathscr{D}(u)$ and $\mathscr{D}'(A_0u) = \omega^{*2}\mathscr{D}(A_0u)$. Thus a common arbitrary phase factor exists for the corepresentation matrices for the antiunitary elements. Those corepresentations having identical matrices $\mathscr{D}(u)$ but matrices $\mathscr{D}(A_0u)$ that differ by a phase factor are equivalent. Similarly, a change of phase for the matrix β for type a or b representations (see Table I) produces an equivalent corepresentation.

In particular, the "identity" corepresentation $\mathscr{D}_1$ has all + 1's for unitary elements and some common phase ω_1 for all antiunitary elements. In identifying invariants, however, Eqs. (4.1) and (4.2) will be used. That is, a $(\mathscr{D}_1,\mathscr{D}_m)$ invariant transforms as that corepresentation $\mathscr{D}_1$ for which $\omega_1 = 1$.

As always in presenting tables of bases, invariants, or covariants, the entries correspond to a particular choice of representation or corepresentation and a transformation to an equivalent representation or corepresentation will result in a transformation of bases [Eq. (4.7)] and a change in the form of invariants or covariants. In particular, for two corepresentations $\mathscr{D}_m$ and $\mathscr{D}_m'$ related by Eq. (4.6) where $V = \omega_m E$, consider an invariant polynomial $\mathscr{P}^{(n)}$ and the transformed polynomial $\mathscr{P}^{(n)'}$. $\mathscr{P}^{(n)}$ and $\mathscr{P}^{(n)'}$ can be written in terms of bases $\{\phi_i^m\}$ and $\{\phi_i^{m'}\}$, respectively. If

$$\mathscr{P}^{(n)} = \sum_{\{k_i\}} C_{\{k_i\}}\left((\phi_1^m)^{k_1}\cdots(\phi_{l_m}^m)^{k_{l_m}}\right), \tag{4.8}$$

where $C_{\{k_i\}}$ is a complex constant (possibly zero) and there is one $C_{\{k_i\}}$ for each set of integers $\{k_i\}, \Sigma_{i=1}^{l_m} k_i = n$, then

$$\mathscr{P}^{(n)'} = (\omega_m^*)^n\sum_k C_{\{k_i\}}\left((\phi_1^{m'})^{k_1}\cdots(\phi_{l_m}^{m'})^{k_{l_m}}\right), \tag{4.9}$$

where the $C_{\{k_i\}}$ are the same as in Eq. (4.8). Thus, if invariants are given for $\mathscr{D}_m$, the invariants for $\mathscr{D}_m'$ are obtained from (4.9). Similar phase relationships will exist for covariants $(\mathscr{D}_r,\mathscr{D}_m)$. For convenience, therefore, all invariants and covariants have been calculated for the particular choice of corepresentations $\mathscr{D}_r$ and $\mathscr{D}_m$ in which $\omega_r = 1$ and $\omega_m = 1$.

Now consider the form of the $(\mathscr{D}_r,\mathscr{D}_m)$ invariants and covariants for different types of corepresentations.

A. Γ_m and Γ_r are both type a

Two situations arise.

1. If β_m and β_r can be chosen such that $\beta_m = \Gamma_m(u)$ for some element $u \in G$ and $\beta_r = \Gamma_r(\bar{u})$ for some element $\bar{u} \in G$, then the $(\mathscr{D}_r, \mathscr{D}_m)$ tensors are identical to the (Γ_r, Γ_m) tensors given in Refs. 4 and 5.

2. If $\beta_m \neq \Gamma_m(u)$ for all $u \in G$, and/or $\beta_r \neq \Gamma_r(u)$ for all $u \in G$, then the $(\mathscr{D}_r, \mathscr{D}_m)$ must be checked for invariance under antiunitary operations. For example, consider the $(\mathscr{D}_1, \mathscr{D}_4)$ invariants of corepresentation $\mathscr{D}_4$ of the grey group $3m1'$ $(M = C_{3v} + \theta C_{3v})$ which is based on the type a representation Γ_4 of C_{3v}. From Ref. 5 the generators, Molien function, and (Γ_1, Γ_4) invariants are

$$\text{Generators: } \begin{pmatrix} e^{\pi i/3} & 0 \\ 0 & e^{-\pi i/3} \end{pmatrix};\quad \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

$$\text{Molien function: } B(\Gamma_1, \Gamma_4; \lambda) = \frac{1 + \lambda^8}{(1 - \lambda^4)(1 - \lambda^6)}.$$

$$(\Gamma_1, \Gamma_4) \text{ invariants: } I^4 = \psi_1^2 \psi_2^2, \quad I^6 = \psi_1^6 - \psi_2^6$$
$$E^8 = \psi_1 \psi_2 (\psi_1^6 + \psi_2^6).$$

However, for this example

$\Gamma_4(u) = \beta \Gamma_4(u)^* \beta^{-1}$, where

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

and $\beta \neq \Gamma_4(u)$.

The $(\mathscr{D}_1, \mathscr{D}_4)$ invariants are then

$$I^4 = \phi_1^2 \phi_2^2; \quad I^6 = i(\phi_1^6 - \phi_2^6); \quad E^8 = i\phi_1 \phi_2 (\phi_1^6 + \phi_2^6).$$

B. Γ_m is type a, Γ_r is type c

In this case the $E^p(\mathscr{D}_r, \mathscr{D}_m)$ tensors are given by

$$E^p(\mathscr{D}_r, \mathscr{D}_m) = \begin{pmatrix} E^p(\Gamma_r, & \Gamma_m) \\ E^p(\bar{\Gamma}_r, & \Gamma_m) \end{pmatrix} \tag{4.10}$$

and $E^p(\Gamma_r, \Gamma_m)$ are available from Refs. 4 and 5. Again if $\beta_m \neq \Gamma_m(u)$ for all $u \in G$, then the $E^p(\Gamma_r, \Gamma_m)$ must be checked for invariance under antiunitary operations.

C. Γ_m is type b

Since $\Gamma_m(u)$ is one dimensional, the corepresentation $\mathscr{D}_m$ has bases $\{\phi_1, \phi_2\}$. The unitary operations do not "mix" ϕ_1 and ϕ_2 so that the invariants under G are monomials $\phi_1^{k_1} \phi_2^{k_2}$ and $\phi_1^{k_2} \phi_2^{k_1}$. However, under the operations of an antiunitary operator, say P_{A_0}, we have (for $\beta = 1$)

$$P_{A_0} C \phi_1^{k_1} \phi_2^{k_2} = (-1)^{k_2} C^* \phi_1^{k_2} \phi_2^{k_1}. \tag{4.11}$$

Thus the $(\mathscr{D}_1, \mathscr{D}_m)$ invariants are polynomials $\mathscr{P}^{(n)}$ of the form

$$\mathscr{P}^{(k_1 + k_2)} = C \phi_1^{k_1} \phi_2^{k_2} + (-1)^{k_2} C^* \phi_1^{k_2} \phi_1^{k_1}. \tag{4.12}$$

As an example consider the grey group $11' = C_1 + \theta C_1$ and, in particular, the corepresentation formed from the type b double-valued representation Γ_2 of C_1. The matrices for this corepresentation are given as follows:

$$\mathscr{D}_2(E) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathscr{D}_2(\bar{E}) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$
$$\mathscr{D}_2(\theta E) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathscr{D}_2(\theta \bar{E}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{4.13}$$

TABLE II. Grey groups: $M = G + \theta G$.

M	G	Type a representations	Type b representations	Type c representations
$11'$	C_1	Γ_1	Γ_2	
$\bar{1}1'$	C_i	$\Gamma_1^{\pm}$	$\Gamma_2^{\pm}$	
$21'$	C_2	Γ_1, Γ_2		Γ_3, Γ_4
$m1'$	C_s	Γ_1, Γ_2		Γ_3, Γ_4
$2221'$	D_2	All		
$2mm1'$	C_{2v}	All		
$31'$	C_3	Γ_1	Γ_6	$\Gamma_2, \Gamma_3; \Gamma_4, \Gamma_5$
$321'$	D_3	$\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4$[a]		Γ_5, Γ_6
$3m1'$	C_{3v}	$\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4$[a]		Γ_5, Γ_6
$41'$	C_4	Γ_1, Γ_2		$\Gamma_3, \Gamma_4; \Gamma_5, \Gamma_6; \Gamma_7, \Gamma_8$
$\bar{4}1'$	S_4	Γ_1, Γ_2		$\Gamma_3, \Gamma_4; \Gamma_5, \Gamma_6; \Gamma_7, \Gamma_8$
$4221'$	D_4	All		
$4m1'$	C_{4v}	All		
$\bar{4}2m1'$	D_{2d}	All		
$61'$	C_6	Γ_1, Γ_4		$\Gamma_2, \Gamma_3; \Gamma_5, \Gamma_6; \Gamma_7, \Gamma_8$ $\Gamma_9, \Gamma_{10}; \Gamma_{11}, \Gamma_{12}$
$\bar{6}1'$	C_{3h}	Γ_1, Γ_4		$\Gamma_2, \Gamma_3; \Gamma_5, \Gamma_6; \Gamma_7, \Gamma_8$ $\Gamma_9, \Gamma_{10}; \Gamma_{11}, \Gamma_{12}$
$6221'$	D_6	All		
$6mm1'$	C_{6v}	All		
$\bar{6}2m1'$	D_{3h}	All		
$231'$	T	$\Gamma_1, \Gamma_4, \Gamma_5$		$\Gamma_2, \Gamma_3; \Gamma_6, \Gamma_7$
$4321'$	0	All, Γ_8[a]		
$\bar{4}3m1'$	T_d	All, Γ_8[a]		

[a] For these representations $\beta \neq \Gamma(u)$.

TABLE III. Molien functions and invariants and covariants for type c grey group representations.

M	G	$\Gamma_{m,n}$	Γ_r	Molien function	Invariants and covariants
$21'$ $m1'$	C_2 C_s	$\Gamma_{3,4}$	Γ_1	$\frac{1+\lambda^4}{(1-\lambda^2)(1-\lambda^4)}$	$I^2=i\phi\bar\phi;\ I^4=\phi^4+\bar\phi^4;\ E^4=i(\phi^4-\bar\phi^4)$
			Γ_2	$\frac{2\lambda^2}{(1-\lambda^2)(1-\lambda^4)}$	$E_a^2=\phi^2+\bar\phi^2;E_b^2=i(\phi^2-\bar\phi^2)$
			$\Gamma_{3,4}$	$\frac{\lambda+\lambda^3}{(1-\lambda^2)(1-\lambda^4)}$	$E^1=\begin{pmatrix}\phi\\ \bar\phi\end{pmatrix};E^3=i\begin{pmatrix}\bar\phi^3\\ \phi^3\end{pmatrix}$
$31'$	C_3	$\Gamma_{2,3}$	Γ_1	$\frac{1+\lambda^3}{(1-\lambda^2)(1-\lambda^3)}$	$I^2=\phi\bar\phi;I^3=\phi^3+\bar\phi^3;E^3=i(\phi^3-\bar\phi^3)$
			$\Gamma_{2,3}$	$\frac{\lambda+\lambda^2}{(1-\lambda^2)(1-\lambda^3)}$	$E^1=\begin{pmatrix}\phi\\ \bar\phi\end{pmatrix};E^2=\begin{pmatrix}\bar\phi^2\\ \phi^2\end{pmatrix}$
		$\Gamma_{4,5}$	Γ_1	$\frac{1+\lambda^6}{(1-\lambda^2)(1-\lambda^6)}$	$I^2=i\phi\bar\phi;I^6=\phi^6+\bar\phi^6;E^6=i(\phi^6-\bar\phi^6)$
			$\Gamma_{2,3}$	$\frac{\lambda^2+\lambda^4}{(1-\lambda^2)(1-\lambda^6)}$	$E^2=\begin{pmatrix}\phi^2\\ \bar\phi^2\end{pmatrix};E^4=\begin{pmatrix}\bar\phi^4\\ \phi^4\end{pmatrix}$
			$\Gamma_{4,5}$	$\frac{\lambda+\lambda^5}{(1-\lambda^2)(1-\lambda^6)}$	$E^1=\begin{pmatrix}\phi\\ \bar\phi\end{pmatrix};E^5=i\begin{pmatrix}\bar\phi^5\\ \phi\end{pmatrix}$
			Γ_6	$\frac{\lambda^3}{(1-\lambda^2)(1-\lambda^6)}$	$E^3=\begin{pmatrix}\phi^3\\ \bar\phi^3\end{pmatrix}$
$321'$ $3m1'$	D_3 C_{3v}	$\Gamma_{5,6}$	Γ_1	$\frac{1+\lambda^4}{(1-\lambda^2)(1-\lambda^4)}$	$I^2=i\phi\bar\phi;I^4=\phi^4+\bar\phi^4;E^4=i(\phi^4-\bar\phi^4)$
			Γ_2	$\frac{2\lambda^2}{(1-\lambda^2)(1-\lambda^4)}$	$E_a^2=\phi^2+\bar\phi^2;E_b^2=i(\phi^2-\bar\phi^2)$
			$\Gamma_{5,6}$	$\frac{\lambda+\lambda^3}{(1-\lambda^2)(1-\lambda^4)}$	$E^1=\begin{pmatrix}\phi\\ \bar\phi\end{pmatrix};E^3=i\begin{pmatrix}\bar\phi^3\\ \phi^3\end{pmatrix}$
$41'$ $\bar41'$	C_4 S_4	$\Gamma_{3,4}$	Γ_1	$\frac{1+\lambda^4}{(1-\lambda^2)(1-\lambda^4)}$	$I^2=\phi\bar\phi;I^4=\phi^4+\bar\phi^4;E^4=i(\phi^4-\bar\phi^4)$
			Γ_2	$\frac{2\lambda^2}{(1-\lambda^2)(1-\lambda^4)}$	$E_a^2=\phi^2+\bar\phi^2;E_b^2=i(\phi^2-\bar\phi^2)$
			$\Gamma_{3,4}$	$\frac{\lambda+\lambda^3}{(1-\lambda^2)(1-\lambda^4)}$	$E^1=\begin{pmatrix}\phi\\ \bar\phi\end{pmatrix};E^3=\begin{pmatrix}\bar\phi^3\\ \phi\end{pmatrix}$
		$\Gamma_{5,6}$	Γ_1	$\frac{1+\lambda^8}{(1-\lambda^2)(1-\lambda^8)}$	$I^2=i\phi\bar\phi;I^8=\phi^8+\bar\phi^8;E^8=i(\phi^8-\bar\phi^8)$
			Γ_2	$\frac{2\lambda^4}{(1-\lambda^2)(1-\lambda^8)}$	$E_a^4=\phi^4+\bar\phi^4;E_b^4=i(\phi^4-\bar\phi^4)$
			$\Gamma_{3,4}$	$\frac{\lambda^2+\lambda^6}{(1-\lambda^2)(1-\lambda^8)}$	$E^2=\begin{pmatrix}\phi^2\\ \bar\phi^2\end{pmatrix};E^6=\begin{pmatrix}\bar\phi^6\\ \phi^6\end{pmatrix}$
			$\Gamma_{5,6}$	$\frac{\lambda+\lambda^7}{(1-\lambda^2)(1-\lambda^8)}$	$E^1=\begin{pmatrix}\phi\\ \bar\phi\end{pmatrix};E^7=i\begin{pmatrix}\bar\phi^7\\ \phi^7\end{pmatrix}$
			$\Gamma_{7,8}$	$\frac{\lambda^3+\lambda^5}{(1-\lambda^2)(1-\lambda^8)}$	$E^3=i\begin{pmatrix}\bar\phi^3\\ \phi^3\end{pmatrix};E^5=\begin{pmatrix}\phi^5\\ \bar\phi^5\end{pmatrix}$
		$\Gamma_{7,8}$	Γ_1	Same as $\Gamma_{m,n}=\Gamma_{5,6},\Gamma_r=\Gamma_1$.	
			Γ_2	Same as $\Gamma_{m,n}=\Gamma_{5,6},\Gamma_r=\Gamma_2$.	
			$\Gamma_{3,4}$	Same as $\Gamma_{m,n}=\Gamma_{5,6},\Gamma_{r\bar r}=\Gamma_{3,4}$.	
			$\Gamma_{5,6}$	Same as $\Gamma_{m,n}=\Gamma_{5,6},\Gamma_{r\bar r}=\Gamma_{7,8}$	
			$\Gamma_{7,8}$	Same as $\Gamma_{m,n}=\Gamma_{5,6},\Gamma_{r\bar r}=\Gamma_{5,6}$.	
$61'$ $\bar61'$	C_6 C_{3h}	$\Gamma_{2,3}$	Γ_1	$\frac{1+\lambda^3}{(1-\lambda^2)(1-\lambda^3)}$	$I^2=\phi\bar\phi;I^3=\phi^3+\bar\phi^3;E^3=i(\phi^3-\bar\phi^3)$
			$\Gamma_{2,3}$	$\frac{\lambda+\lambda^2}{(1-\lambda^2)(1-\lambda^3)}$	$E^1=\begin{pmatrix}\phi\\ \bar\phi\end{pmatrix};E^2=\begin{pmatrix}\bar\phi^2\\ \phi^2\end{pmatrix}$
		$\Gamma_{5,6}$	Γ_1	$\frac{1+\lambda^6}{(1-\lambda^2)(1-\lambda^6)}$	$I^2=\phi\bar\phi;I^6=\phi^6+\bar\phi^6;E^6=i(\phi^6-\bar\phi^6)$
			$\Gamma_{2,3}$	$\frac{\lambda^2+\lambda^4}{(1-\lambda^2)(1-\lambda^6)}$	$E^2=\begin{pmatrix}\bar\phi^2\\ \phi^2\end{pmatrix};E^4=\begin{pmatrix}\phi^4\\ \bar\phi^4\end{pmatrix}$
			Γ_4	$\frac{2\lambda^3}{(1-\lambda^2)(1-\lambda^6)}$	$E_a^3=\phi^3+\bar\phi^3;E_b^3=i(\phi^3-\bar\phi^3)$
			$\Gamma_{5,6}$	$\frac{\lambda+\lambda^5}{(1-\lambda^2)(1-\lambda^6)}$	$E^1=\begin{pmatrix}\phi\\ \bar\phi\end{pmatrix};E^5=\begin{pmatrix}\bar\phi^5\\ \phi^5\end{pmatrix}$
		$\Gamma_{7,8}$	Γ_1	$\frac{1+\lambda^{12}}{(1-\lambda^2)(1-\lambda^{12})}$	$I^2=i\phi\bar\phi;I^{12}=\phi^{12}+\bar\phi^{12};E^{12}=i(\phi^{12}-\bar\phi^{12})$
			$\Gamma_{2,3}$	$\frac{\lambda^4+\lambda^8}{(1-\lambda^2)(1-\lambda^{12})}$	$E^4=\begin{pmatrix}\bar\phi^4\\ \phi^4\end{pmatrix};E^8=\begin{pmatrix}\phi^8\\ \bar\phi^8\end{pmatrix}$
			Γ_4	$\frac{2\lambda^6}{(1-\lambda^2)(1-\lambda^{12})}$	$E_a^6=\phi^6+\bar\phi^6;E_b^6=i(\phi^6-\bar\phi^6)$

TABLE III. (Continued)

M	G	$\Gamma_{m,n}$	Γ_r	Molien function	Invariants and covariants
			$\Gamma_{5,6}$	$\frac{\lambda^2+\lambda^{10}}{(1-\lambda^2)(1-\lambda^{12})}$	$E^2=\begin{pmatrix}\phi^2\\ \bar\phi^2\end{pmatrix};E^{10}=\begin{pmatrix}\bar\phi^{10}\\ \phi^{10}\end{pmatrix}$
			$\Gamma_{7,8}$	$\frac{\lambda+\lambda^{11}}{(1-\lambda^2)(1-\lambda^{12})}$	$E^1=\begin{pmatrix}\phi\\ \bar\phi\end{pmatrix};E^{11}=i\begin{pmatrix}\bar\phi^{11}\\ \phi^{11}\end{pmatrix}$
			$\Gamma_{9,10}$	$\frac{\lambda^5+\lambda^7}{(1-\lambda^2)(1-\lambda^{12})}$	$E^5=i\begin{pmatrix}\bar\phi^5\\ \phi^5\end{pmatrix};E^7=\begin{pmatrix}\phi^7\\ \bar\phi^7\end{pmatrix}$
			$\Gamma_{11,12}$	$\frac{\lambda^3+\lambda^9}{(1-\lambda^2)(1-\lambda^{12})}$	$E^3=\begin{pmatrix}\phi^3\\ \bar\phi^3\end{pmatrix};E^9=i\begin{pmatrix}\bar\phi^9\\ \phi^9\end{pmatrix}$
		$\Gamma_{9,10}$	Γ_1	Same as $\Gamma_{m,n}=\Gamma_{7,8},\Gamma_r=\Gamma_1$.	
			$\Gamma_{2,3}$	Same as $\Gamma_{m,n}=\Gamma_{7,8},\Gamma_{r,\bar r}=\Gamma_{2,3}$	
			Γ_4	Same as $\Gamma_{m,n}=\Gamma_{7,8},\Gamma_r=\Gamma_4$	
			$\Gamma_{5,6}$	Same as $\Gamma_{m,n}=\Gamma_{7,8},\Gamma_{r\bar r}=\Gamma_{5,6}$	
			$\Gamma_{7,8}$	Same as $\Gamma_{m,n}=\Gamma_{7,8},\Gamma_{r\bar r}=\Gamma_{9,10}$	
			$\Gamma_{9,10}$	Same as $\Gamma_{m,n}=\Gamma_{7,8},\Gamma_{r\bar r}=\Gamma_{7,8}$	
			$\Gamma_{11,12}$	Same as $\Gamma_{m,n}=\Gamma_{7,8},\Gamma_{r\bar r}=\Gamma_{11,12}$	
		$\Gamma_{11,12}$	Γ_1	$\frac{1+\lambda^4}{(1-\lambda^2)(1-\lambda^4)}$	$I^2=i\phi\bar\phi;I^4=\phi^4+\bar\phi^4;E^4=i(\phi^4-\bar\phi^4)$
			Γ_4	$\frac{2\lambda^2}{(1-\lambda^2)(1-\lambda^4)}$	$E_a^2=\phi^2+\bar\phi^2;E_b^2=i(\phi^2-\bar\phi^2)$
			$\Gamma_{11,12}$	$\frac{\lambda+\lambda^3}{(1-\lambda^2)(1-\lambda^4)}$	$E^1=\begin{pmatrix}\phi\\ \bar\phi\end{pmatrix}E^3=i\begin{pmatrix}\bar\phi^3\\ \phi^3\end{pmatrix}$
231′	T	$\Gamma_{2,3}$	Γ_1	$\frac{1+\lambda^3}{(1-\lambda^2)(1-\lambda^3)}$	$I^2=\phi\bar\phi;I^3=\phi^3+\bar\phi^3;E^3=i(\phi^3-\bar\phi^3)$
			$\Gamma_{2,3}$	$\frac{\lambda+\lambda^2}{(1-\lambda^2)(1-\lambda^3)}$	$E^1=\begin{pmatrix}\phi\\ \bar\phi\end{pmatrix};\ E^2=\begin{pmatrix}\bar\phi^2\\ \phi^2\end{pmatrix}$
		$\Gamma_{6,7}$	Γ_1	$\frac{1+2\lambda^4+2\lambda^6+2\lambda^8+\lambda^{12}}{(1-\lambda^2)(1-\lambda^4)^2(1-\lambda^6)}$	$I^2=\phi_1\bar\phi_1+\phi_2\bar\phi_2$ $I_a^4=\phi_1^4-2i3^{1/2}\phi_1^2\phi_2^2+\phi_2^4+\bar\phi_1^4+2i3^{1/2}\bar\phi_1^2\bar\phi_2^2+\bar\phi_2^4$ $I_b^4=\phi_1^4-2i3^{1/2}\phi_1^2\phi_2^2+\phi_2^4-i(\bar\phi_1^4+2i3^{1/2}\bar\phi_1^2\bar\phi_2^2+\bar\phi_2^4)$ $I^6=\phi_1\phi_2(\phi_1^4-\phi_2^4)+\bar\phi_1\bar\phi_2(\bar\phi_1^4-\bar\phi_2^4)$
			$\Gamma_{2,3}$	$\frac{3\lambda^4+2\lambda^6+3\lambda^8}{(1-\lambda^2)(1-\lambda^4)^2(1-\lambda^6)}$	
			Γ_4	$\frac{3\lambda^2+5\lambda^4+8\lambda^6+5\lambda^8+3\lambda^{10}}{(1-\lambda^2)(1-\lambda^4)^2(1-\lambda^6)}$	
			Γ_5	$\frac{2\lambda^3+6\lambda^5+6\lambda^7+2\lambda^9}{(1-\lambda^2)(1-\lambda^4)^2(1-\lambda^6)}$	
			$\Gamma_{6,7}$	$\frac{\lambda+3\lambda^3+4\lambda^5+4\lambda^7+3\lambda^9+\lambda^{11}}{(1-\lambda^2)(1-\lambda^4)^2(1-\lambda^6)}$	

The Molien function as given in Eq. (3.8) is

$$B(\mathscr{D}_1,\mathscr{D}_2;\lambda)=(1+\lambda^2)/(1-\lambda^2)^2.$$

The monomials ϕ_1^2,ϕ_2^2, and $\phi_1\phi_2$ are invariant under the unitary operations. The second order invariants under all operations (unitary and anitunitary) of 11′ are then $(\phi_1^2+\phi_2^2)$, $i(\phi_1^2-\phi_2^2)$, and $i\phi_1\phi_2$.

Also

$$B(\mathscr{D}_2,\mathscr{D}_2;\lambda)=\lambda/(1-\lambda^2)^2 \tag{4.14}$$

and

$$E^{(1)}(\mathscr{D}_2,\mathscr{D}_2)=\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}. \tag{4.15}$$

C. Γ_m is type c

With the exception of representations Γ_6 and Γ_7 of T, all other type c representations are one dimensional.

Consider the invariants for these one dimensional cases. The basis functions for the corepresentations are $\{\phi,\bar\phi\}$, where

$$P_{A_0}\phi=\bar\phi \tag{4.16}$$

and

$$P_{A_0}\bar\phi=\Gamma(A_0^2)\phi. \tag{4.17}$$

Since the unitary operations do not "mix" ϕ and $\bar\phi$, the invariants of G are monomials $\phi^k\bar\phi^{\bar k}$, where

$$P_u\phi^k\bar\phi^{\bar k}=\Gamma(u)^k\bar\Gamma(u)^{\bar k}(\phi^k\bar\phi^{\bar k}). \tag{4.18}$$

Under antiunitary operations

$$P_{A_0}c\phi^k\bar\phi^{\bar k}=c^*\Gamma(A_0^2)^{\bar k}\bar\phi^k\phi^{\bar k}. \tag{4.19}$$

In particular, for one-dimensional complex representations from Sec. IIIC the generating function is

$$B(\Gamma_r,\Gamma_m\oplus\Gamma_m^*;\lambda)=(\lambda^t+\lambda^s)/(1-\lambda^2)(1-\lambda^z)$$

and for $r=1$ (4.20)

$$B(\Gamma_1,\Gamma_m\oplus\Gamma_m^*;\lambda)=(1+\lambda^z)/(1-\lambda^2)(1-\lambda^z).$$

TABLE IV. Black-white groups: $M = G + \theta u_0 G, u_0 \notin G$.

M	$M'(G)$	Type a representations	Type b representations	Type c representations
$\bar{1}'$	$C_i(C_1)^a$			
$2'$	$C_2(C_1)$	All		
m'	$C_s(C_1)$	All		
$2/m'$	$C_{2h}(C_2)^a$			
$2'/m$	$C_{2h}(C_s)^a$			
$2'/m$	$C_{2h}(C_i)$	All		
$22'2'$	$D_2(C_2)$	All		
$2m'm'$	$C_{2v}(C_2)$	All		
$2'm'm$	$C_{2v}(C_s)$	All		
$m'm'm'$	$D_{2h}(D_2)^a$			
mmm'	$D_{2h}(C_{2v})^a$			
$m'm'm$	$D_{2h}(C_{2h})$	All		
$4'$	$C_4(C_2)^b$	Γ_1	Γ_2	Γ_3,Γ_4
$\bar{4}'$	$S_4(C_2)^b$	Γ_1	Γ_2	Γ_3,Γ_4
$42'2'$	$D_4(C_4)$	All		
$4'22'$	$D_4(D_2)$	$\Gamma_1,\Gamma_4,\Gamma_5{}^c$		Γ_2,Γ_3
$4/m'$	$C_{4h}(C_4)^a$			
$4'/m'$	$C_{4h}(S_4)^a$			
$4'/m$	$C_{4h}(C_{2h})^b$	Γ_1^+	Γ_2^+	Γ_3^+,Γ_4^+
$4m'm'$	$C_{4v}(C_4)$	All		
$4'mm'$	$C_{4v}(C_{2v})$	$\Gamma_1,\Gamma_4,\Gamma_5{}^c$		Γ_2,Γ_3
$\bar{4}2'm'$	$D_{2d}(S_4)$	All		
$\bar{4}'2m'$	$D_{2d}(D_2)$	$\Gamma_1,\Gamma_4,\Gamma_5{}^c$		Γ_2,Γ_3
$\bar{4}'2'm$	$D_{2d}(C_{2v})$	$\Gamma_1,\Gamma_4,\Gamma_5{}^c$		Γ_2,Γ_3
$4/m'm'm'$	$D_{4h}(D_4)^a$			
$4/m'mm$	$D_{4h}(C_{4v})^a$			
$4'/mmm$	$D_{4h}(D_{2h})$	$\Gamma_1^+,\Gamma_4^+,\Gamma_5^{+\ c}$		Γ_2^+,Γ_3^+
$4'/m'm'm$	$D_{4h}(D_{2d})^a$			
$4/mm'm'$	$D_{4h}(C_{4h})$	All		
$32'$	$D_3(C_3)$	All		
$3m'$	$C_{3v}(C_3)$	All		
$\bar{6}'$	$C_{3h}(C_3)^b$	Γ_1,Γ_6		Γ_2,Γ_3 Γ_4,Γ_5
$\bar{6}m'2'$	$D_{3h}(C_{3h})$	All		
$\bar{6}'m2'$	$D_{3h}(C_{3v})$	All		
$\bar{6}'m'2\cdot$	$D_{3h}(D_3)$	All		
$6'$	$C_6(C_3)^b$	Γ_1,Γ_6		Γ_2,Γ_3 Γ_4,Γ_5
$\bar{3}'$	$C_{3i}(C_3)^a$			
$\bar{3}m'$	$D_{3d}(C_{3i})$	All		
$\bar{3}'m$	$D_{3d}(C_{3v})^a$			
$\bar{3}'m'$	$D_{3d}(D_3)^a$			
$62'2'$	$D_6(C_6)$	All		
$6'22'$	$D_6(D_3)$	All		
$6/m'$	$C_{6h}(C_6)^a$			
$6'/m'$	$C_{6h}(C_{3i})^b$	Γ_1^+,Γ_6^+		Γ_2^+,Γ_3^+ Γ_4^+,Γ_5^+
$6'/m$	$C_{6h}(C_{3h})^a$			
$6m'm'$	$C_{6v}(C_6)$	All		
$6'mm'$	$C_{6v}(C_{3v})$	All		
$6'/mm'm$	$D_{6h}(D_{3h})^a$			
$6'/m'm'm$	$D_{6h}(D_{3d})$	All		
$6/m'm'm'$	$D_{6h}(D_6)^a$			
$6/m'mm$	$D_{6h}(C_{6v})^a$			
$6/mm'm'$	$D_{6h}(C_{6h})$	All		
$m'3$	$T_h(T)^a$			
$\bar{4}'3m'$	$T_d(T)$	All, $\Gamma_4,{}^c\Gamma_5,{}^c\Gamma_6,{}^c\Gamma_7{}^c$		
$4'32'$	$O(T)$	All, $\Gamma_4,{}^c\Gamma_5,{}^c\Gamma_6,{}^c\Gamma_7{}^c$		
$m'3m'$	$O_h(O)^a$			
$m'3m$	$O_h(T_d)^a$			
$m3m'$	$O_h(T_h)$	All, $\Gamma_4,{}^c\Gamma_5,{}^c\Gamma_6,{}^c\Gamma_7{}^c$		

[a] For these groups $u_0 = I$ (the inversion element) and the corepresentations are the same as for the grey group of G.

[b] For these groups u_0 commutes with all u and the Molien functions are the same as for the grey group of G.

[c] For these representations $\beta \neq \Gamma(u)$.

All type c corepresentations $\mathscr{D}_m$ based on one dimensional complex representations Γ_m and Γ_m^* have the following invariants:

$$I^2(\mathscr{D}_1,\mathscr{D}_m) = (c + c^*\Gamma(A_0^2))\phi\bar{\phi},$$
$$I^z(\mathscr{D}_1,\mathscr{D}_m) = \phi^z + \bar{\phi}^z, \tag{4.21}$$

and

$$E^z(\mathscr{D}_1,\mathscr{D}_m) = i(\phi^z - \bar{\phi}^z).$$

Similarly, for $E^p(\mathscr{D}_r,\mathscr{D}_m)$ covariant tensors, the following generalizations can be made.

If $\mathscr{D}_r$ is one dimensional (and, therefore, real) then

$$s = t = z/2$$

and

$$E_a^s(\mathscr{D}_r,\mathscr{D}_m) = \phi^s + \bar{\phi}^s$$

and

$$E_b^s(\mathscr{D}_r,\mathscr{D}_m) = i(\phi^s - \bar{\phi}^s). \tag{4.22}$$

If $\mathscr{D}_r$ is a two-dimensional type c corepresentation the covariants are

$$E^s = c_s\begin{pmatrix}\phi^s\\ \bar{\phi}^s\end{pmatrix} \text{ and } E^t = c_t\begin{pmatrix}\bar{\phi}^t\\ \phi^t\end{pmatrix}. \tag{4.23}$$

For example, consider the black–white group 4′ for which $M = G + \theta C_{4z}^+ G$ and $G = C_2$. The corepresentations formed from the type c double-valued representations Γ_3 and Γ_4 of C_2 are given as follows:

$$\mathscr{D}(E) = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix};\quad \mathscr{D}(c_{2z}) = \begin{pmatrix}i & 0\\ 0 & -i\end{pmatrix};$$
$$\mathscr{D}(\bar{E}) = \begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix};$$
$$\mathscr{D}(\bar{c}_{2z}) = \begin{pmatrix}-i & 0\\ 0 & i\end{pmatrix};\quad \mathscr{D}(A_0) = \begin{pmatrix}0 & -i\\ 1 & 0\end{pmatrix}. \tag{4.24}$$

Note that $\Gamma(A_0^2) = -i$. The Molien function is given by

$$B(\mathscr{D}_1,\mathscr{D}_{3,4};\lambda) = (1+\lambda^4)/(1-\lambda^4)(1-\lambda^2). \tag{4.25}$$

The monomials $\phi\bar{\phi}$, ϕ^4, and $\bar{\phi}^4$ are invariant under the unitary operations of C_2. The invariants under M are then $(1-i)\phi\bar{\phi}$, $(\phi^4 + \bar{\phi}^4)$, and $i(\phi^4 - \bar{\phi}^4)$.

Also,

$$B(\mathscr{D}_{3,4},\mathscr{D}_{3,4};\lambda) = (\lambda + \lambda^3)/(1-\lambda^2)(1-\lambda^4) \tag{4.26}$$

and

$$E^1(\mathscr{D}_{3,4},\mathscr{D}_{3,4}) = \begin{pmatrix}\phi\\ \bar{\phi}\end{pmatrix};\quad E^3(\mathscr{D}_{3,4},\mathscr{D}_{3,4}) = \begin{pmatrix}\bar{\phi}^3\\ \phi^3\end{pmatrix}. \tag{4.27}$$

V. SUMMARY OF RESULTS FOR MAGNETIC POINT GROUPS

The corepresentations of the magnetic point groups have been given by Cracknell[11] for the single-valued representations and Cracknell and Wong[12] for the double-valued representations and are also given in Bradley and Cracknell.[13] Molien functions, invariants and covariants for these corepresentations have been determined using the procedures of the previous sections. The results are summarized in Tables II–V and Secs. VA and VB below.

A. Grey groups

Tables II and III give the results for grey groups ($M = G + \theta G$). These tables do not include those groups obtainable by direct multiplication of the groups given and the inversion group. The Molien functions and invariants for such groups can be obtained from those given as described in Ref. 4. Table II lists the types of representations of the grey groups. The labeling of the representations Γ_j is that of Ref. 14. The Molien functions for type a representations are found in Refs. 4 and 5 and those for type b representations

TABLE V. Molien functions and invariants and covariants for some type c black–white group representations.

M	G	$\Gamma_{m,n}$	Γ_r	Molien function	Invariants and covariants
4′22′	D_2	$\Gamma_{2,3}$	Γ_1	$\frac{1}{(1-\lambda^2)^2}$	$I_a^2 = \phi^2 + \bar{\phi}^2; I_b^2 = i(\phi^2 - \bar{\phi}^2)$
4′mm′	C_{2v}				
$\bar{4}$′2m′	D_2				
$\bar{4}$′2′m	C_{2v}		$\Gamma_{2,3}$	$\frac{\lambda}{(1-\lambda^2)^2}$	$E^1 = \begin{pmatrix}\phi\\ \bar{\phi}\end{pmatrix}.$
			Γ_4	$\frac{\lambda^2}{(1-\lambda^2)^2}$	$E^2 = \phi\bar{\phi}$
4′/mmm	D_{2h}	$\Gamma_{2',3'}$	$\Gamma_{1'}$	$\frac{1}{(1-\lambda^2)^2}$	$I_a^2 = \phi^2 + \bar{\phi}^2; I_b^2 = i(\phi^2 - \bar{\phi}^2)$
			$\Gamma_{2',3'}$	$\frac{\lambda}{(1-\lambda^2)^2}$	$E^1 = \begin{pmatrix}\phi\\ \bar{\phi}\end{pmatrix}$
			$\Gamma_{4'}$	$\frac{\lambda^2}{(1-\lambda^2)^2}$	$E^2 = \phi\bar{\phi}$
		$\Gamma_{2\ ,3}$	$\Gamma_{1'}$	$\frac{1}{(1-\lambda^2)^2}$	$I_a^2 = \phi^2 + \bar{\phi}^2; I_b^2 = i(\phi^2 - \bar{\phi}^2)$
			$\Gamma_{2\ ,3}$	$\frac{\lambda}{(1-\lambda^2)^2}$	$E^1 = \begin{pmatrix}\phi\\ \bar{\phi}\end{pmatrix}$
			$\Gamma_{4'}$	$\frac{\lambda^2}{(1-\lambda^2)^2}$	$E^2 = \phi\bar{\phi}$

are given in Eq. (3.8) of this paper. Invariants and covariants for type a representations are also given in Refs. 4 and 5 except for those representations for which $\beta \neq \Gamma(u)$ when Eq. (4.2) must be satisfied. Such representations are indicated by footnote a in Table II. All type b invariants are as in the example of Sec. IV.

Table III lists the Molien functions and invariants of type c representations of grey groups. Column 1 gives the nonunitary group M while column 2 gives G, the subgroup of unitary operations. Columns 3 and 4 give $\mathscr{D}_m$ and $\mathscr{D}_r$, respectively, where for $\mathscr{D}_m = (\Gamma_m \oplus \overline{\Gamma}_m) = (\Gamma_m \oplus \Gamma_n)$, the entry reads $\Gamma_{m,n}$. Column 5 gives the Molien function $B(\mathscr{D}_r,\mathscr{D}_m;\lambda)$ and column 6 gives the appropriate invariant tensors. $E^p(\mathscr{D}_r,\mathscr{D}_m)$ and $I^q(\mathscr{D}_1,\mathscr{D}_m)$ are written for simplicity as E^p and I^q and the bases of $\mathscr{D}_m = \Gamma_m \oplus \overline{\Gamma}_m$ are always written as $\{\phi,\bar{\phi}\}$. Note that for each $\mathscr{D}_m$ the table includes only those $\mathscr{D}_r$ for which $B(\mathscr{D}_r,\mathscr{D}_m;\lambda)$ is nonzero. In addition for $\Gamma_6 \oplus \Gamma_7$ of $231'(T + \theta T)$, only the denominator invariants, $I^q(\mathscr{D}_1,\mathscr{D}_m)$, have been given for the generating matrices given in Ref. 5.

B. Black–white groups

The black–white groups are listed in Table IV. The groups have been labeled by both International and Schoenflies notation. In particular, for $M = G + u_0\theta G$, the Schoenflies notation is $M'(G)$, where $M' = G + u_0 G$. In order to simplify the table, note that for $u_0 = I$ (the inversion operator), the corepresentations of M and, therefore, the Molien functions and invariants, are the same as for the grey group of G. Twenty-one of the 58 black–white groups have this property[15] and are so indicated by footnote a in column 2 of Table IV. In addition, 26 other black–white groups have only type a representations so that their Molien functions and invariants are obtainable from Refs. 4 and 5 [with the application of Eq. (4.2) when appropriate].

For the remaining groups, if u_0 commutes with all u, the Molien function again is the same as for the grey group of G. However, the coefficients of the invariant polynomials [such as the "c's" in Eq. (4.21)] may be different in order to have invariance under antiunitary operations. These groups are indicated by footnote b in column 2 of Table IV.

Finally, there are five black–white groups wherein u_0 does not commute with all u. The Molien functions and invariants and covariants for these groups are given in Table V.

ACKNOWLEDGMENT

The author wishes to thank Professor Joseph L. Birman for suggesting this problem and for his long discussions and valuable comments.

[1] T. Molien, Sitzungber, König, Preuss, Akad. Wiss. 1152 (1897).
[2] W. Burnside, *Theory of Groups of Finite Order*, 2nd ed. (Dover, New York, 1955).
[3] L. Michel, in *Proceedings of the Vth International Colloquium on Group Theoretical Methods in Physics*, edited by R. T. Sharp and B. Kolman (Academic, New York, 1977).
[4] J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **19**, 2362 (1978).
[5] P. E. Desmier and R. T. Sharp, J. Math. Phys. **20**, 74 (1979).
[6] M. J. Jarić and J. L. Birman, J. Math. Phys. **18**, 1456,1459 (1977).
[7] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
[8] Y. Saint-Aubin, Can. J. Phys. **58**, 1075 (1980).
[9] V. Kopský, J. Phys. A **12**, 943 (1979).
[10] G. F. Karavaev, Soviet Phys.-Solid State **6**, 2943 (1965).
[11] A. P. Cracknell, Prog. Theor. Phys. Kyoto **35**, 196 (1966).
[12] A. P. Cracknell and K. C. Wong, Aust. J. Phys. **20**, 173 (1967).
[13] C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972).
[14] G. F. Koster, J. O. Dimmock, R. G. Wheeler, and H. Statz, *Properties of the Thirty-Two point Groups* (M.I.T., Cambridge, 1963).
[15] J. O. Dimmock and R. G. Wheeler, *The Mathematics of Physics and Chemistry*, Vol. II, edited by H. Margenau and G. M. Murphy (Van Nostrand, New York, 1964).

Generalized SU(2) spherical harmonics

J. Bystricky
Centre d'Etudes Nucléaires, Saclay, Gif-sur-Yvette, France
R. Gaskell
Department of Physics, Lafayette College, Easton, Pennsylvania 18042
J. Patera
Centre de Recherches de Mathématiques Appliquées, Université de Montréal, Montréal, Québec, Canada
R. T. Sharp
Department of Physics, McGill University, Montreal, Quebec, Canada

(Received 5 February 1982; accepted for publication 29 April 1982)

The generating functions for polynomial tensors based on each SU(2) tensor of rank from 7 to 13 (angular momentum 7/2 to 13/2) are given in a "positive" form suitable for interpretation in terms of an integrity basis. An iterative procedure for extending the results to higher rank tensors is indicated.

PACS numbers: 02.30. + g

1. INTRODUCTION

A problem which arises in many contexts in mathematical physics is that of determining all irreducible SU(2) tensors whose components are homogeneous polynomials in the components of a fixed irreducible tensor of rank L (and dimension $L+1$; the corresponding angular momentum is $\frac{1}{2}L$). We call such polynomial tensors generalized spherical harmonics or, more specifically, L-harmonics. For $L=1$ they are Wigner monomials, and for $L=2$, if one discards those containing the quadratic scalar as a factor, they are the familiar spherical harmonics.

Over a century ago Cayley, Sylvester, and Franklin[1-4] gave a sequence of generating functions $F_L(U,A)$, $L=0,1,\ldots,12$ which enumerate L-harmonics. The power series expansion of the rational function

$$F_L(U,A)=\sum_{u,a} m_{ua}^L\ U^u A^a \tag{1.1}$$

provides the number of linearly independent L-harmonics of degree u and rank a as the expansion coefficient m_{ua}^L.[5]

To find the explicit algebraic form of the L-harmonics, an essential step is the determination of their integrity basis, a finite number of L-harmonics, called elementary tensors, in terms of which all can be expressed as stretched products. A serious drawback of the old Cayley–Sylvester–Franklin generating functions is that the value of a particular m_{ua}^L is the result of a cancellation involving terms of both signs. This makes the computation of m_{ua}^L cumbersome, but, more importantly, it obscures the form of the integrity basis. In this paper we rederive the generating functions $F_L(U,A)$ for $L=0,1,\ldots,13$ in a "positive" form. All contributions to each m_{ua}^L are positive, a circumstance which makes it possible to read the degrees and ranks of the integrity basis elements, as well as the the existence of syzygies (polynomial identities) relating them.

In Sec. 2 we present the new forms of the generating functions. In Sec. 3 is found an example of their interpretation and the explicit construction of an integrity basis. Section 4 contains an explanation of their derivation. Some concluding remarks are made in Sec. 5.

Symmetries discovered by Murnaghan[6] imply that m_{ua}^L is also the multiplicity of u-harmonics of degree L and rank a, and, moreover, is the multiplicity of rank-a tensors which are completely antisymmetric in the components of L or u copies of a tensor of rank $L+u-1$.

The L-harmonics provide polynomial bases for symmetric representations $(u,0,0,\ldots,0)$ of SU$(L+1)$, or, for L odd, of Sp$(L+1)$, reduced according to the principal SU(2) subgroup. For L even they play a similar role for O$(L+1)$, or, with $L=6$, for G_2; in these cases they must be rendered traceless by projecting out terms containing the quadratic scalar as a factor.[7] For $L=4$ and $L=6$ such states serve to describe quadrupole and octupole nuclear vibrations.[8-11] Rohoziński and Greiner[12] consider the extension of the problem to higher even L.

Another application of generalized spherical harmonics is the construction of missing label operators for any semisimple group G reduced to its SU(2) subgroup. Such a missing label operator is an SU(2) scalar polynomial in the generators of G, independent of the G and SU(2) Casimir invariants. The G generators, reduced according to SU(2), consist of an $L=2$ tensor [the SU(2) generators] and a second SU(2) tensor T which may or may not be reducible. Then the missing label operators, and G Casimir invariants, correspond precisely to the SU(2) polynomial tensors form from T [a $2a$-tensor from T must be contracted with the $2a$-tensor of degree a in the SU(2) generators]. If T is a single irreducible L-tensor, the missing label operators are enumerated by $F_L(U,A)$; this is the case for SU(3) with $L=4$, for O(5) with $L=6$, and for G_2 with $L=10$. If T is reducible the generating functions for the L's which comprise it must be combined by a procedure described in Sec. 5.

Generating functions $F_L(U,A)$ based on the representations $L=4,8,12,\ldots$ are needed in the study of bifurcations in the Bénard problem.[13-16]

2. THE GENERATING FUNCTIONS

The generating function for SU(2) tensors contained in the symmetric product of an arbitrary number of identical SU(2) tensors is defined in Eq. (1.1). With the methods out-

0022-2488/82/091560-06$02.50

TABLE I. Invariant exponents n_i^L.

L \ i	1	2	3	4	5	6	7	8	9	10	11
7	4	12	8	20	12						
8	2	3	4	5	6	7					
9	4	12	8	10	12	14	16				
10	2	6	4	10	6	14	8	9			
11	4	12	8	10	12	14	16	18	20		
12	2	3	4	5	6	7	8	9	10	11	
13	4	12	8	10	12	14	16	18	20	22	24

lined in Sec. 4 we obtain, for $L > 2$,

$$F_L(U,A) = \sum_{k=1}^{l} N_k^L \Big[(1 - UA^L) \prod_{i=1}^{\lambda+k-2} \times (1 - U^{n_i^L}) \prod_{j=1}^{l-k+1} (1 - U^2 A^{2L-4j}) \Big]^{-1}, \quad (2.1)$$

where l and $\lambda = L - l$ are the greatest integers in $(L-1)/2$ and $(L+2)/2$, respectively. For $k < l$ the numerators have the form

$$N_k^L = \prod_{m=1}^{l-k+1} (1 + UA^{L-2m}) \sum B_{ij}^{Lk} U^i A^j \quad (k < l), \quad (2.2)$$

while for $k = l$,

$$N_l^L = (1 + U^3 U^{3L-6}) \sum B_{i0}^{Ll} U^i + (1 + UA^{L-2}) \sum_{j \neq 0} B_{ij}^{Ll} U^i A^j. \quad (2.3)$$

The exponents n_i^L and the $\Sigma B_{ij}^{Lk} U^i A^j$ with $k = 1$ are tabulated in Tables I and II. The remaining B_{ij}^{Lk} are tabulated in Tables III–VII.[17] The tabulation for $k = l$ can be simplified by two symmetry relations. First,

$$B_{i+2,2L-4-j}^{Ll} = B_{ij}^{Ll} \quad (j = 1,2,\ldots,L-3) \quad (2.4)$$

allows us to terminate the tabulation at $j = L - 2$ without loss of information. Second, the B_{ij}^{Ll} can be shown to have reflection symmetry about some $i = i_0$,[18] that is,

$$B_{i_0+i,j}^{Ll} = B_{i_0-i,j}^{Ll}. \quad (2.5)$$

A further simplification occurs for odd L. There, the exponents of U and A must have the same parity so that $i + j$ is always even. We have therefore reduced the odd L tables by tabulating B_{ij}^{Lk} for even j and $B_{i-1,j}^{Lk}$ for odd j. This means that 1 must be subtracted from the indicated i when j is odd.

TABLE II. Summations appearing in first term numerators.

L	$\sum B_{ij}^{L1} U^i A^j$
7	$U^6(1+U^6)A^6$
8	U^4A^8
9	$U^8(1+U^6)A^8$
10	$U^5(1+U^3)(1+U^5)A^{10}$
11	$U^{10}(1+U^6)A^{10}$
12	U^6A^{12}
13	$U^{12}(1+U^6)A^{12}$

TABLE III. B_{ij}^{Lk} for $L = 7$. Subtract 1 from i for odd j.

	k=2			k=3					
i \ j	6	7	8	0	1	2	3	4	5
0	0	0	0	1	0	0	0	0	0
2	1	0	0	0	0	1	0	0	0
4	1	0	1	0	0	0	1	3	1
6	1	2	2	0	1	3	2	2	3
8	3	2	3	2	3	3	3	7	6
10	5	4	4	0	3	7	8	11	10
12	6	6	6	4	5	9	11	16	16
14	7	6	7	4	9	13	14	22	23
16	8	8	8	5	9	18	21	28	30
18	9	8	9	9	12	19	24	33	37
20	9	9	8	6	15	23	28	37	42
22	8	9	9	9	14	23	32	42	47
24	8	8	8	8	17	26	32	38	49
26	7	7	7	9	17	23	32	42	49
28	6	7	6	6	14	23	32	37	47
30	4	5	4	9	15	19	28	33	42
32	3	3	3	5	12	18	24	28	37
34	2	2	2	4	9	13	21	22	30
36	0	1	1	4	9	9	14	16	23
38	0	1	0	0	5	7	11	11	16
40				2	3	3	8	7	10
42				0	3	3	3	2	6
44				0	1	0	2	3	3
46				0	0	1	1	0	1
48				1	0	0	0	0	0

It is clear from (2.1) that there are exactly L denominator factors, $L - 2$ in the scalar $(A = 0)$ part.

3. AN EXAMPLE

As a simple example of the use of the generating function $F_L(U,A)$, for $L = 3$, we interpret it in terms of an integrity basis and find the algebraic form of the elementary tensors. For tensor $L = 4,6$, the states describe quadrupole and octupole nuclear vibrations and the integrity bases have been discussed in that context.[9–11]

For $L = 3$ we have the generating function

$$F_L(U,A) = (1 + U^3A^3)[(1 - U^4)(1 - UA^3)(1 - U^2A^2)]^{-1} \quad (3.1)$$

for which the power series expansion is

TABLE IV. B_{ij}^{Lk} for $L = 8$.

	k=2		k=3			
i \ j	8	10	0	2	4	6
0	0	0	1	0	0	0
1	0	0	0	0	0	0
2	1	0	0	0	1	0
3	1	0	0	0	1	1
4	0	1	0	0	2	1
5	1	1	0	1	2	2
6	1	1	0	1	2	3
7	1	2	0	2	2	4
8	1	2	1	2	2	4
9	1	1	1	3	1	5
10	1	1	1	2	2	4
11	1	1	0	2	2	4
12	1	0	0	1	2	3
13			0	1	2	2
14			0	0	2	1
15			0	0	1	1
16			0	0	1	0
17			0	0	0	0
18			1	0	0	0

TABLE V. B_{ij}^{Lk} for $L = 9$. Subtract 1 from i for odd j.

	k=2			k=3					k=4							
i \ j	8	9	10	8	9	10	11	12	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0
4	1	1	0	2	0	1	1	1	1	0	0	1	3	1	2	1
6	2	1	2	4	3	4	2	4	0	1	5	3	4	5	11	5
8	4	3	3	11	6	11	8	11	5	4	8	9	18	12	20	16
10	6	4	6	25	18	23	15	22	4	10	25	21	33	29	53	38
12	11	8	9	45	33	43	32	44	17	21	39	42	75	62	97	82
14	13	11	14	74	59	74	57	72	20	39	81	78	125	117	182	154
16	17	14	16	115	94	112	91	110	47	67	123	133	216	199	288	264
18	18	17	18	159	136	158	131	156	61	108	199	209	318	314	454	420
20	19	19	19	207	181	206	182	202	97	157	272	309	469	466	639	621
22	18	19	18	251	230	249	226	245	120	215	378	430	620	648	874	862
24	14	17	15	283	265	281	266	281	165	284	473	563	797	848	1106	1132
26	12	14	11	300	291	300	293	297	189	353	580	701	957	1053	1348	1407
28	7	10	9	300	302	297	300	297	223	413	660	831	1109	1245	1549	1660
30	5	7	5	278	289	277	291	281	241	464	729	934	1211	1401	1707	1869
32	2	3	3	242	263	245	264	245	254	499	762	1002	1270	1502	1784	2004
34	1	2	1	199	223	199	223	202	254	510	762	1028	1270	1536	1784	2050
36	0	0	1	149	176	152	176	156	241	499	729	1002	1211	1502	1707	2004
38				105	129	109	132	110	223	464	660	934	1109	1401	1549	1869
40				68	88	70	88	72	189	413	580	831	957	1245	1348	1660
42				39	52	41	56	44	165	353	473	701	797	1053	1106	1407
44				20	30	23	31	22	120	284	378	563	620	848	874	1132
46				10	14	9	16	11	97	215	272	430	469	648	639	862
48				3	6	4	6	4	61	157	199	309	318	466	454	621
50				1	1	1	3	1	47	108	123	209	216	314	288	420
52				0	1	0	0	0	20	67	81	133	125	199	182	264
54									17	39	39	78	75	117	97	154
56									4	21	25	42	33	62	53	82
58									5	10	8	21	18	29	20	38
60									0	4	5	9	4	12	11	16
62									1	1	0	3	3	5	2	5
64									0	0	1	1	0	1	1	1
66									1	0	0	0	0	0	0	0

$$1 + UA^3 + U^2(A^2 + A^6) + U^3(A^3 + A^5 + A^9)$$
$$+ U^4(1 + A^4 + A^6 + A^8 + A^{12}) + \cdots .$$

One reads that, for example, there are five tensors of degree 4, whose ranks are 0,4,6,8,12. The elementary tensors can be read from (3.1); they are (1,3), (2,2), (3,3), and (4,0) in an obvious notation. Any polynomial tensor based on $L = 3$ is a stretched product of powers of these; its highest component is the product of highest components of elementary tensors. The fact that only the first power of (3,3) appears in the expansion of the generating function implies a syzygy (polynomial relationship) expressing its square as a sum of products of powers of the other elementary tensors. In terms of elementary tensors, the tensors of degree 4 listed above are, respectively, (4,0), $(2,2)^2$, (1,3) (3,3), $(1,3)^2$ (2,2), $(1,3)^4$.

Since their degrees and ranks are known, it is straightforward to determine the analytic form of the elementary tensors. Their highest components are, respectively,

$$(1,3)\sim\alpha_3,\ \ (2,2)\sim\alpha_1^2 - \sqrt{3}\,\alpha_3\,\alpha_{-1},$$
$$(3,3)\sim 3\sqrt{3}\,\alpha_3^2\,\alpha_{-3} - 3\sqrt{3}\,\alpha_3\,\alpha_1\,\alpha_{-1} + 2\alpha_1^3,$$
$$(4,0)\sim 3\sqrt{3}\,\alpha_3^2\,\alpha^2_{-3} + 4\,\alpha_3\,\alpha^3_{-1}$$
$$-6\sqrt{3}\,\alpha_3\,\alpha_1\,\alpha_{-1}\,\alpha_{-3} + 4\,\alpha_1^3\,\alpha_{-3} - \sqrt{3}\,\alpha_1^2\,\alpha^2_{-1},$$

where $\alpha_3, \alpha_1, \alpha_{-1}, \alpha_{-3}$ are the components of the basic rank-3 tensor. The highest components are related by the syzygy

$$(3,3)^2 - 4(2,2)^3)3\sqrt{3}\,(1,3)^2(4,0) = 0.$$

For higher L the interpretation of $F_L(U,A)$ is straightforward but more tedious.

4. CONSTRUCTION OF $F_L(U,A)$

The construction of the generating function $F_L(U,A)$ begins with the generator for weights of symmetric products of an arbitrary number of copies of a single SU(2) tensor of rank L

$$G_L(U,\eta) = \sum a^L_{um}\, U^u \eta^m = \left[\prod_{i=0}^{L}(1 - U\eta^{L-2i})\right]^{-1}, \tag{4.1}$$

where a^L_{um} is the multiplicity of the weight m in a product of u rank L tensors (the weights are double the spin projections). Of the a^L_{um} weights m, some will arise from SU(2) tensors of rank m while some will come from higher rank tensors. Since these higher tensors each contain the weight $m+2$, the difference $a^L_{um} - a^L_{um+2}$ is just the number of tensors of rank m contained in the symmetric product of u rank L tensors. It is easy to see, then, that the desired generating function $F_L(U,\eta)$ can be obtained by multiplying $G_L(U,\eta)$ by $(1-\eta^{-2})$ and retaining only non-negative powers of η. Replacing η by A we have

$$F_L(U,A) = P_A\left[(1 - A^{-2}) \Big/ \prod_{i=0}^{L}(1 - UA^{L-2i})\right], \tag{4.2}$$

TABLE VI. B_{ij}^{Lk} for $L = 10$.

	k=2		k=3			k=4				
j i	10	12	10	12	14	0	2	4	6	8
0	0	1	0	0	0	1	0	0	0	0
1	1	1	0	0	0	0	0	0	0	0
2	1	1	0	1	0	0	0	1	0	1
3	2	2	1	0	1	0	1	0	2	1
4	4	2	1	1	1	0	0	3	1	4
5	4	4	3	2	3	0	3	3	5	5
6	6	5	4	4	4	2	2	6	8	11
7	7	6	8	8	7	0	7	10	14	16
8	8	8	12	11	11	4	8	16	20	28
9	10	9	17	17	17	4	15	21	34	40
10	11	11	24	25	24	7	20	35	45	59
11	12	11	33	33	31	8	28	45	65	81
12	13	12	44	43	42	15	38	61	87	111
13	14	13	52	53	53	15	49	80	112	143
14	13	13	66	65	65	20	61	102	142	182
15	13	14	77	77	75	27	75	123	176	223
16	13	13	89	87	87	29	90	149	208	268
17	12	13	98	99	97	35	105	175	243	314
18	11	12	108	106	108	40	119	197	281	357
19	10	10	115	116	113	44	132	224	309	402
20	9	9	120	118	119	47	147	242	340	437
21	7	8	122	121	122	55	154	260	364	471
22	6	7	121	122	122	52	164	273	383	491
23	4	5	120	119	119	57	166	284	391	509
24	3	4	114	113	113	56	172	280	400	512
25	3	3	106	105	108	57	166	284	391	509
26	1	2	96	98	97	52	164	273	383	491
27	1	1	87	86	87	55	154	260	364	471
28	1	0	74	76	75	47	147	242	340	437
29			63	62	65	44	132	224	309	402
30			51	53	53	40	119	197	281	357
31			41	41	42	35	105	175	243	314
32			31	32	31	29	90	149	208	268
33			23	23	24	27	75	123	176	223
34			15	17	17	20	61	102	142	182
35			10	12	11	15	49	80	112	143
36			7	6	7	15	38	61	87	111
37			3	5	4	8	28	45	65	81
38			2	2	3	7	20	35	45	59
39			1	1	1	4	15	21	34	40
40			1	0	1	4	8	16	20	28
41						0	7	10	14	16
42						2	2	6	8	11
43						0	3	3	5	5
44						0	0	3	1	4
45						0	1	0	2	1
46						0	0	1	0	1
47						0	0	0	0	0
48						1	0	0	0	0

where P_A means non-negative powers part of. The quantity in square brackets is the same as that for the case $L-2$ but with extra denominators $(1 - UA^{\pm L})$. This allows us to simplify the evaluation of (4.2) by use of a recursion procedure described below which generates F_L from F_{L-2}. The derivation of this procedure is given elsewhere.[18]

As the first step in determining F_L we construct the function

$$R_L(U,A) = P_A\, F_{L-2}(U,A)/(1 - UA^{-L}) = \sum_{i=0}^{L-2} r_i(U)A^i + \tilde{R}_L(U,A), \qquad (4.3)$$

where the expansion of $\tilde{R}_L$ contains only terms with exponent of A [denoted hereafter by EX(A)] $> L-2$. We use the r_i obtained above to construct

$$q_i(U) = (U^2 r_i - U r_{L-2-i})/(1 - U^2). \qquad (4.4)$$

The generating function F_L is then given by

$$F_L(U,A) = \Big[R_L(U,A) + \sum_{i=0}^{L-2} q_i(U)A^i \Big] \Big/ (1 - UA^L). \qquad (4.5)$$

The solutions to this procedure are best put in the standard form

$$F_L(U,A) = \sum_{k=0}^{l} S_k^L \Big[\prod_{i=2}^{\lambda+k-1} (1 - U^{2i}) \times \prod_{j=0}^{l-k+1} (1 - UA^{L-2j;} \Big]^{-1}, \qquad (4.6)$$

where

$$S_k^L = \sum_{ij} s_{ij}^{Lk}\, U^i A^j. \qquad (4.7)$$

For $k < l$ the range of EX(A) in S_k^L is $L - 1 \leqslant j \leqslant 2\lambda + 2k - 4$ while for $k = l$ we have $0 \leqslant j \leqslant 2L - 4$. In addition, since $F_L(U,A) \sim U^{-L-1}$ as $U \rightarrow \infty$, we find that i must lie in the range $0 \leqslant i \leqslant (\lambda + k - 1)^2 - 2$. The form (4.6) has the advantage that when the k th term of $F_{L-2}/(1 - UA^{-L})$ is split into one term with no negative powers of A in its expansion and a remainder term, the first term can be arranged to have EX(A) $> L - 2$ and will not contribute to the q_i while the remainder will have the same form as the $(k+1)$st term of $F_{L-2}/(1 - UA^{-L})$.

The k th term of $F_{L-2}/(1 - UA^{-L})$ plus the remainder from the reduction of the $(k-1)$st term contains a factor $(1 - UA^{L+2k-2l-2})$ in the denominator which is not contained in the $(k+1)$st term. The $(k+1)$st denominator contains the factor $(1 - U^{2(\lambda+k-1)}) = (1 - U^{2(L+k-l-1)})$ which is not included in the k th denominator. To split the k th term as described above we multiply and divide by this latter factor and use for the numerator factor

$$(1 - U^{2(L+k-l-1)}) = (1 - (UA^{-L})^{L+2k-2l-2}) + (UA^{-L})^{L+2k-2l-2} \times (1 - (UA^{L+2k-2l-2})^L). \qquad (4.8)$$

Some rearrangement is required to obtain EX(A) $> L - 2$ in the expansion of the first of these terms. After we have applied this procedure to the $l - 1$ terms of $F_{L-2}/(1 - UA^{-L})$ we will be left with a remainder having factors $(1 - UA^{L-2})(1 - UA^{-L})$ in the denominator. This remainder can be split by the method described above into one term whose expansion contains only non-negative powers of A and one with negative powers only. The latter is discarded while the former is the only term contributing to the q_i. The result of this procedure is again of the form (4.6).

The form (4.6) is not yet the proper one for a generating function. First, it may contain some negative coefficients in the numerators. Second, the denominator factors $(1 - UA^n)$ for $n < l$ do not correspond to elementary tensors. Finally, it may be possible to cancel some common factors between numerator and denominator. Adjustments are easily made which lead to the form of Sec. 2.

5. DISCUSSION

Sometimes one needs to enumerate and construct tensors which are functions of the components of two or more

TABLE VII. B_{ij}^{Lk} for $L = 12$.

	k=2		k=3			k=4				k=5					
i \ j	12	14	12	14	16	12	14	16	18	0	2	4	6	8	10
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0
3	1	0	0	1	0	1	0	1	0	0	0	1	1	2	1
4	1	0	1	1	1	2	2	1	1	1	0	3	2	5	3
5	1	1	2	3	2	4	3	4	3	1	2	5	6	8	9
6	1	2	5	4	5	8	7	7	8	3	4	10	11	17	17
7	2	2	8	8	7	15	14	15	14	4	10	17	23	29	35
8	3	3	14	12	13	25	26	26	26	7	16	31	38	53	61
9	4	4	20	20	18	44	43	44	44	9	30	48	68	85	104
10	5	5	30	27	27	67	70	68	69	17	45	78	105	137	164
11	5	6	38	37	35	102	104	103	103	21	73	115	163	205	254
12	6	6	51	46	47	147	149	147	150	36	102	170	236	303	368
13	6	5	60	60	55	202	206	204	204	45	148	235	338	423	524
14	6	5	72	68	67	268	272	270	270	65	196	326	455	583	711
15	5	5	78	79	74	346	346	344	344	81	264	424	608	766	946
16	4	4	86	83	84	425	428	423	425	110	332	551	773	992	1209
17	3	3	86	89	85	509	507	508	506	131	419	683	971	1233	1519
18	2	2	88	86	88	590	588	586	587	168	501	838	1174	1509	1839
19	1	1	82	85	83	663	660	658	656	193	601	987	1398	1784	2190
20	0	1	77	77	80	723	717	719	715	232	686	1153	1607	2074	2529
21	0	1	66	70	70	766	758	760	759	256	783	1295	1824	2338	2866
22			58	57	62	786	781	782	779	293	854	1441	2005	2592	3158
23			45	48	49	785	780	783	779	307	931	1549	2171	2792	3421
24			36	36	40	764	757	760	759	336	974	1645	2284	2959	3603
25			25	28	28	718	717	717	715	339	1015	1691	2368	3048	3733
26			19	18	21	658	657	659	656	351	1017	1721	2383	3091	3765
27			11	13	13	585	585	585	587	339	1015	1691	2368	3048	3733
28			7	7	9	503	506	506	506	336	974	1645	2284	2959	3603
29			3	5	4	419	423	424	425	307	931	1549	2171	2792	3421
30			2	2	2	340	343	342	344	293	854	1441	2005	2592	3158
31			0	1	1	264	270	266	270	256	783	1295	1824	2338	2866
32			0	0	1	199	201	202	204	232	686	1153	1607	2074	2529
33						143	147	145	150	193	601	987	1398	1784	2190
34						99	102	102	103	168	501	838	1174	1509	1839
35						65	67	68	69	131	419	683	971	1233	1519
36						42	41	42	44	110	332	551	773	992	1209
37						24	25	25	26	81	264	424	608	766	946
38						13	13	15	14	65	196	326	455	583	711
39						7	7	7	8	45	148	235	338	423	524
40						3	3	4	3	36	102	170	236	303	368
41						1	1	2	1	21	73	115	163	205	254
42						1	0	1	0	17	45	78	105	137	164
43										9	30	48	68	85	104
44										7	16	31	38	53	61
45										4	10	17	23	29	35
46										3	4	10	11	17	17
47										1	2	5	6	8	9
48										1	0	3	2	5	3
49										0	0	1	1	2	1
50										0	0	1	0	1	0
51										0	0	0	0	0	0
52										1	0	0	0	0	0

irreducible tensors. Sylvester and Franklin[19] gave generating functions based on two irreducible tensors of ranks L_1, L_2 with $L_1 + L_2 \leqslant 10$. They may now be constructed straightforwardly, in a "positive" form, by combining the generating functions for L_1 and L_2 with the help of the SU(2) Clebsch–Gordan generating function[20]:

$$F_{L_1L_2}(U_1,U_2,A) = P_{A_1A_2}\,F_{L_1}(U_1,A_1)\,F_{L_2}(U_2,A_2) \times[(1-A_1^{-1}A_2^{-1}) \times(1-A_1^{-1}A)(A_2^{-1}A)]^{-1}. \tag{5.1}$$

U_1 and U_2 carry as exponents the degrees in the L_1 and L_2 tensors, respectively, whiel A carries the rank of the polynomial tensor. Further base tensors may be introduced consecutively by iteration of the procedure. Such states are needed for interacting boson nuclear models, with bosons of different L. To construct Sp$(2n)\supset$SU(2) labeling operators, for example, one needs polynomial tensors based on $L = 2,6,\ldots,4n-2$.

As mentioned in Sec. 1, the generators of G_2 decompose under its maximal O(3) subgroup into tensors of rank 10 and 2. Hence, the generating function $F_{10}(U,A)$ enumerates and defines an integrity basis for $G_2\supset$O(3) missing label operators. The denominator factors $[(1-U^2)(1-U^6)]^{-1}$ correspond to the G_2 Casimir invariants and should be ignored. A term $c_{ua}\,U^u A^a$ in the expansion of $(1-U^2)(1-U^6)$ $\times F_{10}(U,A)$ implies the existence of just c_{ua} linearly independent labeling operators of degree u in the rank 10 tensor and degree $\frac{1}{2}a$ in the rank 2 tensor. The fact that there are eight denominator factors accords with the known fact that there are twice as many functionally independent missing label operators as there are missing labels (four). Given their degrees it is straightforward to construct these operators. Finding four which mutually commute is, however, an unsolved problem.

Two directions for extending the work come to mind. One is the computation of polynomial (symmetric) tensors for higher groups. The second is the calculation of tensors with exchange symmetries corresponding to all representations of the permutation group; in principle one could find a single generating function for tensors of all degrees and symmetries based on a given tensor.

A computer program has now been developed which can assist in the extension of the results of this paper to arbitrarily high L.[21]

ACKNOWLEDGMENTS

We thank R. P. Stanley for helpful discussions.

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and by the Ministère de l'Éducation du Québec.

[1] A. Cayley, Am. J. Math. **2**, 71 (1879).

[2] J. J. Sylvester and F. Franklin, Am. J. Math. **2**, 223 (1879).

[3] F. Franklin, Am. J. Math. **3**, 128 (1880).

[4] J. J. Sylvester, Am. J. Math. **4**, 41 (1881).

[5] An equivalent formulation of the problem of counting L-harmonics of degree u is to require the decomposition of the symmetric part S_L^u of the direct product $R_L \otimes R_L \otimes \cdots \otimes R_L$ of u copies of the irreducible representation R_L into a direct sum of irreducible representations $S_L^u = \oplus m_{ua}^L R_a$. The reduction is called the computation of a symmetric plethysm.

[6] F. D. Murnaghan, Proc. Natl. Acad. Sci. (USA) **37**, 439 (1951).

[7] M. A. Lohe and C. A. Hurst, J. Math. Phys. **12**, 1882 (1971).

[8] R. Gaskell, A. Peccia, and R. T. Sharp, J. Math. Phys. **19**, 727 (1978).

[9] E. Chacón, M. Moshinsky, and R. T. Sharp, J. Math. Phys. **17**, 668 (1976).

[10] G. vanden Berghe and H. D. de Meyer, Nucl. Phys. A **323**, 302 (1979).

[11] S. Rohoziński, J. Phys. G **4**, 1075 (1978).

[12] S. Rohoziński and W. Greiner, J. Phys. G **6**, 969 (1980).

[13] F. Busse, J. Fluid Mech. **72**, 67 (1975).

[14] P. Chossat, SIAM J. Appl. Math. **37**, 21 (1979).

[15] D. H. Sattinger, J. Math. Phys. **19**, 1720 (1978).

[16] M. Golubitsky and D. Schaeffer, "Bifurcations with O(3) symmetry including applications to the Benard problem," Duke University preprint (1980).

[17] Tables of B_{ij}^{Lk} for $k > 1$ and $L = 11$ and 13 have not been included here due to their length. Copies may be obtained from R. Gaskell.

[18] R. W. Gaskell, "Polynomials in the components of a single SU(2) tensor—A new recursion technique," Proceedings of the Tenth International Colloquium on Group-Theoretical Methods in Physics, Physica (to appear).

[19] J. J. Sylvester and F. Franklin, Am. J. Math. **2**, 293 (1879).

[20] J. Patera and R. T. Sharp, *Lecture Notes in Physics*, Vol. 94 (Springer, New York, 1979).

[21] R. Gaskell, Comput. Phys. Commun. **24**, 191 (1981).

Bäcklund generated solutions of Liouville's equation and their graphical representations in three spatial dimensions

George Leibbrandt
Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada
Shein-shion Wang
Institute of Computer Science, University of Guelph, Guelph, Ontario, Canada and Department of Geophysics, Colorado School of Mines, Golden,[a] Colorado, 80401
Nader Zamani
Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada and Research Department, Dominion Engineering Works Ltd., Lachine, Quebec,[a] Canada

(Received 15 December 1981; accepted for publication 23 April 1982)

The Bäcklund–Bianchi method is employed to generate, in three spatial dimensions, the following multiple solutions of Liouville's equation $\nabla^2\alpha = \exp\alpha$: The three-wave interaction function α_3 and the five-wave interaction function α_5. It is verified numerically that α_3 satisfies Liouville's equation to an accuracy of one part in 10^{14}, while α_5 satisfies it to one part in 10^6. The construction of α_5 is conditional upon solving ten nonlinear constraint equations. We analyze the complicated structures of α_3 and α_5 with the help of a three-dimensional plotting routine. It is found that α_3 is, surprisingly enough, only characterized by a single ring singularity, while α_5 exhibits three ring singularities. It is speculated that the function tanh α_3 represents a ring soliton whose shape appears to be preserved in the nonlinear superposition of similar ring solitons. The derivation of Liouville's solutions α_3 and α_5 is intimately connected with the auxiliary functions β_2 and β_4 which solve Laplace's equation. The latter are also derived and plotted in the paper.

PACS numbers: 02.30.Jr

I. INTRODUCTION

The powerful method of Bäcklund transformations dates back over a hundred years to the pioneering work of Lie, Bianchi and Bäcklund, who investigated the transformation properties of certain surfaces, especially pseudospherical surfaces, in a series of fundamental papers between 1873 and 1883.[1] The reason for the revival of and sustained interest in Bäcklund transformations during the past two decades is fairly obvious: Judicious application of Bäcklund's theory permits, in many instances, the solution of physically relevant second-order nonlinear partial differential equations, among which the nonlinear Schrödinger and Korteweg–de Vries equations and the sine–Gordon system have been analyzed with particular zest and thoroughness.

Another nonlinear equation which is soluble by the Bäcklund technique and which has become prominent of late is Liouville's equation

$$(\partial_x{}^2 + \partial_y{}^2)\chi = k\exp(a\chi), \quad \partial_x \equiv \partial/\partial x, \text{ etc.,} \tag{1}$$

with suitable boundary conditions on χ, where χ is a scalar field and a,k are real constants. It was initially solved by Liouville[2] in 1853 in two spatial dimensions and has since been studied by several well-known mathematicians, including Picard, Poincaré, and Bierberbach.[3] Liouville's equation is known to possess significant applications in electrostatics,[4] hydrodynamics,[5–7] and cosmology.[8] In recent years, Eq. (1) has also attracted the attention of particle physicists in connection with monopole theories.[9]

Liouville's equation has been analyzed not only in 1 + 1 and 2 + 0 dimensions,[10] but also in 3 + 0 dimensions.[11] One of the present authors employed a Bäcklund-like transformation to generate an exact three-wave interaction solution of Liouville's equation in three spatial dimensions,

$$\nabla^2\alpha = \exp\alpha, \quad \nabla^2 \equiv \partial_x{}^2 + \partial_y{}^2 + \partial_z{}^2, \tag{2}$$

satisfying the boundary conditions $\alpha \to -\infty$ and $d\alpha/dr \to 0^-$, for $r \equiv (x^2+y^2+z^2)^{1/2} \to +\infty$. A plot of this three-wave solution, labeled α_3 in Ref. 11, is given here for the first time and will be discussed in Sec. 3.

The purpose of the present article is two-fold. (i) First, we wish to verify numerically that the Bäcklund-like transformation in 3 + 0 dimensions, originally derived in Ref. 11, is sufficiently powerful to generate multiple solutions of (2) beyond α_3, namely up to and including α_5. This is no easy task, since the construction of the five-wave interaction function $\alpha_5(x,y,z)$ depends decisively on solving numerically ten algebraic and transcendental constraint equations to a high degree of accuracy. (ii) Our second aim is to present three-dimensional graphical representations of both α_3 and α_5. The result of this plotting exercise was somewhat surprising; the global, i.e., asymptotic, features of α_3 and α_5 turned out to be quite different from those normally expected from the superposition of "single" solutions, as in the sine–Gordon theory, for example. In the latter case we know[12] that superposing two (four) single solitons in a nonlinear fashion does not alter asymptotically the shape or direction of the individual solitons. This is no longer true for Liouville's multiple solutions; for instance, when the three lowest-order solutions $\alpha_1{}^{(i)}$,1,2,3, are combined by the usual Bäcklund–Bianchi technique, the final structure of the resulting α_3 solution has no resemblance to any one of the original α_1's, even from a global perspective. A similar situation exists for

[a] Present address.

0022-2488/82/091566-07$02.50

α_5, implying that Liouville's solutions do not represent solitons of the conventional type.

The outline of our paper is as follows: In Sec. 2 we review, for the sake of completeness and in order to establish the nomenclature, the Bäcklund transformation method for Liouville's system. In Sec. 3 we analyze and plot the three-wave interaction function $\alpha_3(x,y,z)$. Section 4 is devoted to the auxiliary function $\beta_4(x,y,z)$ which satisfies Laplace's equation $\nabla^2\beta_4 = 0$ in a certain domain $\mathscr{D}$, and is essential for the construction of the next-highest Liouville solution $\alpha_5(x,y,z)$. The latter function, superposed from five "single" solutions $\alpha_1{}^{(i)}, i = 1,2,...5$, and subject to ten constraint equations, is examined in considerable detail in Secs. 5 and 6, where we also give its graphical representation. The article concludes with a summary and discussion.

2. REVIEW AND NOTATION

The "Bäcklund transformation" for Liouville's equation (2) in three spatial dimensions was shown to be of the form[11]

$$K(i\beta - \alpha) = \sqrt{2}\exp((\alpha + i\beta)/2)\exp i\theta\mathscr{S},$$
$$K \equiv I\partial_x + i(\sigma_1\partial_y + \sigma_3\partial_z), \mathscr{S} = \sigma_1\exp(-i\lambda\sigma_2), \tag{3}$$

where α and β satisfy Liouville's and Laplace's equations, respectively,

$$\nabla^2\alpha = \exp\ \alpha, \tag{4}$$
$$\nabla^2\beta = 0; \tag{5}$$

$\sigma_1,\sigma_2,\sigma_3$ are the Pauli matrices, I is the unit matrix and θ,λ $(0\leqslant\theta\leqslant 2\pi,\ 0\leqslant\lambda\leqslant 2\pi)$ are the Bäcklund transformation parameters. As shown in Ref. 11, Eq. (3) is equivalent to eight real scalar equations which, in turn, are subject to six integrability conditions [cf. Eqs. (5) and (6) of Ref. 11]. We may replace system (3) by two *real* matrix equations

$$I\partial_x\alpha + \mathscr{P}\beta$$
$$= \sqrt{2}[\mathscr{S}\sin\theta\sin(\beta/2) - I\cos\theta\cos(\beta/2)]\exp(\alpha/2), \tag{6a}$$

$$I\partial_x\beta - \mathscr{P}\alpha$$
$$= \sqrt{2}[\mathscr{S}\sin\theta\cos(\beta/2) + I\cos\theta\sin(\beta/2)]\exp(\alpha/2), \tag{6b}$$

where $\mathscr{P} \equiv \sigma_1\partial_y + \sigma_3\partial_z$.

Let us summarize the results for α_1,β_2, and α_3.[11] Setting $\beta \equiv \beta_0 = 0$ in Eq. (6a) we get the simplest nontrivial Liouville solution

$$\alpha_1{}^{(i)}(x,y,z) = \ln(2/T_i{}^2), \quad i = 1,2,...,5,$$
$$T_i(x,y,z) = x\cos\theta_i + \sin\theta_i(y\cos\lambda_i + z\sin\lambda_i) + b, \tag{7}$$

where b is a constant of integration (we took $b = +40$), possessing the dimension of a length, while the index i in (7) labels the five different α_1 solutions which are needed in the construction of α_5. See Fig. 1. Observe that formula (7) is equivalent to

$$\tanh(\alpha_1{}^{(i)}/4) = (\surd 2 - T_i)/(\surd 2 + T_i), i = 1,2,...,5, \tag{8}$$

provided $T_i > 0$. Since $\tanh(\alpha_1{}^{(i)}/4) \to 1^-$ as $T_i \to 0^+$, we see that $\alpha_1{}^{(i)}$ develops a *line singularity* for $T_i \sim 0$, as indicated by Fig. 2.

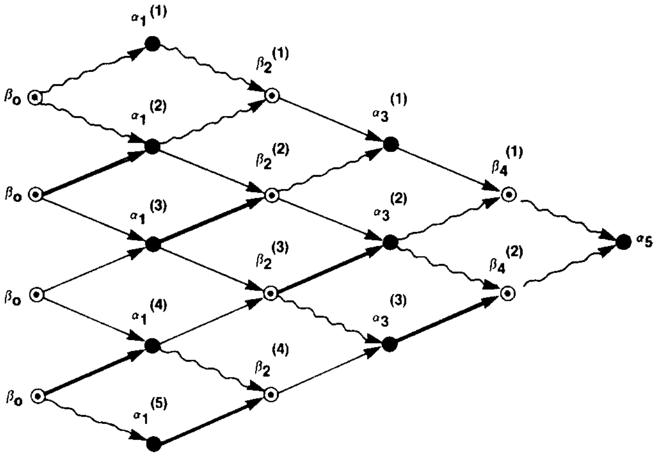


FIG. 1. Extended Bianchi diagram showing the generation of α_N solutions, $N = 1,3,5$, of Liouville's equation (4) and of β_j solutions, $j = 0,2,4$ of Laplace's equation (5).

According to Fig. 1, the next highest Liouville solution is α_3,

$$\tanh((\alpha_3{}^{(1)} - \alpha_1{}^{(2)})/4) = R_{13}\tan((\beta_2{}^{(1)} - \beta_2{}^{(2)})/4),$$
$$R_{13} = -[(1 + \mathscr{L}_{13})/(1 - \mathscr{L}_{13})]^{1/2}, \quad |\mathscr{L}_{13}| < 1, \tag{9}$$

which is subject to the constraint equation

$$1 + 2\mathscr{L}_{12}\mathscr{L}_{23}\mathscr{L}_{31} = (\mathscr{L}_{12})^2 + (\mathscr{L}_{23})^2 + (\mathscr{L}_{31})^2,$$

with

$$\mathscr{L}_{pq} \equiv \cos\theta_p\cos\theta_q + \sin\theta_p\sin\theta_q\cos(\lambda_p - \lambda_q),$$
$$p,q \text{ integer}, \tag{10a}$$

and

$$\nabla^2\alpha_3 = \exp\alpha_3, \quad \text{domain } \mathscr{D}_3. \tag{10b}$$

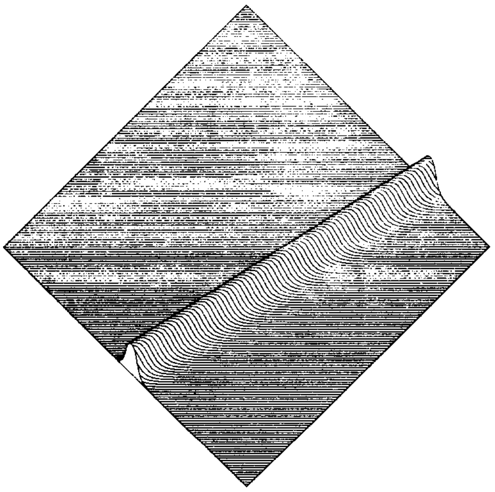


FIG. 2. The function $\tanh(\alpha_1{}^{(2)}/4)$ in the domain $\mathscr{D}_1 = \{(x,y)| -45<x<15, -55<y<5\}$. The line singularity is clearly evident. Here $\phi = 80°$ and $\gamma = 135°$.

The *auxiliary* function $\beta_2^{(1)}$ satisfies Laplace's equation

$$\nabla^2\beta_2^{(1)} = 0, \quad \text{domain } \mathscr{D}_2, \tag{11}$$

and is given by

$$\tan((\beta_2^{(1)} - \beta_0)/4) = R_{12} \tanh((\alpha_1^{(1)} - \alpha_1^{(2)})/4),$$
$$R_{12} = +((1 + \mathscr{L}_{12})/(1 - \mathscr{L}_{12}))^{1/2}, \quad |\mathscr{L}_{12}| < 1. \tag{12}$$

Figure 3 depicts a three-dimensional graph of $\sin\beta_2^{(1)}$.

3. ANALYSIS OF α_3

To study the singularity structure of $\alpha_3^{(1)}$ it is convenient to express Eq. (9) as

$$\tanh((\alpha_3^{(1)} - \alpha_1^{(2)})/4) = R_{13}N(x,y,z)/D(x,y,z),$$
$$|R_{13}N/D| \leqslant 1, \tag{13}$$

where

$$N \equiv +[R_{12}(T_3 + T_2)(T_2 - T_1) - R_{23}(T_3 - T_2) \times (T_2 + T_1)](T_1T_3T_2^2)^{-1}, \tag{14a}$$
$$D \equiv [(T_3 + T_2)(T_2 + T_1) + R_{12}R_{23}(T_3 - T_2) \times (T_2 - T_1)](T_1T_3T_2^2)^{-1}, \tag{14b}$$

and to observe that for fixed z, both $N(x,y,z)$ and $D(x,y,z)$ are *quadratic* functions of x,y. Eq. (13) implies that

$$\exp((\alpha_3^{(1)} - \alpha_1^{(2)})/2) = [1 + R_{13}N/D] \times [1 - R_{13}N/D]^{-1}, \tag{15}$$

provided $T_2 \geqslant 0$.

To help us identify the two types of singularities in system (13)–(14), we have plotted $\tanh((\alpha_3^{(1)} - \alpha_1^{(2)})/4)$, as shown in Fig. 4, and compared it with $\tanh\alpha_3^{(1)}$ in Fig. 5. The former figure suggests the presence of two singularities: a *ring-like* singularity associated with $\alpha_3^{(1)}$, and a *line* singularity which is connected with $\alpha_1^{(2)}$ and emerges in the limit as $T_2 \to 0^+$. The function $\tanh\alpha_3^{(1)}$, on the other hand, is even more amazing; its ring-shaped structure implies that $\alpha_3^{(1)}$ contains only *one* singularity, the ring singularity depicted in Fig. 5. These results also follow from algebraic considerations. What appears to have happened in this: the three line singularities of $\alpha_1^{(1)}$,$\alpha_1^{(2)}$ and $\alpha_1^{(3)}$ seem to have been transformed by nonlinear superposition into one single ring-shaped singularity which bears no resemblance to the original α_1's.

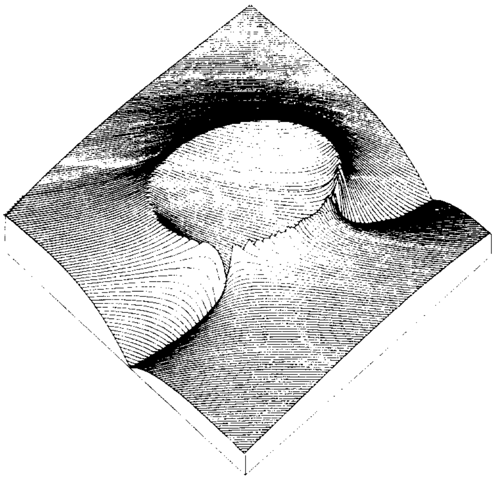


FIG. 4. The function $\alpha_3^{(1)}$ in the form $\tanh((\alpha_3 - \alpha_1^{(2)})/4)$, the domain being $\mathscr{D}_1$ (as in Fig. 2); $\phi = 65°$ and $\gamma = 135°$.

As mentioned in Ref. 12, the function $\alpha_3^{(1)}$ is a coplanar solution of $\nabla^2\alpha_3^{(1)} = \exp\alpha_3^{(1)}$ which has been solved subject to the constraint

$$1 + 2\mathscr{L}_{12}\mathscr{L}_{23}\mathscr{L}_{31} = (\mathscr{L}_{12})^2 + (\mathscr{L}_{23})^2 + (\mathscr{L}_{31})^2. \tag{16}$$

We have verified numerically that for points away from the ring singularity, $\alpha_3^{(1)}$ satisfies Liouville's equation (10) to one part in 10^{14} for the parameters $(\theta_1,\theta_2,\theta_3)$ and $(\lambda_1,\lambda_2,\lambda_3)$ listed in Eq. (27). The constraint equation (16) for the $\mathscr{L}_{ij}$'s is satisfied to an accuracy of one part in 10^{20}. All numerical work has been carried out in quadruple precision.

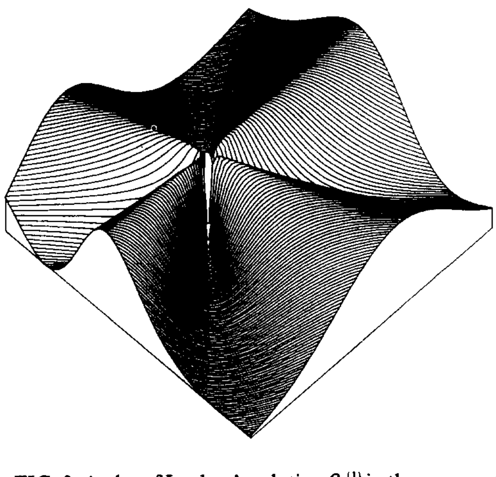


FIG. 3. A plot of Laplace's solution $\beta_2^{(1)}$ in the representation $\sin\beta_2^{(1)}$. The domain is $\mathscr{D}_2 = \{(x,y)| -50 < x < 10, -65 < y < -5\}$, with $\phi = 60°, \gamma = 135°$.

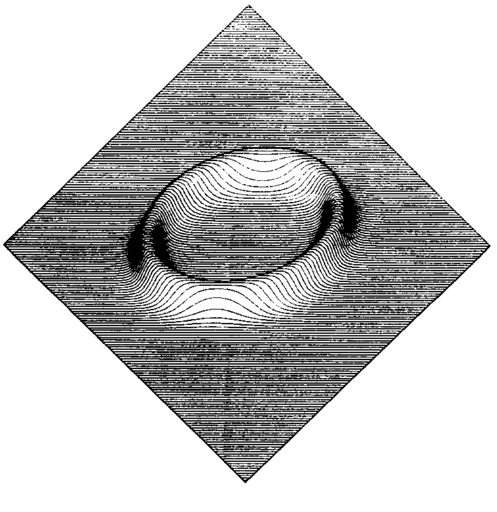


FIG. 5. A plot of the function $\tanh\alpha_3^{(1)}$ in the domain $\mathscr{D}_1$, for $\phi = 80°$ and $\gamma = 135°$.

4. THE AUXILIARY SOLUTION $\beta_4(x,y,z;\theta_1,...,\theta_4;\lambda_1,...,\lambda_4)$

As seen from Fig. 1, the construction of α_5 requires two β_4 functions which must be solutions of Laplace's equation

$$\nabla^2\beta_4^{(i)} = 0, \quad i = 1,2, \quad \text{domain } \mathscr{D}_4. \tag{17a}$$

It can be shown that $\beta_4^{(i)}$ may be represented by

$$\tan((\beta_4^{(i)} - \beta_2^{(i+1)})/4) = R_{i,i+1} \tanh((\alpha_3^{(i)} - \alpha_3^{(i+1)})/4), \tag{18}$$

where

$$\nabla^2\beta_2^{(i+1)} = 0, \tag{17b}$$

while

$$\nabla^2\alpha_3^{(\sigma)} = \exp \alpha_3^{(\sigma)}, \quad \sigma = 1,2,3. \tag{19}$$

the coefficient $R_{i,i+1}$ is defined by

$$R_{pq} = (-1)^{(1+p+q)}|R_{pq}|, \quad p,q \text{ integer}, \quad p \neq q, \tag{20a}$$

$$|R_{pq}| = ((1 + \mathscr{L}_{pq})/(1 - \mathscr{L}_{pq}))^{1/2}, \quad |\mathscr{L}_{pq}| < 1,$$

with

$$\mathscr{L}_{pq} = \cos\theta_p \cos\theta_q + \sin\theta_p \sin\theta_q \cos(\lambda_p - \lambda_q). \tag{20b}$$

The Bäcklund parameters θ_p,λ_p are confined to the domains $0 \leqslant \theta_p \leqslant 2\pi$, $0 \leqslant \lambda_p \leqslant 2\pi$. For each $\alpha_3^{(j)}, j = 1,2,3$[cf. Eq. (16)], there is only *one* constraint equation, namely

$$\begin{aligned} 1 + 2&\mathscr{L}_{j,j+1}\mathscr{L}_{j+1,j+2}\mathscr{L}_{j+2,j} \\ &= (\mathscr{L}_{j,j+1})^2 + (\mathscr{L}_{j+1,j+2})^2 + (\mathscr{L}_{j+2,j})^2, \end{aligned} \tag{21}$$

while for each $\beta_4^{(i)}, i = 1,2$, we have these *four* constraints:

$$\begin{aligned} 1 + 2&\mathscr{L}_{i,i+1}\mathscr{L}_{i+1,i+2}\mathscr{L}_{i+2,i} \\ &= (\mathscr{L}_{i,i+1})^2 + (\mathscr{L}_{i+1,i+2})^2 + \mathscr{L}_{i+2,i})^2, \\ 1 + 2&\mathscr{L}_{i,i+1}\mathscr{L}_{i+1,i+3}\mathscr{L}_{i+3,i} \\ &= (\mathscr{L}_{i,i+1})^2 + (\mathscr{L}_{i+1,i+3})^2 + (\mathscr{L}_{i+3,i})^2, \\ 1 + 2&\mathscr{L}_{i+1,i+2}\mathscr{L}_{i+2,i+3}\mathscr{L}_{i+3,i+1} \\ &= (\mathscr{L}_{i+1,i+2})^2 + (\mathscr{L}_{i+2,i+3})^2 + (\mathscr{L}_{i+3,i+1})^2, \\ 1 + 2&\mathscr{L}_{i,i+2}\mathscr{L}_{i+2,i+3}\mathscr{L}_{i+3,i} \\ &= (\mathscr{L}_{i,i+2})^2 + \mathscr{L}_{i+2,i+3})^2 + (\mathscr{L}_{i+3,i})^2. \end{aligned} \tag{22}$$

The function $\beta_4(x,y,z = -20)$, which depends on a total of eight Bäcklund parameters, has been checked to an accuracy of one part in 10^{10}. We note that system (22) is characterized for each $i = 1,2$, by six distinct $\mathscr{L}$'s which are functions of the eight Bäcklund parameters: $(\theta_1,...,\theta_4;\lambda_1,...,\lambda_4)$ for $i = 1$, and $(\theta_2,...,\theta_5;\lambda_2,...,\lambda_5)$ for $i = 2$. Figure 6 shows a plot of the four-wave interaction $\beta_4^{(1)}$.

5. THE INTERACTION FUNCTION $\alpha_5(x,y,z)$

(a) It is easily demonstrated with the aid of Fig. 1 that superposition of the single Liouville solutions $\alpha_1^{(i)}, i = 1,2,3,4,5$ in the spirit of Bäcklund-Bianchi, leads to the five-wave interaction function $\alpha_5(x,y,z)$

$$\begin{aligned} &\tanh((\alpha_5 - \alpha_3^{(2)})/4) = R_{15} \tan((\beta_4^{(1)} - \beta_4^{(2)})/4), \\ &R_{15} = -((1 + \mathscr{L}_{15})/(1 - \mathscr{L}_{15}))^{1/2}, \quad |\mathscr{L}_{15}| < 1, \end{aligned} \tag{23}$$

with $\mathscr{L}_{15}$ defined bys Eq. (20b) and

$$\nabla^2\alpha_5 = \exp\alpha_5. \tag{24}$$

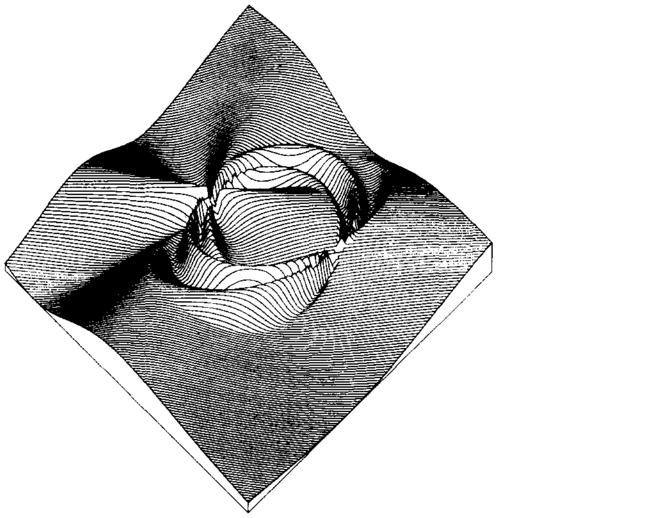


FIG. 6. The function $\beta_4^{(1)}$ in the form $\sin\beta_4^{(1)}$ for $\phi = 80°$ and $\gamma = 135°$ in the domain $\mathscr{D}_1$.

The function $\alpha_5(x,y,z;\theta_1,...,\theta_5;\lambda_1,...,\lambda_5)$ will solve Eq. (24) provided the following constraint system is solved consistently for the ten Bäcklund parameters $\theta_i,\lambda_i, i = 1,2,3,4,5$:

$$\begin{aligned} 1 + 2\mathscr{L}_{12}\mathscr{L}_{23}\mathscr{L}_{31} &= \mathscr{L}_{12}^2 + \mathscr{L}_{23}^2 + \mathscr{L}_{31}^2, \\ 1 + 2\mathscr{L}_{12}\mathscr{L}_{24}\mathscr{L}_{41} &= \mathscr{L}_{12}^2 + \mathscr{L}_{24}^2 + \mathscr{L}_{41}^2, \\ 1 + 2\mathscr{L}_{23}\mathscr{L}_{34}\mathscr{L}_{42} &= \mathscr{L}_{23}^2 + \mathscr{L}_{34}^2 + \mathscr{L}_{42}^2, \\ 1 + 2\mathscr{L}_{13}\mathscr{L}_{34}\mathscr{L}_{41} &= \mathscr{L}_{13}^2 + \mathscr{L}_{34}^2 + \mathscr{L}_{41}^2, \\ 1 + 2\mathscr{L}_{23}\mathscr{L}_{35}\mathscr{L}_{52} &= \mathscr{L}_{23}^2 + \mathscr{L}_{35}^2 + \mathscr{L}_{52}^2, \\ 1 + 2\mathscr{L}_{34}\mathscr{L}_{45}\mathscr{L}_{53} &= \mathscr{L}_{34}^2 + \mathscr{L}_{45}^2 + \mathscr{L}_{53}^2, \\ 1 + 2\mathscr{L}_{24}\mathscr{L}_{45}\mathscr{L}_{52} &= \mathscr{L}_{24}^2 + \mathscr{L}_{45}^2 + \mathscr{L}_{52}^2, \\ 1 + 2\mathscr{L}_{12}\mathscr{L}_{25}\mathscr{L}_{51} &= \mathscr{L}_{12}^2 + \mathscr{L}_{25}^2 + \mathscr{L}_{51}^2, \\ 1 + 2\mathscr{L}_{13}\mathscr{L}_{35}\mathscr{L}_{51} &= \mathscr{L}_{13}^2 + \mathscr{L}_{35}^2 + \mathscr{L}_{51}^2, \\ 1 + 2\mathscr{L}_{14}\mathscr{L}_{45}\mathscr{L}_{51} &= \mathscr{L}_{14}^2 + \mathscr{L}_{45}^2 + \mathscr{L}_{51}^2. \end{aligned} \tag{25}$$

As mentioned previously,[12] the effect of these constraint equations is to force α_3,β_4, and α_5 to become coplanar.

System (25) consists of ten equations in precisely ten unknowns $\mathscr{L}_{12},...,\mathscr{L}_{45}$; it was solved to an accuracy of one part in 10^{20} to yield:

$$\begin{aligned} \mathscr{L}_{12} &= 0.956\ 786\ 286\ 485\ 616\ 000\ 00, \\ \mathscr{L}_{13} &= 0.721\ 114\ 935\ 626\ 545\ 100\ 00, \\ \mathscr{L}_{14} &= 0.593\ 187\ 193\ 522\ 227\ 000\ 00, \\ \mathscr{L}_{15} &= 0.568\ 140\ 304\ 086\ 613\ 400\ 00, \\ \mathscr{L}_{23} &= 0.891\ 418\ 089\ 642\ 861\ 600\ 00, \\ \mathscr{L}_{24} &= 0.801\ 659\ 727\ 311\ 817\ 700\ 00, \\ \mathscr{L}_{25} &= 0.782\ 890\ 829\ 596\ 234\ 700\ 00, \\ \mathscr{L}_{34} &= 0.985\ 517\ 314\ 624\ 865\ 400\ 00, \\ \mathscr{L}_{35} &= 0.979\ 834\ 261\ 020\ 515\ 000\ 00, \\ \mathscr{L}_{45} &= 0.999\ 526\ 709\ 229\ 567\ 800\ 00. \end{aligned} \tag{26}$$

The constraint equations (25) for the $\mathscr{L}_{ij}$'s were solved, using quadruple precision, to give the following values to an accuracy of one part in 10^{20} (the λ 's are in radians, the θ 's in degrees):

$$\begin{aligned}
\lambda_1 &= 0.66\ 391\ 602\ 978\ 664\ 308\ 98,\\
\lambda_2 &= 0.747\ 431\ 140\ 737\ 200\ 265\ 76,\\
\lambda_3 &= 0.606\ 370\ 474\ 119\ 483\ 888\ 08,\\
\lambda_4 &= 0.287\ 144\ 414\ 680\ 506\ 957\ 00,\\
\lambda_5 &= 0.725\ 240\ 492\ 717\ 253\ 436\ 05,\\
\theta_1 &= 65.740\ 776\ 764\ 056\ 272\ 193\ 29^\circ, \qquad (27)\\
\theta_2 &= 105.949\ 931\ 184\ 388\ 188\ 536\ 08^\circ,\\
\theta_3 &= 97.561\ 341\ 430\ 862\ 637\ 974\ 37^\circ,\\
\theta_4 &= 77.670\ 945\ 869\ 960\ 531\ 352\ 14^\circ,\\
\theta_5 &= 104.683\ 428\ 006\ 073\ 160\ 110\ 62^\circ.
\end{aligned}$$

As in the sine–Gordon case,[12] it is essential that the λ 's and θ 's be arranged in ascending order when the accuracy of α_5, as a solution of Eq. (24), is being checked. Substituting the values for θ_i and λ_i from (27) into the expression for R_{pq} in Eq. (20), we find that

$$\begin{aligned}
R_{12} &= +\ 6.729\ 160\ 761\ 177\ 823\ 000\ 00,\\
R_{13} &= -\ 2.484\ 232\ 896\ 793\ 022\ 000\ 00,\\
R_{14} &= +\ 1.978\ 955\ 804\ 601\ 952\ 000\ 00,\\
R_{15} &= -\ 1.905\ 553\ 389\ 405\ 890\ 000\ 00,\\
R_{23} &= +\ 4.173\ 640\ 423\ 411\ 595\ 000\ 00, \qquad (28)\\
R_{24} &= -\ 3.013\ 914\ 531\ 328\ 807\ 000\ 00,\\
R_{25} &= +\ 2.865\ 650\ 959\ 983\ 701\ 000\ 00,\\
R_{34} &= +\ 11.708\ 797\ 397\ 936\ 890\ 000\ 00,\\
R_{35} &= -\ 9.908\ 487\ 069\ 842\ 504\ 000\ 00,\\
R_{45} &= +\ 64.997\ 939\ 177\ 800\ 580\ 000\ 00.
\end{aligned}$$

Before examining the α_5 function in greater detail, we ought to mention that all our plots were obtained by fixing the third spatial component at $z = -20$, and allowing the other two coordinates x,y to vary over the indicated domains. Thus tanh $\alpha_1^{(2)},\ldots,$tanh α_5 are all plotted in the XY plane. A change in z to $z = 10$, for example, merely causes a *translation* of the original plot—it does not lead to a change in structure. Moreover, our decision to plot tanh $\alpha_3^{(1)}$ rather than $\alpha_3^{(1)}$,sin$\beta_4^{(1)}$ rather than $\beta_4^{(1)}$, and so forth, was based strictly on convenience.

(*b*) *Numerical calculations*: One of the major tasks of generating the α and β solutions is an accurate determination of the Bäcklund parameters. In order to do so, one has to calculate the $\mathscr{L}_{ij}$'s from Eq. (25) and use them in Eq. (10a) to obtain the parameters θ_i and λ_i. In both steps we are faced with solving a system of ten nonlinear equations for ten unknowns. Let us briefly outline the method which was used to obtain a solution to the above systems.

For the sake of brevity we assume that the system of nonlinear equations is given by

$$f_i(x_1,x_2,\ldots,x_{10}) = 0, \quad i = 1,2,\ldots,10. \qquad (29)$$

This is equivalent to finding a vector $(x_1,x_2,\ldots,x_{10})$ which minimizes the functional J, given by

$$J(x_1x_2,\ldots,x_{10}) = \sum_{i=1}^{10} [f_i(x_1,x_2,\ldots,x_{10})]^2. \qquad (30)$$

A straightforward procedure to find a minimizer is to use the steepest descent algorithm, which is an iterative technique represented by

$$X_{n+1} = X_n - \rho_n \nabla J(X_n) \quad n = 1,2,\cdots, \qquad (31)$$

$$X_0 = (x_1^0,x_2^0,\ldots,x_{10}^0) \quad \text{specified.} \qquad (32)$$

Here ρ_n is a measure of the step size; in our calculations it was taken to be a positive, decreasing function of n. The choice of X_0 is crucial for convergence of the algorithm. Our selection of X_0 was based on our knowledge of the eight Bäcklund parameters required to generate β_4.[12]

Once the ten parameters (θ_i,λ_i) have been calculated, the accuracy of the α and β solutions is checked using a seven-point discretization of the Laplacian operator which has the symbolic form

$$\nabla^2 u\big|_{ijk} \sim \frac{6u_{i,j,k} - u_{i+1,j,k} - u_{i-1,j,k} - u_{i,j+1,k} - u_{i,j-1,k} - u_{i,j,k+1} - u_{i,j,k-1}}{h^2}, \qquad (33)$$

where

$$h = \Delta_x = \Delta_y = \Delta_z, \text{ and } u \equiv u(x,y,z).$$

In our calculations, $h = 10^{-4}$, which is a reasonable value for the "size" of the α and β solutions. All calculations have been performed in quadruple precision on an AMDAHL 470/ V5 computer. Finally, the plotting was done with the aide of a SYMVU package on a CALCOMP 770 plotter.[13]

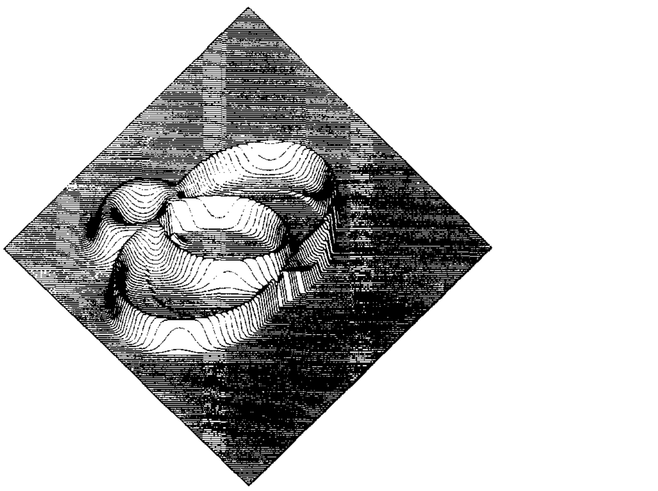


FIG. 7. The five-wave interaction function α_5 in the convenient representation tanh α_5 for $\phi = 80°$ and $\gamma = 135°$ in the domain $\alpha_5 = \{(x,y)| -30 \leqslant x \leqslant 30, -55 \leqslant y \leqslant 5\}$.

6. DISCUSSION OF tanh α_5 PLOT

(i) *Fig. 7*: This figure depicts the nonlinear superposition of three α_3 solutions to produce α_5, whose functional dependence reads $\alpha_5(x,y,z = -20;\theta_1,\ldots,\theta_5;\lambda_1,\ldots,\lambda_5)$, where x,y lie in the rectangular domain $\mathscr{D}_5 = \{(x,y)| -30 \leqslant x \leqslant 30, -55 \leqslant y \leqslant 5\}$. The altitude, or viewing angle ϕ, is 80°, while the angle of rotation in the XY-plane, or the azimuth γ, is 135°. Since each α_3 solution is characterized by *one* "ring" singularity, it seems reasonable that α_5 should contain *three*

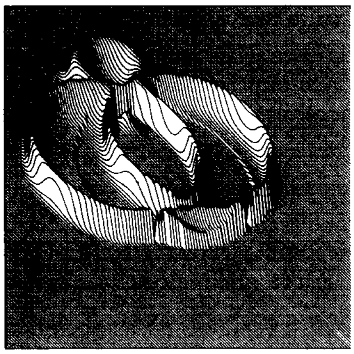


FIG. 8. The function tanh α_5 in the domain $\mathscr{D}_5$ for $\phi = 80°$ and $\gamma = 90°$.

more or less circularly shaped singularities. The function tanhα_5 is seen to possess exactly three such rings. One of these rings is clearly much smaller than the other two. It is amusing to note (at least in retrospect!) that the small ring didn't even show up in our earlier plots; it appeared only after we had managed to increase the accuracy of the Bäcklund parameters $\theta_1,...,\theta_5$ and $\lambda_1,...,\lambda_5$ to a sufficiently high level.

(ii)*Figs. 8,9 and 10*: The function tanhα_5 is depicted in the same domain $\mathscr{D}_5$ as in Fig. 7, but at different altitudes ϕ and azimuth angles γ.

(iii)*Fig.11*: This plot shows a *cut* view of tanhα_5 in the domain $\mathscr{D}_6 = \{(x,y)| -30 \leqslant x \leqslant 30, -25 \leqslant y \leqslant 35\}$ at an azimuth angle of $\gamma = 135°$ and an altitude of $\phi = 60°$.

7. DISCUSSION

We have illustrated that the Bäcklund–Bianchi method can be employed to generate, in three spatial dimensions, a most remarkable five-wave interaction α_5 for Liouville's equation. Our analysis included (a) the mathematical derivation of α_5 and its representation in closed form, (b) the graphical analysis of α_5 by means of three-dimensional computer plots, and (c) numerical studies and accuracy checks. Here are the principal results.

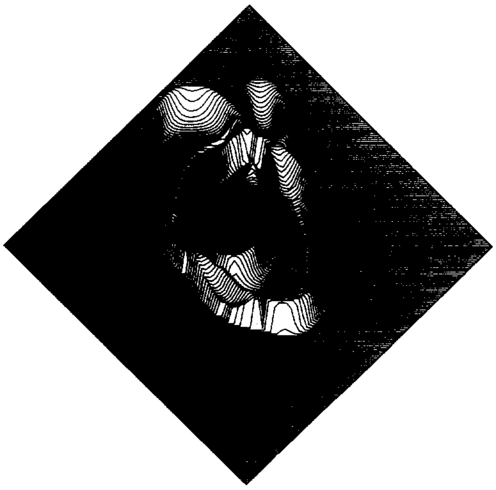


FIG. 9. The function tanh α_5 in the domain $\mathscr{D}_5$ for $\phi = 80°$ and $\gamma = 45°$.

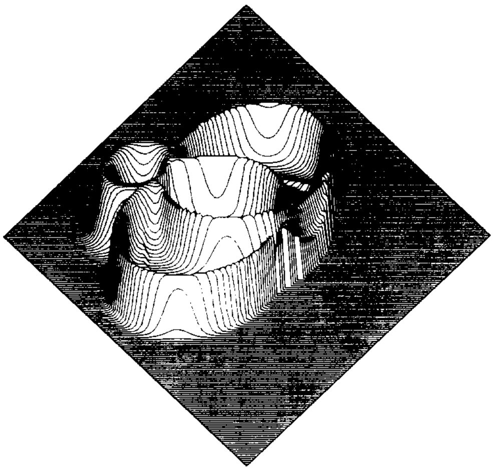


FIG. 10. The function tanh α_5 in the domain $\mathscr{D}_5$ for $\phi = 70°$ and $\gamma = 135°$.

(a) The mathematical derivation of α_5 proceeds from a Bäcklund–like transformation containing two essentially different functions α and β; the α's are solutions of Liouville's equation, while the auxiliary functions β must solve Laplace's equation. The construction of multiple solutions is aided considerably by the Bianchi diagram in Fig. 1.

(b) As pointed out in the Introduction, our second aim in this project was to employ three-dimensional plots to learn what α_3 and α_5 really "looked like." Our graphing techniques led to several surprising results, even for the relatively simple three-wave interaction $\alpha_3(x,y,z = -20)$, reported previously in the literature.[11] The latter was found to possess only a *single* singularity, a *ring* singularity, whose amazing symmetry can best be displayed by plotting, instead of α_3, the bounded function tanh α_3. We have also verified numerically that α_3 *is* a solution of Liouville's equation. Our latest accuracy figure stands at one part in 10^{14}.

Let us look now at α_5. In view of the ring structure of α_3, it should not be surprising to find that the five-wave interaction α_5 is dominated by three singularities which manifest themselves as rings in the functional representation of tanh α_5. We shall continue our discussion of these ring singularities at the end of this section. Before leaving this topic

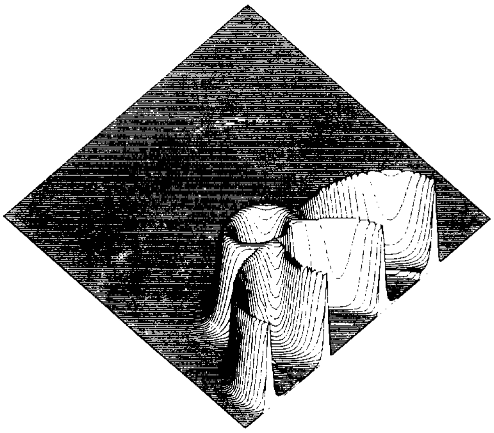


FIG. 11. A "cut view" of tanh α_5 in the domain $\mathscr{D}_6 = \{(x,y)| -30 \leqslant x \leqslant 30, -25 \leqslant y \leqslant 35\}$, $\gamma = 135°$, and $\phi = 60°$.

of 3D-plots, let us just reiterate that the above construction of α_5 is intimately associated with the auxiliary functions β_2 and β_4 which have likewise been derived and plotted here.

(c) There is no doubt that numerical studies are indispensable for projects of the present kind. This is particularly true for the verification of α_5 as an exact solution of Liouville's equation. The function α_5 is so complicated, with its nested dependence on trigonometric and hyperbolic functions, that it is virtually impossible to deduce its detailed structure from purely algebraic considerations. Using quadruple precision, we have been able to show that our $\alpha_5(x,y,z = -20)$ satisfies Liouville's equation to an accuracy of at least one part in 10^6. While this figure may not seem particularly impressive, it is nevertheless, remarkably good when considering (i) that the construction of α_5 is conditional upon solving *ten* nonlinear constraint equations [Eqs. (25)] and (ii) that accuracy checks on differential equations, having an exponential dependence, are *a priori* trickier than checks on equations with a sinusoidal dependence.

The accuracy of α_5 hinges decisively on the attainable values for the ten Bäcklund parameters θ_i and $\lambda_i, i = 1,2,...,5$, in Eq. (27). To obtain the accuracy for α_5 mentioned above, it was necessary to calculate the θ's and λ's to twenty significant figures and to solve these parameters from the $\mathscr{L}_{ij}$'s to an accuracy of one part in 10^{20}. Subsequent numerical studies have convinced us that, given enough computing time, the accuracy figure of 10^{-6} can be improved consistently.

Finally, let us speculate for a moment on the possible significance of the circular objects discovered in connection with α_3 and α_5. While it is safe to say that Liouville's basic solutions α_1 do not represent solitons of the conventional type—we recall that the three line singularities have "disappeared" in α_3!—the "composite" solution "tanh α_3" does seem to portray some kind of stable entity, a type of ring soliton, whose shape appears to be preserved in the nonlinear superposition of similar ring solitons. This is certainly evident from the graph of tanh α_5 in Fig. 7, which shows three distinct ring solitons. It would be interesting to know, of course, to what extent, if any, our tanh α_3- and tanh α_5-solutions are related to the monopole solutions arising in certain gauge theories.[9]

ACKNOWLEDGMENT

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grant No. A8063.

[1] A. V. Bäcklund, Math. Ann. **IX**, 297 (1876); **XVII**, 285 (1880);**XIX**, 387(1882); see also Lunds Univ. Årsskr. **19** (1883); L. Bianchi, *Lezioni di Geometria Differentiale*, Vol. I (Enrico Spoerri, Pisa, 1922), pp. 743–747; S. Lie, Math. Ann. VIII (3), 215 (1874); also Arch. Math. Naturvidenskab V (3), 282 (1880).

[2] J. Liouville, J. Math.**1** (18), 71 (1853).

[3] E. Picard, J. Math. **4** (9), 273 (1893); H. Poincaré, J. Math. **5** (4), 137 (1898); L. Bierberbach, Math. Ann. **77**, 173 (1916).

[4] O. W. Richardson, *The Emission of Electricity from Hot Bodies* (Longman Green, New York, 1921), pp. 50–54; M. v. Laue, Jahrb. Radio akt. Elektron. **XV**, 205 (1918).

[5] H. Bateman, *Partial Differential Equations of Mathematical Physics* (Cambridge U. P., Cambridge, England, 1964), pp. 166–168.

[6] W. F. Ames, *Nonlinear Partial Differential Equations in Engineering* (Academic, New York, 1965), pp. 180–183.

[7] S. Richardson, Mathematika **27**, 321 (1980).

[8] G.W. Walker, Proc. Roy. Soc. A **91**, 410 (1915).

[9] D. Olive, "Classical Solutions in Gauge Theories—Spherically Symmetric Monopoles—Lax Pairs and Toda Lattices," Imperial College Preprint ICTP/80/81 (Imperial College, London).

[10] G. P. Dzhorzhadze, A. K. Pogrebkov, and M. K. Polivanov, Theor. Math. Phys. **40**, 706 (1979).

[11] G.Leibbrandt, Lett. Math. Phys. **4**, 317 (1980).

[12] G. Leibbrandt, R.Morf, and S. S. Wang, J. Math. Phys. **21**, 1613 (1980).

[13] SYMVU is a three-dimensional plotting routine, developed by the Laboratory for Computer Graphics and Spatial Analysis, Harvard University, Cambridge, Mass.

On the projections of representation spaces of the symmetry group on the Minkowski space

Jan Rzewuski[a)]
CERN, Geneva, Switzerland

(Received 13 May 1981; accepted for publication 19 June 1981)

In this paper we generalize the projection of the representation space of the symmetry group SU(2,2)×U(2) on the Minkowski space to arbitrary internal symmetries U(m). The procedure involves certain restrictions on the coordinates of the representation space. Representations of the symmetry group in the restricted space and in the corresponding restricted Hilbert space are constructed.

PACS numbers: 02.40.Dr

The problem of projecting the representation space of the full symmetry group of the physical system on the Minkowski space arises in attempts to provide a common geometrical basis for internal as well as external symmetries[1,2] (cf. also Refs. 3–5 for further literature).

Already the lowest nontrivial linear representation space of the symmetry group provides such a basis. The physical Minkowski space must be embedded in this representation space in a way which is consistent with the group. This means that transformations of the group in the representation space should induce, via the projection, correct transformations of the Minkowski space. In particular, the last one should be invariant with respect to internal symmetries.

We assume that the physical symmetry group G is the direct product $G = G_1 \times G_2$ of the external and internal symmetry group in accordance with a theorem by Coleman–Mandula and Lopuszanski [cf., e.g., Ref. 6]. G is the conformal group or its Poincaré subgroup whereas a sufficiently general candidate for G_2 (internal symmetries) is U(m).

Since we shall be interested in the most economic extension of Minkowski space, we start with the lowest linear representation space of G which is $\mathbb{C}^{4m}$. The coordinates $\xi_{a;\alpha}$ ($\xi_{a;\alpha} \in \mathbb{C}, a = 1,...,4, \alpha = 1,...,m$) of this space form a complex $4 \times m$ matrix the elements of which transform with respect to $G = G_1 \times G_2$ according to

$$\xi'_{a;\alpha} = g^{(1)}{}_a{}^b g^{(2)}{}_\alpha{}^\beta \xi_{b;\beta}. \tag{1}$$

Let us first consider the case when $G_2 = \mathrm{GL}(2,\mathbb{C})$ or one of its subgroups [in the uninteresting case $G_2 = \mathrm{GL}(1,\mathbb{C})$, no projection on M_4 exists]. In this case the solution is known.[7] One first uses the Penrose projection $\mathbb{C}^8 \rightarrow M_4^c$[2]

$$z_\mu = x_\mu - iy_\mu = i\lambda \frac{\xi_{a;1}(\gamma^T\gamma_\mu)^{ab}\xi_{b;2}}{\xi_{a;1}[\gamma^T(1-\gamma_5)]^{ab}\xi_{b;2}}, \tag{2}$$

where the coordinates of a complex Minkowski space M_4^c appear as ratios of antisymmetric forms [$(\gamma_\mu)_a{}^b$ are Dirac matrices, $(\gamma^T)^{ab} = -(\gamma^T)_{ab}$ is a transposition matrix raising and lowering the indices according to $(\gamma^T\gamma_\mu)^{ab} = (\gamma^T)^{ac}(\gamma_\mu)_c{}^b$ and satisfying $(\gamma^T)_{ac}(\gamma^T)^{cb} = -\delta_a^b$, $(\gamma^T)^{ab} = -(\gamma^T)^{ba}$, $(\gamma^T\gamma_\mu)^{ab} = -(\gamma^T\gamma_\mu)^{ba}$]. The notation used here is explained in full detail in Ref. 8.

From (2) it is seen that the z_μ are invariant with respect to $G_2 = \mathrm{GL}(2,\mathbb{C})$ because the antisymmetric forms in the numerator and denominator of (2) are multiplied under $G_2 = \mathrm{GL}(2,\mathbb{C})$ by the common factor det GL(2,$\mathbb{C}$). It can also be shown [cf., e.g., Refs. 2–4 or 9] that conformal linear transformations $G_1 = \mathrm{SU}(2,2)$ of the complex variables $\xi_{a;\alpha}$ given by the generators

$$\left.\begin{aligned} d &= -\tfrac{1}{2}i\gamma_5, \\ p_\mu &= -\lambda^{-1}\gamma^+\gamma_\mu, \\ k_\mu &= -\lambda\gamma^-\gamma_\mu, \\ m_{\mu\nu} &= \tfrac{1}{4}i[\gamma_\mu,\gamma_\nu], \end{aligned}\right\} \quad \gamma^\pm = \tfrac{1}{2}(1 \pm \gamma_5), \tag{3}$$

(acting on the first index) induce the correct nonlinear conformal transformations of the complex Minkowski variables

$$\begin{aligned} Dz_\lambda &= iz_\lambda, \\ P_\mu z_\lambda &= ig_{\mu\lambda}, \\ K_\mu z_\lambda &= -ig_{\mu\lambda}z^2 + 2iz_\mu z_\lambda, \\ M_{\mu\nu}z_\lambda &= -ig_{\mu\lambda}z_\nu + ig_{\nu\lambda}z_\mu \end{aligned} \tag{4}$$

[λ in (3) is a parameter with dimension of length].

It is seen from (4) that the real part x_μ of z_μ transforms with respect to translations (P_μ) like a vector whereas the imaginary part y_μ is translationally invariant (behaves like a coordinate difference). We are obliged, therefore, to identify x_μ with the coordinates of the real Minkowski space M_4.

It can be shown[4,9] that projection (2) is the only projection in terms of antisymmetric forms consistent with the group $\mathscr{P} \times \mathrm{GL}(2,\mathbb{C})$, where $\mathscr{P} \subset \mathrm{SU}(2,2)$ is the Poincaré subgroup of the conformal group. Consistency with the whole conformal group is a consequence of consistence with $\mathscr{P}$. It can also be shown [4,9] that the two other projections in terms of symmetric and Hermitian tensors are identical with (2) in the case when these tensors are simple (i.e., when they are represented by bilinear forms).

To approach the homogeneous space of the group in the case when $G_2 = \mathrm{SL}(2,\mathbb{C})$ or $G_2 = \mathrm{SU}(2)$, we parametrize the projection $\mathbb{C}^8 \rightarrow M_4$, provided by the real part of (2), by introducing the four real conformal invariants

$$r_{;\mu} := (\sigma_\mu)^{\dot\alpha\beta}\xi^*_{a;\alpha}f^{\dot a b}\xi_{b;\beta}, \tag{5}$$

[a)] On leave of absence from the Institute of Theoretical Physics, Wroclaw University, Cybulskiego 36, Wroclaw, Poland.

where σ_i $(i=1,2,3)$ are Pauli matrices, $\sigma_0 = 1$, and $f^{\dot a b}$ is the Hermitian 4×4 matrix with eigenvalues $+1$, $+1$, -1, -1, determining SU(2,2) [in representation (3) $f^{\dot a b} = i(\gamma_4)^{\dot a c}(\gamma_5)_c{}^b$]. We first split the variables $\xi_{a;\alpha}$ into two parts,

$$\xi^{\pm} = \gamma^{\pm}\xi, \tag{6}$$

by means of the projectors

$$\gamma^{\pm} = \tfrac{1}{2}(1 \pm \gamma_5). \tag{7}$$

From (3) it follows that ξ^- is translationally invariant. It is seen from (2) that

$$z_\mu = i\lambda \frac{\xi^{+}_{a;1}(\gamma^T\gamma_\mu)^{ab}\xi^{-}_{b;2} + \xi^{-}_{a;1}(\gamma^T\gamma_\mu)^{ab}\xi^{+}_{b;2}}{2\xi^{-}_{a;1}(\gamma^T)^{ab}\xi^{-}_{b;2}} \tag{8}$$

is a linear combination of ξ^+ with coefficients depending on ξ^-. Relations (8) can be inversed with respect to ξ^+, and we obtain

$$\xi^{+} = -i\lambda^{-1}z^\mu\gamma_\mu\xi^{-}. \tag{9}$$

It is seen that we can perform a change of variables

$$\{\xi_{a;\alpha}\}\rightarrow\{x_\mu, y_\mu, \xi^{-}_{a;\alpha}\} \tag{10}$$

in which y_μ and $\xi^{-}_{a;\alpha}$ represent the translationally invariant parameters of the particular embedding of M_4 in $\mathbb{C}^8$.

It can be shown further that $y_{\mu;}$ is connected with $r_{;\mu}$ (from now on we use the semicolon whenever it is necessary to distinguish vectors with respect to external and internal symmetries) by means of an inversible relation

$$r_{;\nu} = -2\lambda^{-1}y^{\mu;}r_{\mu;\nu} \tag{11}$$

in which the coefficients $r_{\mu;\nu}$ depend on ξ^- only and satisfy the orthogonality relations[7,8]

$$r_{\mu;\lambda}r^{\mu;}{}_{;\zeta} = 4g_{;\lambda\zeta}|\xi^{-}_{a;1}(\gamma^T)^{ab}\xi^{-}_{b;2}|^2,$$
$$r_{\mu;\lambda}r_{;\nu}{}^{;\lambda} = 4g_{\mu\nu;}|\xi^{-}_{a;1}(\gamma^T)^{ab}\xi^{-}_{b;2}|^2. \tag{12}$$

We obtain the desired parametrization by replacing $y_{\mu;}$ by $r_{;\mu}$ and $\xi^{-}_{a;\mu}$

$$\{\xi_{a;\alpha}\}\rightarrow\{x_{\mu;}, r_{;\mu}, \xi^{-}_{a;\alpha}\}. \tag{13}$$

The submanifolds

$$\mathscr{I}_1 := \tfrac{1}{4}r_{;\mu}r^{;\mu} = \text{const} \tag{14}$$

and

$$\mathscr{I}_1 = \text{const}, \quad \mathscr{I}_2 := r_{;0} = \text{const},$$
$$\mathscr{I}_1 + \mathscr{I}_2{}^2 \geqslant 0 \tag{15}$$

of $\mathbb{C}^8$ are invariant with respect to SU(2,2)×SL(2,$\mathbb{C}$) or SU(2,2)×SU(2), respectively. It is certain, therefore, that they contain the homogeneous manifolds of these groups and, if there are no other independent invariants, coincide with these manifolds. They also contain in a consistent and unique way the Minkowski space and provide therefore the most economic embedding.

In the case when we restrict the conformal symmetry SU(2,2) to its Poincaré subgroup, another invariant appears, namely $|\xi^{-}_{a;1}(\gamma^T)^{ab}\xi^{-}_{b;2}|^2$. We obtain, therefore, 14- or 13-dimensional submanifolds of $\mathbb{C}^8$ containing M in a way consistent with the groups $\mathscr{P}\times$SL(2,$\mathbb{C}$) or $\mathscr{P}\times$SU(2), respectively.

Due to the fact that $r_{;\mu}$ is invariant with respect to a phase transformation $\xi\rightarrow\xi' = e^{i\phi}\xi$, the above statements are valid also in the case of $G_2 = \text{U}(2)$.

Let us go over now to the general case $G_2 = \text{GL}(m,\mathbb{C})$ or one of its subgroups. Now, the representation space is $\mathbb{C}^{4m}$ and the coordinates $\xi_{a;\alpha}$ form a $4\times m$ complex matrix. We can, therefore, construct, according to (2), $\frac{1}{2}m(m-1)$ complex vectors

$$z_\mu^{(\alpha,\beta)} = \frac{i\lambda}{2}\frac{\xi_{a;\alpha}(\gamma^T\gamma_\mu)^{ab}\xi_{b;\beta}}{\xi^{-}_{a;\alpha}(\gamma^T)^{ab}\xi^{-}_{b;\beta}}, \tag{16}$$

which, according to the statement made after formula (2), transform in the correct way (4) if the $\xi_{a;\alpha}$ undergo a conformal transformation $G_1 = \text{SU}(2,2)$. In fact, we have infinitely many such vectors given by the basic antisymmetric forms appearing in (16):

$$z_\mu(\kappa) = \frac{i\lambda}{2}\frac{\kappa^{\alpha\beta}\xi_{a;\alpha}(\gamma^T\gamma_\mu)^{ab}\xi_{b;\beta}}{\kappa^{\alpha\beta}\xi^{-}_{a;\alpha}(\gamma^T)^{ab}\xi^{-}_{b;\beta}}. \tag{17}$$

Moreover neither (16) nor (17) are invariant with respect to $G_2 = \text{GL}(2,\mathbb{C})$ or its subgroups SL(m,$\mathbb{C}$) or U(m) because the antisymmetric bilinear forms

$$\xi_{a_1a_2;\alpha_1\alpha_2} := \begin{vmatrix} \xi_{a_1;\alpha_1} & \xi_{a_1;\alpha_2} \\ \xi_{a_2;\alpha_1} & \xi_{a_2;\alpha_2} \end{vmatrix} \tag{18}$$

provide a $[\binom{4}{2}\times\binom{m}{2}]$-dimensional irreducible representation of GL(4,$\mathbb{C}$)×GL(m,$\mathbb{C}$) and, therefore, the $z_\mu(\kappa)$ transform into each other like ratios of linear combinations of vectors of such a representation. This situation is unsatisfactory because the physical space M_4 must be invariant with respect to internal symmetries.

A way out of this difficulty is provided by imposing on the coordinates $\xi_{a;\alpha}$ constraints which would reduce the representation space of G, and simultaneously, the variety of z_μ. We shall consider here $4(m-2)$ linear independent constraints

$$\lambda_{ia} = 0, \quad i = 3,4,\ldots,m, \quad a = 1,\ldots,4, \tag{19}$$

where

$$\lambda_{ia} = C_i{}^\alpha\xi_{a;\alpha}, \tag{20}$$

and $C_i{}^\alpha$ are $(m-2)$ independent vectors. Conditions (19) are invariant with respect to the group G in the sense that

$$\lambda'_{ia} := C'_i{}^\alpha\xi'_{a;\alpha} = g^{(1)}{}_a{}^b\lambda_{ib}, \tag{21}$$

where

$$C'_i{}^\alpha = C_i{}^\beta(g^{(2)\,-1})_\beta{}^\alpha. \tag{22}$$

It follows from (21) that

$$\lambda_{ia} = 0 \Longleftrightarrow \lambda'_{ia} = 0. \tag{23}$$

We shall prove in this note that conditions (19) imply that all the points $z^{(\alpha,\beta)}$, $\alpha,\beta = 1,\ldots,m$, defined in (16) as well as the points $z_\mu(\kappa)$ defined in (17) coincide and are invariant with respect to the group $G_2 = \text{GL}(2,\mathbb{C})$ of internal symmetries. The proof is immediate, but we shall postpone it for convenience to a later place, first deriving some general results concerning representations of G in the restricted space.

To this end, we introduce new variables

$$\begin{aligned} &\eta_{a;A} = \xi_{a;A}, \quad a = 1,\ldots,4, \quad A = 1,2, \\ &\lambda_{ia} = C_i{}^\alpha \xi_{a;\alpha}, \quad a = 1,\ldots,4, \quad i = 3,\ldots,m. \end{aligned} \tag{24}$$

We can assume, without loss of generality, that the subdeterminant $\det\{C_i{}^{\bar A}\}$ $(i,\bar A = 3,\ldots,m)$ of the matrix $\{C_i{}^\alpha\}$ is different from zero

$$\det\{C_i{}^{\bar A}\} \neq 0. \tag{25}$$

We can solve, therefore, Eqs. (24) with respect to $\xi_{a;\alpha}$ and obtain

$$\xi_{a;\alpha} = a_\alpha{}^A \eta_{a;A} + b_\alpha{}^i \lambda_{ia}, \tag{26}$$

where

$$\begin{aligned} &a_A{}^B = \delta^B_A, \quad b_A{}^i = 0, \quad A,B = 1,2, \\ &a_{\bar A}{}^B = -b_{\bar A}{}^i C_i{}^B, \quad b_{\bar A}{}^i = [\det\{C_i{}^{\bar A}\}]^{-1} A_{\bar A}{}^i, \\ &(B = 1,2, \ i,\bar A = 3,\ldots,m) \end{aligned} \tag{27}$$

and $A_{\bar A}{}^i$ is the adjoint determinant of order $m - 3$ corresponding to the element $C_i{}^{\bar A}$ of the matrix $\{C_i{}^{\bar A}\}$. The coefficients $b_{\bar A}{}^i$ satisfy the relations (following from general properties of determinants)

$$b_{\bar A}{}^i C_i{}^{\bar B} = \delta^{\bar B}_{\bar A}, \quad C_i{}^{\bar A} b_{\bar A}{}^k = \delta^k_i. \tag{28}$$

In a new coordinate system $\xi'_{a;\alpha} = g_\alpha{}^\beta \xi_{a;\beta}$, $g \in G_2$, we have

$$\begin{aligned} &\eta'_{a;A} = \xi'_{a;A}, \\ &\lambda'_{ia} = C'_i{}^\alpha \xi'_{a;\alpha} = \lambda_{ia}, \end{aligned} \tag{29}$$

and the inverse relations

$$\xi'_{a;\alpha} = a(ga)_\alpha{}^A \eta'_{a;A} + b\,(ga,gb\,)_\alpha{}^i \lambda'_{ia}, \tag{30}$$

where $a(ga)_\alpha{}^A$ and $b\,(ga,gb\,)_\alpha{}^i$ are solutions of the equations

$$\begin{aligned} &(ga)_\alpha{}^A = a(ga)_\alpha{}^B (ga)_B{}^A, \\ &(gb\,)_\alpha{}^i = b\,(ga,gb\,)_\alpha{}^i + a(ga)_\alpha{}^B (gb\,)_B{}^i. \end{aligned} \tag{31}$$

The coefficients a and b satisfy relations

$$\begin{aligned} &a(hga) = a(ha(ga)), \\ &b\,(hga,hgb\,) = hb\,(ga,gb\,) - a(ha(ga))hb\,(ga,gb\,), \end{aligned} \tag{32}$$

which ensure uniqueness of the functions $a(ga)$ and $b\,(ga,gb\,)$ defined in (31) on the group G_2 (we do not consider here the external symmetries G_1 because the transformation properties of η and λ under this group are obvious). One also easily verifies that $a(ga)$ and $b\,(ga,gb\,)$ are the same functions of $C'_i{}^\alpha$ as a and b are of $C_i{}^\alpha$.

From (26) and (29) we derive finally the relations between η and η':

$$\begin{aligned} &\eta'_{a;A} = (ga)_A{}^B \eta_{a;B} + (gb\,)_A{}^i \lambda_{ia}, \quad \lambda'_{ia} = \lambda_{ia}, \\ &\eta_{a;A} = [g^{-1}a(ga)]_A{}^B \eta'_{a;B} + [g^{-1}b\,(ga,gb\,)]_A{}^i \lambda'_{ia}. \end{aligned} \tag{33}$$

This closes the system of mutual relations of the variables $\xi_{a;\alpha}$ and $\eta_{a;A}$, λ_{ia} in different coordinate systems. One has to keep in mind, however, that all these relations are local and valid only in the neighborhood of $g = e$ for which $\det\{C'_i{}^{\bar A}\} \neq 0$.

Imposing of the invariant conditions $\lambda_{ia} = 0$ $(a = 1,\ldots,4; i = 3,\ldots,m)$ replaces the linear representation in $\mathbb{C}^{4m}$ by a nonlinear local representation in the space $\mathbb{C}^8 \times \mathbb{C}^{2(m-2)} = \mathbb{C}^{2(m+2)}$ of the variables $\eta_{a;A}$ and $a_{\bar A}{}^B$ given by

$$\begin{aligned} &\eta'_{a;A} = (ga)_A{}^B \eta_{a;B}, \\ &a'_\alpha{}^A = a(ga)_\alpha{}^A. \end{aligned} \tag{34}$$

Relation (30) now takes the form

$$\xi'_{a;\alpha} = a(ga)_\alpha{}^A \eta'_{a;A} = (ga)_\alpha{}^A \eta_{a;A}. \tag{35}$$

Introducing this into (18), we prove both statements concerning (16) and (17). We obtain first

$$\xi'_{ab;\alpha\beta} = (ga)_{\alpha\beta}{}^{12} \eta_{ab;12}, \tag{36}$$

where

$$\xi'_{ab;\alpha\beta} := \begin{vmatrix} \xi'_{a;\alpha} & \xi'_{a;\beta} \\ \xi'_{b;\alpha} & \xi'_{b;\beta} \end{vmatrix}, \quad \eta_{ab;12} := \begin{vmatrix} \eta_{a;1} & \eta_{a;2} \\ \eta_{b;1} & \eta_{b;2} \end{vmatrix}, \tag{37}$$

$$(ga)_{\alpha\beta}{}^{12} := \begin{vmatrix} (ga)_\alpha{}^1 & (ga)_\alpha{}^2 \\ (ga)_\beta{}^1 & (ga)_\beta{}^2 \end{vmatrix},$$

and consequently

$$\xi'_{ab;\alpha\beta}/\xi'_{cd;\alpha\beta} = \eta_{ab;12}/\eta_{cd;12}. \tag{38}$$

From (38) the first part of the statement (coincidence) follows for $g = e$ and the second part (invariance) for arbitrary g (satisfying condition $\det\{C'_i{}^{\bar A}\} \neq 0$) and $\alpha,\beta = 1, 2$.

By imposing additional invariant conditions on the matrix $\{\xi_{a;\alpha}\}$ we have obtained a unique and consistent projection $\mathbb{C}^{2(m+2)} \to M_4$ in terms of the new variables $\eta_{a;A}$ and a nonlinear local representation of G in the space of the variables $\eta_{a;A}$ and $a_{\bar A}{}^B$.

For applications, it is important to construct representations of G in spaces of functions which take into account conditions (19) (first quantization). We start with the infinitesimal representation of G in $L^2(\mathbb{C}^{4m})$, which is given by the infinitesimal generators

$$X := \frac{\partial}{\partial \xi_{a;\alpha}} X_a{}^b \xi_{b;\alpha}, \quad X' := \frac{\partial}{\partial \xi_{a;\alpha}} X'_\alpha{}^\beta \xi_{a;\beta}, \tag{39}$$

for G_1 and G_2, respectively, where $X_a{}^b$ and $X'_\alpha{}^\beta$ are the corresponding generators in $\mathbb{C}^{4m}$.

To introduce condition (19), we carry out the change of variables described in (24) and (26) and the corresponding change in derivatives

$$\begin{aligned} &\frac{\partial}{\partial \xi_{a;A}} = \frac{\partial}{\partial \eta_{a;A}} + \frac{\partial}{\partial \lambda_{ia}} C_i{}^A, \quad \frac{\partial}{\partial \xi_{a;\bar A}} = \frac{\partial}{\partial \lambda_{ia}} C_i{}^{\bar A}, \\ &\frac{\partial}{\partial \eta_{a;A}} = \frac{\partial}{\partial \xi_{a;\alpha}} a_\alpha{}^A, \quad \frac{\partial}{\partial \lambda_{ia}} = \frac{\partial}{\partial \xi_{a;\alpha}} b_\alpha{}^i, \end{aligned} \tag{40}$$

to obtain

$$X = \frac{\partial}{\partial \eta_{a;A}} X_a{}^b \eta_{b;A} + \frac{\partial}{\partial \lambda_{ia}} X_a{}^b \lambda_{ib}, \tag{41}$$

$$\begin{aligned} X' = &\frac{\partial}{\partial \eta_{a;A}} (X'a)_A{}^B \eta_{a;B} + \frac{\partial}{\partial \eta_{a;A}} (X'b\,)_A{}^i \lambda_{ia} \\ &+ \frac{\partial}{\partial \lambda_{ia}} (CX'a)_i{}^B \eta_{a;B} + \frac{\partial}{\partial \lambda_{ia}} (CX'b\,)_i{}^k \lambda_{ka}. \end{aligned} \tag{42}$$

To restrict the representation to functions on the plane $\lambda_{ia} = 0$, it is convenient to use the relation

$$C_i{}^\alpha X'{}_\alpha{}^\beta \frac{\partial}{\partial C_i{}^\beta} = - \frac{\partial}{\partial \lambda_{ia}} (CX'a)_i{}^B \eta_{a;B} - \frac{\partial}{\partial \lambda_{ia}} (CX'b)_i{}^k \lambda_{ka}, \tag{43}$$

by means of which we can rewrite the generator (42) in the form

$$X' = \frac{\partial}{\partial \eta_{a;A}} \{(X'a)_A{}^B \eta_{a;B} + (X'b)_A{}^i \lambda_{ia}\} - C_i{}^\alpha X'{}_\alpha{}^\beta \frac{\partial}{\partial C_i{}^\beta}. \tag{44}$$

The functions on the plane $\lambda_{ia} = 0$ can be written in the form

$$\phi(\eta,c) := \int f(\xi_{a;\alpha}) \prod_{a=1}^{4} \prod_{i=3}^{m} \delta(\lambda_{ia}) d\lambda_{ia} = f(a_\alpha{}^A \eta_{a;A}) \tag{45}$$

[the integral representation does not involve the assumption (25) and is, therefore, more general]. The generators (41) and (44), when restricted to such functions, are

$$X|_{\lambda=0} = \frac{\partial}{\partial \eta_{a;A}} X_a{}^b \eta_{b;A},$$
$$X'|_{\lambda=0} = \frac{\partial}{\partial \eta_{a;A}} (X'a)_A{}^B \eta_{a;B} - C_i{}^\alpha X'{}_\alpha{}^\beta \frac{\partial}{\partial C_i{}^\beta}. \tag{46}$$

The dependence on $C_i{}^\alpha$ of the function $\phi(\eta,c) = f(a_\alpha{}^A \eta_{a;A})$ occurs through the variables $a_{\bar{A}}{}^A$ and we can put, therefore, for the last term in the second expression (46)

$$- C_i{}^\alpha X'{}_\alpha{}^\beta \frac{\partial}{\partial C_i{}^\beta} = [(X'a) - (aX'a)]_{\bar{A}}{}^B \frac{\partial}{\partial a_{\bar{A}}{}^B}. \tag{47}$$

Second quantization in the case of $G_2 = \mathrm{GL}(2,\mathbb{C})$ or its subgroups was given a preliminary treatment in Ref. 10, invariant differential operators in the same case were investigated in Ref. 7. The present construction is intended to provide a basis to extend those investigations to the physically more interesting cases of SU(3) and higher internal symmetries.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor J. Prentki for his warm hospitality extended to me during my stay at the Theoretical Division of CERN. I would also like to express my thanks to Professor V. Glaser and Dr. L. Turko for many stimulating discussions.

[1] J. Rzewuski, Bull. Acad. Polon. Sci. **6**, 261, 335 (1958); **7**, 571 (1959); **8**, 777, 783 (1960); Nuovo Cimento **5**, 942 (1958); Acta Phys. Polon. **17**, 417 (1958); **18**, 549 (1959).

[2] R. Penrose, Ann. Phys. **10**, 171 (1960); J. Math. Phys. **8**, 345 (1967); *The Structure of Space-Time*, Batelle Rencontres 1967, edited by C. M. deWitt and J. A. Wheeler (Benjamin, New York, 1968); R. Penrose and M. A. H. MacCallum, Phys. Rep. C **6**, 241 (1972).

[3] R. Ablamowicz, J. Mozrzymas, Z. Oziewicz, and J. Rzewuski, Rep. Math. Phys. **14**, 89 (1978).

[4] R. Ablamowicz, Z. Oziewicz, and J. Rzewuski, J. Math. Phys. **23**, 231 (1982).

[5] L. P. Hughston, *Twistors and Particles*, in Lecture Notes in Physics, Vol. 97 (Springer, New York, 1979).

[6] J. Lopuszanski, J. Math. Phys. **12**, 2401 (1971).

[7] J. Rzewuski, Bull. Acad. Polon. Sci. **28**, 63 (1980).

[8] J. Rzewuski, Acta Univ. Wratislaviensis (to be published).

[9] R. Ablamowicz, Z. Oziewicz, and J. Rzewuski, Bull. Acad. Polon. Sci. **27**, 201 (1979).

[10] J. Rzewuski, Bull. Acad. Polon. Sci. **27**, 177 (1979).

On a certain class of two-sided continuous local semimartingales: Toward a sample-wise characterization of the Nelson process

Kunio Yasue
Department of Theoretical Physics, University of Geneva, CH-1211 Geneva 4, Switzerland

(Received 1 May 1981; accepted for publication 25 June 1981)

Working with the extended framework of stochastic integrals recently discovered by Itô, a complex of stochastic processes inherent in quantum mechanics, the Nelson process, is characterized in terms of sample paths. It is shown that the Nelson process belongs to a certain class of two-sided continuous local semimartingales. Several basics of stochastic calculus in this class are presented. Stochastic calculus of variations is applied in this class to construct the Nelson process and to further illustrate some details of its sample paths. Examples are the bound states, the two-slit interference, and the gravity in quantum mechanics.

PACS numbers: 02.50.Ey, 03.65. – w

INTRODUCTION

A mathematical formulation of quantum mechanics in which the notion of stochastic processes plays an essential role was first presented in its systematic form by Nelson.[1–4] It gave rise to a new quantization scheme which has been called the stochastic quantization procedure and applied extensively to many physical problems.[5–16] Conceptual enlargements have been also taken into account.[17–27] Nowadays it can be understood as one of the representatives of quantum dynamics.[23] Unfortunately there seem to be some elementary confusions arising from improper criticisms of the use of stochastic processes in quantum mechanics. The origin of the confusions is the conceptual gap between the mathematician's refined notion of stochastic processes and the physicist's classical notion of them. It is not so easy to overcome this gap and to evaluate the stochastic quantization procedure properly if one is used to the classical notion of stochastic processes such as appear in macroscopic classical statistical physics. Therefore it seems meaningful to make the mathematical characterization of the Nelson process clear for the purposes of making the gap as small as possible and facilitating the physicist's understanding of the stochastic quantization procedure.

In the present paper I will clarify the class of stochastic processes to which the Nelson process belongs, working with the extended framework of stochastic integrals recently discovered by Itô. It is a class of two-sided continuous local semimartingales. This class provides a wide ring in which the stochastic quantization in terms of the Nelson process can safely play a role. (Sec. 1) Several basics of stochastic calculus in this class are presented. (Sec. 1). Stochastic calculus of variations is applied in this class to reconsider the Nelson process globally, which permits us to illustrate the details of its sample path behavior. (Sec. 2). I will take the bound state, the two-slits interference, and the gravity in quantum mechanics as illustrating examples. (Sec. 3).

The final introductory remarks are the following: Although I will classify and represent the Nelson process in the modern language of mar tingale integrals, the essentials do not differ otherwise from Nelson's original analysis. It seems surprising that Nelson's observation and analysis were so refined and close to reality as to still survive in the modern theory of stochastic integrals. Chronological order proves this. In 1942 stochastic integrals were first discovered by Itô. Doob suggested the use of martingale theory approach to stochastic integrals in his book of 1953. On the basis of the work of Itô and Doob, Nelson developed his idea before 1966, just a sunrise epoch of the modern theory of stochastic integrals (Fisk in 1963, Courrège in 1963, Kunita and Watanabe in 1967, and Meyer in 1967).

1. STOCHASTIC INTEGRALS AND TWO-SIDED CONTINUOUS LOCAL SEMIMARTINGALES

This section will be devoted to an exposition of a certain class of continuous local semimartingales in which the stochastic processes appearing in stochastic quantization can safely be treated. The framework is Itô's extended stochastic integrals, and the main source for this section is his recent paper.[28]

Let $(\Omega, \mathfrak{A}, Pr)$ be a base probability space, where Ω is a certain nonempty set and a σ-algebra $\mathfrak{A}$ of subsets of Ω is the domain of a probability measure Pr. A measurable mapping from Ω to $\mathbb{R}^n$ is a random variable. Its image of each element $\omega\in\Omega$ is a sample. A family of random variables indexed by a continuous time parameter, $X = \{X_t | -\infty < t < \infty\} = \{X_t(\omega) | \omega\in\Omega, -\infty < t < \infty\}$, is a stochastic process. For each $\omega\in\Omega$ there is a family $\{X_t(\omega) | -\infty < t < \infty\}$ which defines a function $X(\omega)$ from $(-\infty, \infty)$ to $\mathbb{R}^n$. This is called a sample path or sample function. A stochastic process is thus a complex of sample paths. A mathematical property of a stochastic process should be understood as that of every sample path Pr-almost surely. Here "Pr-almost surely" means "except sample paths corresponding to Pr-null sets" and will be abbreviated (a.s.). For example, a stochastic process whose sample paths are continuous (a.s.) is said to be continuous and one with sample functions of bounded variation on any finite interval (a.s.) is said to be of locally bounded variation.

Let us fix a right-continuous increasing family of sub-σ-algebras of $\mathfrak{A}$, $\mathscr{P} = \{\mathscr{P}_t | -\infty < t < \infty\}$ such that $\mathscr{P}_t$ contains every Pr-null set. This is called a reference family or a filtration on the time interval $(-\infty, \infty)$. Take an arbitrary q in the interval $(-\infty, \infty)$; then $\mathscr{P}^{(q)} = \{\mathscr{P}_t^{(q)} = \mathscr{P}_{q+t} | 0 \leqslant t < \infty\}$ defines a reference family on the time interval

$[0,\infty)$. A stochastic process $M^{(q)} = \{M_t^{(q)} | 0 \leqslant t < \infty\}$ is said to be a local $\mathscr{P}^{(q)}$ martingale if there exists an increasing sequence of stopping times, $\{\theta_k\}_{k=1}^{\infty}$, adapted to $\mathscr{P}^{(q)}$ such that the stopped processes $M_{(k)}^{(q)} = \{M_{\min(t,\theta_k)}^{(q)} | 0 \leqslant t < \infty\}$, $k = 1,2,\ldots$, are $\mathscr{P}^{(q)}$ martingales, that is, $E[M_{(k)t+u}^{(q)} | \mathscr{P}_t^{(q)}] = M_{(k)t}^{(q)}$ (a.s.), where $E[\cdot | \mathscr{B}]$ is the conditional expectation with respect to a sub-σ-algebra $\mathscr{B} \subset \mathfrak{A}$. Now we reach an important class of stochastic processes. A stochastic process $X = \{X_t | -\infty < t < \infty\}$ is a continuous local $\mathscr{P}$ semimartingale (or $\mathscr{P}$ quasimartingale) if $X^{(q)} = \{X_t^{(q)} = X_{q+t} | 0 \leqslant t < \infty\}$, for every q in the interval $(-\infty,\infty)$, admits a decomposition

$$X^{(q)} = M^{(q)} + V^{(q)}, \quad V_0^{(q)} = 0, \tag{1.1}$$

where $M^{(q)}$ is a continuous local $\mathscr{P}^{(q)}$ martingale and $V^{(q)}$ is a $\mathscr{P}^{(q)}$-adapted process (i.e., $V_t^{(q)}$ is $\mathscr{P}_t^{(q)}$-measurable for each t in $[0,\infty)$) of locally bounded variation. This decomposition is unique and called a canonical $\mathscr{P}^{(q)}$ decomposition. The totality of continuous local $\mathscr{P}$ semimartingales is denoted by $Q(\mathscr{P})$. Let us denote by $\mathbb{L}(\mathscr{P},dX)$ for each X in $Q(\mathscr{P})$ the totality of all $\mathscr{P}$-adapted stochastic processes $Y = \{Y_t | -\infty < t < \infty\}$'s such that $\int_0^u |Y_t^{(q)}| (dM_t^{(q)})^2 < \infty$ and $\int_0^u |Y_t^{(q)}|\, |dV_t^{(q)}| < \infty$ for every q in $(-\infty,\infty)$ and u in $[0,\infty)$, where $(dM_t^{(q)})^2$ is the Lebesgue–Stieltjes measure induced by the quadratic variation of $M^{(q)}$, and $dV_t^{(q)}$ is that induced by the absolute variation of $V^{(q)}$. Here we are in the position to introduce the notion of stochastic integrals.

Let $X \in Q(\mathscr{P})$, then for any $Y \in \mathbb{L}(\mathscr{P},dX)$ the stochastic $\mathscr{P}$ integral is defined by

$$\begin{aligned} \mathscr{P}\text{-}\int_r^t Y_s\, dX_s &= \mathscr{P}^{(r)}\text{-}\int_0^{t-r} Y_s^{(r)}\, dX_s^{(r)}, \\ &= \mathscr{P}^{(r)}\text{-}\int_0^{t-r} Y_s^{(r)} dM_s^{(r)} + \int_0^{t-r} Y_s^{(r)} dV_s^{(r)}, \end{aligned} \tag{1.2}$$

for any $r \leqslant t$ in $(-\infty,\infty)$, where the first term of the right-hand side is the usual $\mathscr{P}^{(r)}$-martingale integral and the second term the sample-wise Lebesgue–Stieltjes integral.

The stochastic $\mathscr{P}$ integral (1.2) defines a stochastic process $Z = \{Z_t | -\infty < t < \infty\}$ such that $Z_t^{(q)} = \mathscr{P}\text{-}\int_q^{q+t} Y_s\, dX_s$. It is evident that Z belongs to $Q(\mathscr{P})$. If $s \leqslant u \leqslant v$, then

$$\mathscr{P}\text{-}\int_s^u Y_t\, dX_t + \mathscr{P}\text{-}\int_u^v Y_t\, dX_t = \mathscr{P}\text{-}\int_s^v Y_t\, dX_t. \tag{1.3}$$

There are other notions of stochastic integrals. A family of sub-σ-algebras $\mathscr{F} = \{\mathscr{F}_t | -\infty < t < \infty\}$ is said to be a time-reversed reference family if $\mathscr{F}^* = \{\mathscr{F}_t^* = \mathscr{F}_{-t} | -\infty < t < \infty\}$ is a reference family. In other words, the time-reversed reference family $\mathscr{F}$ is a left-continuous decreasing family of sub-σ-algebras of $\mathfrak{A}$ such that $\mathscr{F}_t$ contains every Pr-null set. A stochastic process $X = \{X_t | -\infty < t < \infty\}$ is said to be time-reversed continuous local $\mathscr{F}$ semimartingale if $X^* = \{X_t^* = X_{-t} | -\infty < t < \infty\}$ belongs to $Q(\mathscr{F}^*)$. The totality of time-reversed continuous local $\mathscr{F}$ semimartingales is denoted by $Q(\mathscr{F})$. A stochastic process $Y = \{Y_t | -\infty < t < \infty\}$ is said to belong to $\mathbb{L}(\mathscr{F},dX)$ if $Y^* = \{Y_t^* = Y_{-t} | -\infty < t < \infty\}$ belongs to $\mathbb{L}(\mathscr{F}^*,dX^*)$.

Let $X \in Q(\mathscr{F})$, then for any $Y \in \mathbb{L}(\mathscr{F},dX)$ the stochastic $\mathscr{F}$ integral is defined by

$$\mathscr{F}\text{-}\int_t^u Y_s\, dX_s = \mathscr{F}^*\text{-}\int_{-t}^{-u} Y_s^*\, dX_s^*, \tag{1.4}$$

for any $u \leqslant t$ in $(-\infty,\infty)$. The stochastic $\mathscr{F}$ integral (1.4) also defines a stochastic process of class $Q(\mathscr{F})$.

Let $\mathscr{S} = (\mathscr{P},\mathscr{F})$, $Q(\mathscr{S}) = Q(\mathscr{P}) \cap Q(\mathscr{F})$, and $\mathbb{L}(\mathscr{S},dX) = \mathbb{L}(\mathscr{P},dX) \cap \mathbb{L}(\mathscr{F},dX)$ for $X \in Q(\mathscr{S})$. If $X \in Q(\mathscr{S})$, then X is called a two-sided continuous local semimartingale on the interval $(-\infty,\infty)$. Now the most faithful notion of stochastic integrals is ready. Let $X \in Q(\mathscr{S})$, then for any $Y \in \mathbb{L}(\mathscr{S},dX)$ the symmetric stochastic $\mathscr{S}$ integral is defined by

$$\mathscr{S}\text{-}\int_r^t Y_s \circ dX_s = \frac{1}{2}\left(\mathscr{P}\text{-}\int_r^t Y_s\, dX_s - \mathscr{F}\text{-}\int_t^r Y_s\, dX_s\right) \tag{1.5}$$

for $r \leqslant t$, and by

$$\mathscr{S}\text{-}\int_r^t Y_s \circ dX_s = \frac{1}{2}\left(\mathscr{F}\text{-}\int_r^t Y_s\, dX_s - \mathscr{P}\text{-}\int_t^r Y_s\, dX_s\right) \tag{1.6}$$

for $t \leqslant r$.

The symmetric stochastic $\mathscr{S}$ integral (1.5) and (1.6) defines a stochastic process $Z = \{Z_t | -\infty < t < \infty\}$ by putting $Z_t^{(q)} = \mathscr{L}\text{-}\int_q^{q+t} Y_s \circ dX_s$. If $\mathscr{P} = \mathscr{F}^*$, then Z belongs to $Q(\mathscr{S})$. If $\mathscr{F}^*$ is finer than $\mathscr{P}$, that is, $\mathscr{F}_t^* \supset \mathscr{P}_t$ for every t in $(-\infty,\infty)$, then Z belongs to $Q(\mathscr{P})$. Z belongs to $Q(\mathscr{F})$ in the opposite case. The followings are some of the useful formulae for the symmetric stochastic $\mathscr{S}$ integrals. For $X \in Q(\mathscr{S})$, $Y \in \mathbb{L}(\mathscr{S},dX)$ and r, t, u in $(-\infty,\infty)$,

$$\mathscr{S}\text{-}\int_r^t Y_s \circ dX_s + \mathscr{S}\text{-}\int_t^u Y_s \circ dX_s = \mathscr{S}\text{-}\int_r^u Y_s \circ dX_s, \tag{1.7}$$

and so

$$\mathscr{S}\text{-}\int_r^t Y_s \circ dX_s = -\mathscr{S}\text{-}\int_t^r Y_s \circ dX_s. \tag{1.8}$$

If Y is continuous furthermore, then

$$\mathscr{S}\text{-}\int_r^t Y_s \circ dX_s = \underset{N\to\infty}{\text{l.i.p.}} \sum_{j=1}^N \tfrac{1}{2}(Y_{t_j} + Y_{t_{j-1}})(X_{t_j} - X_{t_{j-1}}) \tag{1.9}$$

for any $r \leqslant t$ in $(-\infty,\infty)$, where $t_j = r + j(t-r)/N$, $j = 0, 1, 2,\ldots,N$, is a division of the interval $[r,t]$ and l.i.p. is the limit in probability. The last formula is extremely important for extending the symmetric stochastic $\mathscr{S}$ integral as also being defined a little outside of $Q(\mathscr{S})$. If X and Y belong to $Q(\mathscr{P})$, then the limit in the right-hand side of Eq. (1.9) becomes

$$\mathscr{P}\text{-}\int_r^t Y_s \circ dX_s + \frac{1}{2}\int_r^t dY_s\, dX_s, \tag{1.10}$$

which will be denoted by $\int_r^t Y_s \circ dX_s$, where the second term (i.e., the quadratic variation) is defined by

$$\int_r^t dY_s\, dX_s = \underset{N\to\infty}{\text{l.i.p.}} \sum_{j=1}^N (Y_{t_j} - Y_{t_{j-1}})(X_{t_j} - X_{t_{j-1}}). \tag{1.11}$$

Itô calls Eq. (1.10) the forward symmetric stochastic integral which is one of the extensions of the symmetric stochastic $\mathscr{S}$

integral to the region $Q(\mathscr{P})\backslash Q(\mathscr{S})$. The extension to the opposite region $Q(\mathscr{F})\backslash Q(\mathscr{S})$ is also given by the right-hand side of Eq. (1.9), resulting in

$$\mathscr{F}\text{ - }\int_r^t Y_s\, dX_s - \frac{1}{2}\int_r^t dY_s\, dX_s, \tag{1.12}$$

which will be denoted also by $\int_r^t Y_s \circ dX_s$. This is the backward symmetric stochastic integral of Itô. Consequently we have the cyclic equalities between different notions of stochastic integrals if X and Y are restricted to lie in $Q(\mathscr{S})$.

$$\begin{aligned}\mathscr{S}\text{ - }\int_r^t Y_s \circ dX_s &= \mathscr{P}\text{ - }\int_r^t Y_s\, dX_s + \frac{1}{2}\int_r^t dY_s\, dX_s \\ &= \mathscr{F}\text{ - }\int_r^t Y_s dX_s - \frac{1}{2}\int_r^t dY_s dX_s \\ &= \underset{N\to\infty}{\text{l.i.p.}} \sum_{j=1}^{N} \tfrac{1}{2}(Y_{t_j} + Y_{t_{j-1}})(X_{t_j} - X_{t_{j-1}}).\end{aligned} \tag{1.13}$$

The stochastic process Z given by the symmetric stochastic $\mathscr{S}$ integral (1.13) on $Q(\mathscr{S})$ by putting $Z_t^{(q)} = \mathscr{S}$ - $\int_q^{q+t} Y_s \circ dX_s$ certainly belongs to $Q(\mathscr{S})$. In other words, $Q(\mathscr{S})$ is closed under the symmetric stochastic $\mathscr{S}$ integral. Immediate outputs of those equalities are

$$\int_r^t d(X_s Y_s) = \mathscr{S}\text{ - }\int_r^t Y_s \circ dX_s + \mathscr{S}\text{ - }\int_r^t X_s \circ dY_s, \tag{1.14}$$

$$\begin{aligned}X_t Y_t - X_r Y_r &= \mathscr{P}\text{ - }\int_r^t Y_s\, dX_s + \mathscr{F}\text{ - }\int_r^t X_s\, dY_s \\ &= \mathscr{F}\text{ - }\int_r^t Y_s\, dX_s + \mathscr{P}\text{ - }\int_r^t X_s\, dY_s.\end{aligned} \tag{1.15}$$

The last result (1.15) is a generalization of a theorem of Nelson;[1] notice that it holds for stochastic processes of class $Q(\mathscr{S})$. Again Eq. (1.14) can be extended to the region $Q(\mathscr{P})\Delta Q(\mathscr{F}) = [Q(\mathscr{P})\backslash Q(\mathscr{S})]\cup[Q(\mathscr{F})\backslash Q(\mathscr{S})]$. Namely for X and Y in $Q(\mathscr{P})\backslash Q(\mathscr{S})$ or those in $Q(\mathscr{F})\backslash Q(\mathscr{S})$, we have

$$\int_r^t d(X_s Y_s) = \int_r^t Y_s \circ dX_s + \int_r^t X_s \circ dY_s. \tag{1.16}$$

It is convenient to present here the basic formulae in a differential form.

$$\begin{aligned}\mathscr{S}\text{ - } Y_t \circ dX_t &= \mathscr{P}\text{ - } Y_t dX_t + \tfrac{1}{2} dY_t dX_t \\ &= \mathscr{F}\text{ - } Y_t dX_t - \tfrac{1}{2} dY_t dX_t,\end{aligned} \tag{1.17}$$

$$\begin{aligned}d(X_t Y_t) &= \mathscr{S}\text{ - } Y_t \circ dX_t + \mathscr{S}\text{ - } X_t \circ dY_t \\ &= \mathscr{P}\text{ - } Y_t dX_t + \mathscr{F}\text{ - } X_t dY_t \\ &= \mathscr{F}\text{ - } Y_t dX_t + \mathscr{P}\text{ - } X_t dY_t.\end{aligned} \tag{1.18}$$

Now we will consider the mean Lebesgue differentiability in a certain sense of continuous local semimartingales. Let $X = \{X_t \mid -\infty < t < \infty\}$ be a stochastic process belonging to $Q(\mathscr{P})$, then $X^{(q)} = \{X_t^{(q)} = X_{q+t} \mid 0 \leqslant t < \infty\}$ has a unique canonical decomposition $X^{(q)} = V^{(q)} + M^{(q)}$ for any q in $(-\infty, \infty)$. If $V^{(q)}$ is an absolutely continuous process for any q in $(-\infty, \infty)$, then X is said to be mean $\mathscr{P}$ differentiable. Let $V^{(q)\prime}$ be its Lebesgue derivative process and define a stochastic process $DX = \{DX_t \mid -\infty < t < \infty\}$ such that $DX^{(q)} = \{DX_t^{(q)} = DX_{q+t} = V_t^{(q)\prime} \mid -\infty < t < \infty\}$ for any q in $(-\infty, \infty)$. DX is a locally integrable process adapted to $\mathscr{P}$ and called the mean (Lebesgue) $\mathscr{P}$ derivative. Let $X \in Q(\mathscr{F})$, then X^*, the time-reversed version of X belongs to $Q(\mathscr{F}^*)$. X is said to be mean $\mathscr{F}$ differentiable if X^* is mean $\mathscr{F}^*$ differentiable. The mean Lebesque $\mathscr{F}$ derivative of X is defined to be DX^* with opposite sign and denoted by D_*X. This is a locally integrable process adapted to $\mathscr{F}$. Let us denote by $\mathbb{D}(\mathscr{P})$, $\mathbb{D}(\mathscr{F})$, and $\mathbb{D}(\mathscr{S})$, respectively, the totality of mean $\mathscr{P}$-differentiable processes in $Q(\mathscr{P})$, that of mean $\mathscr{F}$-differentiable processes in $Q(\mathscr{F})$, and their intersection $\mathbb{D}(\mathscr{P})\cap\mathbb{D}(\mathscr{F})$ in $Q(\mathscr{S})$. It is straightforward to see that $\lim_{h\downarrow 0}|E[X_{t+h} - X_t|\mathscr{P}_t] - DX_t h| = 0$ (a.s.) if $X \in \mathbb{D}(\mathscr{P})$ and $\lim_{h\downarrow 0}|E[X_t - X_{t-h}|\mathscr{F}_t] - D_*X_t h| = 0$ (a.s.) if $X \in \mathbb{D}(\mathscr{F})$. The formula (1.15) yields, for X and Y in $\mathbb{D}(\mathscr{S})$, the formula of integration by parts,

$$\begin{aligned}E[X_t Y_t - X_r Y_r] &= E\left[\int_r^t (Y_s DX_s + X_s D_* Y_s) ds\right] \\ &= E\left[\int_r^t (X_s DY_s + Y_s D_* X_s) ds\right],\end{aligned} \tag{1.19}$$

for any $r \leqslant t$ in $(-\infty, \infty)$. We have also

$$E\left[\mathscr{S}\text{ - }\int_r^t Y_s \circ dX_s\right] = E\left[\int_r^t Y_s \frac{1}{2}(DX_s + D_*X_s) ds\right]. \tag{1.20}$$

Finally we reach a class of two-sided continuous local semimartingales in which the stochastic quantization procedure of Nelson works safely. Let us denote by $\mathbb{N}(\mathscr{S})$ a subclass of stochastic processes in $\mathbb{D}(\mathscr{S})$ such that $X \in \mathbb{N}(\mathscr{S})$ implies ($X \in \mathbb{D}(\mathscr{S})$, of course) $DX \in \mathbb{D}(\mathscr{F})$ and $D_*X \in \mathbb{D}(\mathscr{P})$. A stochastic process belonging to $\mathbb{N}(\mathscr{S})$ is called a Nelson process. This choice of the name seems appropriate because Nelson considered such stochastic processes first, though his class is wider than the present one. He worked with the $L_1(\Omega, Pr)$ and/or $L_2(\Omega, Pr)$ analyses, whereas we work with the sample paths analysis. If X is a Nelson process, then D_*DX and DD_*X are locally integrable processes. A measurable function F: $\mathbb{R}^n \to \mathbb{R}^n$ is said to be N admissible if $F(X)$, for any $X \in \mathbb{N}(\mathscr{S})$, defines a locally integrable process. For example, a locally bounded function is N admissible. Let X be a Nelson process and F be an N-admissible function; then the equation

$$\tfrac{1}{2}(DD_*X + D_*DX) = F(X) \text{ (a.s.)} \tag{1.21}$$

is well posed since this is a relation connecting two locally integrable processes. Because the stochastic processes employed in the stochastic quantization are assumed to satisfy the equation of motion of this type, the class $\mathbb{N}(\mathscr{S})$ is sufficient to include them. A stochastic process $X \in \mathbb{N}(\mathscr{S})$ satisfying Eq. (1.21), though this would not be unique, is said to be an F trajectory. The totality of Nelson processes X's such that DX and D_*X belong to $\mathbb{D}(\mathscr{S})$ and $DDX = D_*D_*X$ holds (a.s.) is denoted by $\mathbb{N}'(\mathscr{S})$. A stochastic process in this class is called a locally probability conserving Nelson process, or simply a Nelson process hereafter. This class was first noticed by Etim.[9]

2. STOCHASTIC CALCULUS OF VARIATIONS AND HOW TO SEE THE SAMPLE PATHS

This section will be devoted to showing how one can illustrate the sample path behavior of a Nelson process which represents quantum dynamics by working with stochastic calculus of variations.[29–31] developed recently. First we will present the stochastic quantization procedure in the terminology of $Q(\mathscr{S})$. Second the basics of stochastic calculus of variations will be presented also in the framework of $Q(\mathscr{S})$. Then we will proceed to the main exposition of this section.

Let us start with a differentiable dynamical system in $\mathbb{R}^n$, $x = \{x_t \mid -\infty < t < \infty\}$. Together with Newton's equation of motion

$$m\ddot{x} = F(x), \tag{2.1}$$

where m is a mass parameter and a continuous locally bounded function F: $\mathbb{R}^n \rightarrow \mathbb{R}^n$ represents forces acting on the system, it defines a classical dynamical system. A function of class C^2, $t \mapsto x_t$, is said to be a classical F trajectory if it satisfies Eq. (2.1). If F is integrable, that is, if there is a function V: $\mathbb{R}^n \rightarrow \mathbb{R}$ of class C^1 such that $F = -\nabla V$ holds, the classical dynamical system is said to be conservative, where ∇ is the gradient. A classical F trajectory x: $(-\infty, \infty) \mapsto \mathbb{R}^n$ of a conservative system can be characterized by Hamilton's principle of least action

$$\int_r^t \left\{ \tfrac{1}{2} m |\dot{x}_s|^2 - V(x_s) \right\} ds = \text{stationary}, \tag{2.2}$$

for any r and t in $(-\infty, \infty)$. Indeed the Euler equation equivalent to the variational condition (2.2) becomes

$$m\ddot{x} = -\nabla V(x). \tag{2.3}$$

A family of hypotheses and substitutions is called a quantization procedure (or simply, a quantization) if it joins each conservative classical dynamical system (2.3) and the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \left(-\frac{\hbar^2}{2m} \Delta + V \right) \psi, \tag{2.4}$$

where $\hbar$ is the Planck constant divided by 2π, Δ is the Laplacian. Once a solution $\psi = \{\psi_t \in L_2(\mathbb{R}^n, C) \mid -\infty < t < \infty\}$ to Eq. (2.4) is given, the probability distribution of the system at any time t is assumed to be $|\psi_t|^2$ times the Lebesgue measure provided that $\|\psi_t\| = 1$ for every t in $(-\infty, \infty)$.

The stochastic quantization procedure originally proposed by Nelson consists of the following hypotheses and substitutions. First, consider the classical conservative dynamical system in $\mathbb{R}^n$ $x = \{x_t \mid -\infty < t < \infty\}$, subject to the equation of motion (2.3). Second, replace it by a Nelson process $X = \{X_t \mid -\infty < t < \infty\}$ subject to the equation.

$$\tfrac{1}{2} m (DD_*X + D_*DX) = -\nabla V(X). \tag{2.5}$$

This may be understood as a minimal extension of Newton's equation of motion (2.3). Third, assume X belongs to $\mathbb{N}'(\mathscr{S})$. Namely, $DDX = D_*D_*X$ holds (a.s.). Fourth, assume X to be a Markov process and that there exist two functions b, b_*: $\mathbb{R}^n \times (-\infty, \infty) \rightarrow \mathbb{R}^n$ of class C^2 such that $DX_t = b(X_t, t)$ and $D_*X_t = b_*(X_t, t)$ hold for every t in $(-\infty, \infty)$. Fifth, the quadratic variation of X is assumed to be $(\hbar/2m)$ times the Lebesgue measure, that is,

$$\lim_{h \downarrow 0} \left| E\left[(X_{t+h} - X_t)^{\otimes 2} \middle| \mathscr{P}_t \right] - \frac{\hbar}{2m} h \right| = 0 \ \text{(a.s.)}. \tag{2.6}$$

Sixth, the existence of the Lebesgue density of the probability distribution is assumed, that is, $Pr(\{\omega \in \Omega \mid X_t(\omega) \in d^n x\}) = p(x,t) d^n x$ for every t in $(-\infty, \infty)$. Last, $(b + b_*)$ is assumed to be integrable, that is, $b + b_* = (\hbar/m) \nabla S$ for a certain function S: $\mathbb{R}^n \times (-\infty, \infty) \mapsto \mathbb{R}$. This procedure is enough to reach quantum mechanics. These seven hypotheses and substitutions are really equivalent to the Schrödinger equation (2.4) for the wavefunction $\psi = p^{1/2} \times \exp(iS)$.[1–4] It is worthwhile to notice that the third hypothesis is not necessary since it comes from the fourth and fifth hypotheses and the formula (1.9). They also yield $b_* = b - (\hbar/m) \nabla p/p$ and $\lim_{h \downarrow 0} | E[(X_t - X_{t-h})^{\otimes 2} | \mathscr{F}_t] - (\hbar/2m) h | = 0$ (a.s.).

Next, we shall proceed to the exposition of stochastic calculus of variations. In our original formulation of stochastic calculus of variations we worked with the $L_2(\Omega, Pr)$ functional analysis.[29,30] Here we will present it in terms of two-sided continuous local semimartingales and so work with the sample paths analysis. This has the merit of defining the action integral in a sample-wise way, whereas we had to define it as a Bochner integral in the $L_2(\Omega, Pr)$ analysis. Let L: $(\mathbb{R}^n)^3 \rightarrow \mathbb{R}$ be a function of class C^1, and consider a functional I on $\mathbb{D}(\mathscr{S})$ defined by

$$I(X) = E\left[\int_r^t L(X_s, DX_s, D_*X_s) \, ds \right] \tag{2.7}$$

for any $X \in \mathbb{D}(\mathscr{S})$. If $L(X, DX, D_*X) = \{L(X_t, DX_t, D_*X_t) \mid -\infty < t < \infty\}$ is a locally integrable process, then $I(X)$ is well defined. If not, we simply put $I(X) = \infty$. A typical example of such a functional is the action integral

$$I_{QM}(X) = E\left[\int_r^t \left\{ \tfrac{1}{2} \left(\tfrac{1}{2} m |DX_s|^2 + \tfrac{1}{2} m |D_*X_s|^2 \right) - V(X_s) \right\} ds \right], \tag{2.8}$$

where m and V are same as in Eq. (2.2). For X's in $\mathbb{D}(\mathscr{S})$ such that DX and D_*X are locally square integrable processes, I_{QM} is well defined. Let X and Y be two stochastic processes belonging to $\mathbb{D}(\mathscr{S})$. The Y component of the differential of the functional I at X is given by

$$dI(X, Y) = \frac{d}{da} I(X + aY) \Big|_{a=0}. \tag{2.9}$$

The integration by parts formula (1.19) leads to

$$\begin{aligned} dI(X,Y) &= E\left[\int_r^t \left\{ \frac{\partial L}{\partial X_s} - D_*\left(\frac{\partial L}{\partial DX_s} \right) - D\left(\frac{\partial L}{\partial D_*X_s} \right) \right\} \cdot Y_s \, ds \right] \\ &\quad + E\left[\left(\frac{\partial L}{\partial DX_s} + \frac{\partial L}{\partial D_*X_s} \right) \cdot Y_s \Big|_r^t \right] \end{aligned} \tag{2.10}$$

if $\partial L / \partial DX \in \mathbb{D}(\mathscr{F})$ and $\partial L / \partial D_*X \in \mathbb{D}(\mathscr{P})$. From now on we assume that Y is a conditional stochastic process in $\mathbb{D}(\mathscr{S})$ such that $Y_r = Y_t = 0$ (a.s.). For such Y's the second term of the right-hand side of Eq. (2.10) disappears and we get the fundamental theorem of stochastic calculus of variations: The differential of the functional I at $X \in \mathbb{D}(\mathscr{S})$ vanishes if and only if X is subject to the equation

$$D_*\left(\frac{\partial L}{\partial DX}\right)+D\left(\frac{\partial L}{\partial D_*X}\right)-\frac{\partial L}{\partial X}=0. \tag{2.11}$$

Such X makes the functional I stationary. This extends the Euler equation in ordinary calculus of variations and will be called the Euler–Nelson equation since it also extends Eq. (2.5) proposed by Nelson. Indeed, the Euler–Nelson equation for the action integral I_{QM} given by Eq. (2.8) coincides with Eq. (2.5). This will provide the possibility of reformulating the stochastic quantization procedure in terms of the least action principle $I_{QM}(X) =$ stationary.

Let X be a conditional stochastic process in $\mathbb{D}(\mathscr{S})$ subject to the Euler–Nelson equation (2.11) such that $X_t = x\in\mathbb{R}^n$ (a.s.), and Y be an arbitrary conditional process in $\mathbb{D}(\mathscr{S})$ such that $Y_r = 0$ and $Y_t = y\in\mathbb{R}^n$ (a.s.). Then Eq. (2.10) claims $dI(X,Y) = E\,[(\partial L/\partial DX_t) + (\partial L/\partial D_*X_t)|X_t = x]\cdot y$, and therefore we obtain

$$E\left[\frac{\partial L}{\partial DX_t}+\frac{\partial L}{\partial D_*X_t}\,\middle|\,X_t=x\right]$$
$$=\nabla E\left[\int_r^t L(X_s,DX_s,D_*X_s)ds\,\middle|\,X_t=x\right]. \tag{2.12}$$

For more about stochastic calculus of variations, see Ref. 29.

We shall now show how stochastic calculus of variations works in getting informations about the sample path behavior of the Nelson process X. Because the seven hypotheses on X of the stochastic quantization procedure are equivalent to the Schrödinger equation (2.4), we may say that the Nelson process representing the quantum dynamics is determined by the Schrödinger equation. Therefore, the wavefunction ψ seems to contain some information about the sample path behavior. Since the probability distribution $p(x,t)d^nx = Pr(\{\omega\in\Omega\,|X_t(\omega)\in d^nx\})$ does not help us to see sample path behavior, we must look for it in the phase of the wavefunction ψ. It is quite interesting to see that the key concept of quantum mechanics—the phase of wavefunction—has a close relation to the sample path behavior of the quantum dynamics.

The stochastic quantization procedure can also be formulated within the realm of stochastic calculus of variations. The first, third, fifth, and sixth hypotheses remain unchanged, but the second one (i.e., the dynamical assumption) is replaced by the least action principle $I_{QM}(X) =$ stationary. The last one is not necessary, for it comes from this least action principle. Indeed Eq. (2.12) yields

$$m\tfrac{1}{2}(b(x,t)+b_*(x,t))=E\left[\int_r^t L(X_s,DX_s,D_*X_s)ds\,\middle|\,X_t=x\right]. \tag{2.13}$$

We see immediately that those six hypotheses and substitutions are equivalent to the Schrödinger equation (2.4), where the wavefunction ψ is expressed explicitly in terms of sample paths;

$$\psi(x,t)=p(x,t)^{1/2}$$
$$\times\exp\left[\frac{i}{\hbar}E\left[\int_r^t L(X_s,DX_s,D_*X_s)\,ds\,\middle|\,X_t=x\right]\right]. \tag{2.14}$$

Here we have a program to visualize the quantum dynamics as the sample path behavior of the Nelson process: Let $X\in\mathrm{N}(\mathscr{S})$ be the Nelson process completely determined by the basic hypotheses of the stochastic quantization procedure. Since $X\in Q(\mathscr{S})$, $X^{(r)} = \{X_t^{(r)} = X_{r+t}|0<t<\infty\}$ becomes a usual continuous local semimartingale on the interval $[0,\infty)$ and we can think of its sample paths easily. Almost all sample paths $X^{(r)}(\omega)$'s are chosen in such a way that the basic dynamical assumption (2.5) holds. The sample path behavior thus defined can affect the phase of the wavefunction through Eq. (2.14). Conversely the wavefunction ψ—a solution to the Schrödinger equation (2.4)—provides the information about the sample path behavior through its phase. This fact will help us to illustrate the quantum dynamics in terms of the sample paths of the Nelson process.

3. SAMPLE PATH ANALYSIS OF QUANTUM DYNAMICS: EXAMPLES

A. The bound state

Let us consider a solution to the Schrödinger equation (2.4) of the form $\psi(x,t) = u(x)\exp(-i\lambda t/h)$, $t>0$, where $u\in L_2(\mathbb{R}^n,C)$ is an eigenfunction of the stationary Schrödinger equation

$$\left(-\frac{\hbar^2}{2m}\Delta+V\right)u=\lambda u, \tag{3.1}$$

and λ is the eigenvalue. This solution represents a quantum mechanical bound state and defines a cross-section $X^{(0)}$ of the Nelson process $X\in\mathrm{N}(\mathscr{S})$ as a continuous local semimartingale on the interval $[0,\infty)$. Since b, b_*, and p given by u do not depend on t, the cross-section $X^{(0)}$ is a stationary diffusion process on the interval $[0,\infty)$ with invariant measure $p(x)d^nx = |u(x)|^2d^nx$ generated by a diffusion equation

$$\frac{\partial f}{\partial t}=\left(b(x)\cdot\nabla+\frac{\hbar}{2m}\Delta\right)f. \tag{3.2}$$

In such a case of the bound state, fortunately, we can construct the Nelson process X directly from its cross-section $X^{(0)}$. Namely the Nelson process X for the bound state u is a stationary Markov process on the interval $(-\infty,\infty)$ with invariant measure $p(x)d^nx = |u(x)|^2d^nx$ which has the same transition probability law as the diffusion process $X^{(0)}$.[25] Therefore we can illustrate freely the sample path behavior of the Nelson process (i.e., the quantum dynamics) in the bound state by applying the transition probability law of the diffusion process (3.2). The success of the recent probabilistic approach to quantum mechanical tunneling and instanton analysis certainly owes much to this fact.[7–12,24] As the function $b(x)$ is a logarithmic derivative of the stationary distribution density in the bound state case, it may take an infinite value. From a mathematical point of view, however, this is harmless provided that the set $\{x\in\mathbb{R}^n|b(x)=\infty\}$ has vanishing Lebesgue measure.[25] The stationary Markov process X (i.e., the Nelson process for the bound state) can be generated by the Dirichlet form approach of Fukushima even if the probability distribution density p has nodal surfaces.[24] We can also investigate the sample path behavior of the Nelson process for the bound state with nodal surfaces by Fukushima's Dirichlet form approach. It has been shown quite recently that there exists a *Pr*-nonnull set of sample paths which go across the nodal surfaces depending on the degree of zero.[32]

B. Quantum mechanical interference

Let us consider the famous thought experiment of quantum mechanical interferences. A quantum mechanical particle (e.g., a nonrelativistic electron) is emitted from a certain source at a certain time r, and it reaches a certain point of a detecting film at a time t certainly later than r. Between the source and the detecting film one places an infinite plate with two parallel slits separated by a small distance a. Therefore at a certain time after r and before t the particle goes across the slits. The probability distribution of the particle on the detecting film, which can be obtained by successive emissions of the particles, shows the interference pattern when the two slits are open and one does not know through which one the particle goes. Once one knows that the particle goes through one of the slits, say the slit A (for example, by closing the other one, say the slit B) the interference pattern does not arise. Here we present the interpretation of this quantum mechanical interference with the use of the Nelson process and its sample path behavior. Let V_A, V_B, and V_{AB} be the potential energies representing the existence of the infinite plate with the slit A open and B closed, that with A closed and B open, and that with both A and B open, respectively. In each case there corresponds a Nelson process representing the quantum dynamics of the particle. Let X^A, X^B, and X^{AB} be the Nelson processes for those three cases, which make the action integral (2.8) with respect to V_A, V_B, and V_{AB}, respectively, stationary. They are different stochastic processes in the class $\mathrm{N}(\mathscr{S})$. Since we are interested in the case in which the particle certainly reaches some point on the detecting film, we see only the sample paths going through the two slits.

Case 1: The slit A is open and B closed. Let x be an arbitrary point on the detecting film and Ω^A_x be the totality of sample paths of the Nelson process X^A which reach x at a certain time t. By the formula (2.14) the wavefunction $\psi_A(x,t)$ is given in the form

$$\psi_A(x,t) = p_A(x,t)^{1/2} \times\exp\Big[\frac{i}{\hbar}E^A_x\Big[\int_r^t L(X^A_s, DX^A_s, D_*X^A_s)\,ds\Big]\Big], \tag{3.3}$$

where $E^A_x[\cdot]$ is the expectation over the sample paths belonging to Ω^A_x, and $p_A(x,t)d^nx = Pr(\{\omega\in\Omega\,|X^A_t(\omega)\in d^nx\})$ is the probability of finding the particle in the vicinity of x. This gives the detecting pattern on the film for the successive emissions of the particles.

Case 2: The slit A is closed and B open. The wavefunction $\psi_B(x,t)$ is

$$\psi_B(x,t) = p_B(x,t)^{1/2} \times\exp\Big[\frac{i}{\hbar}E^B_x\Big[\int_r^t L(X^B_s, DX^B_s, D_*X^B_s)\,ds\Big]\Big], \tag{3.4}$$

where $E^B_x[\cdot]$ is the expectation over the totality of sample paths of X^B, denoted by Ω^B_x, reaching x at t and $p_B(x,t)d^nx = Pr(\{\omega\in\Omega\,|X^B_t(\omega)\in d^nx\})$. If the distance between the source and the slits is long enough compared with that between the slits A and B, $\psi_B(x,t)$ coincides with the wavefunction of case 1 shifted by a, that is,

$$\begin{aligned}\psi_B(x,t) &= \psi_A(x-a,t)\\ &= p_A(x-a,t)^{1/2}\\ &\quad\times\exp\Big[\frac{i}{\hbar}E^A_{x-a}\Big[\int_r^t L(X^A_s, DX^A_s, D_*X^A_s)ds\Big]\Big],\end{aligned} \tag{3.5}$$

where a here should be understood as a vector parallel to the film and the plate with norm a. The detecting pattern $p_B(x,t)\,d^nx$ also becomes a shifted one $p_A(x-a,t)d^nx$.

Case 3: Both A and B are open. Let Ω^{AB}_x be the totality of sample paths of the Nelson process X^{AB} which reach x at t. Then the wavefunction $\psi_{AB}(x,t)$ becomes

$$\psi_{AB}(x,t) = p_{AB}(x,t)^{1/2} \times\exp\Big[\frac{i}{\hbar}E^{AB}_x\Big[\int_r^t L(X^{AB}_s, DX^{AB}_s, D_*X^{AB}_s)\,ds\Big]\Big], \tag{3.6}$$

where $E^{AB}_x[\cdot]$ is the expectation over Ω^{AB}_x and $p_{AB}(x,t)d^nx = Pr(\{\omega\in\Omega\,|X^{AB}_t(\omega)\in d^nx\})$. Notice that $\Omega^{AB}_x \neq \Omega^A_x\cup\Omega^B_x$. However Ω^{AB}_x admits a decomposition $\Omega^{AB}_x = \Omega^{ABA}_x\cup\Omega^{ABB}_x$, where Ω^{ABA}_x and Ω^{ABB}_x are the totalities of sample paths of the Nelson process X^{AB} which go through the slit A and the slit B, respectively. This decomposition can be used to get another expression for the wavefunction $\psi_{AB}(x,t)$. Since we have the decomposition $\Omega^{AB}_x = \Omega^{ABA}_x\cup\Omega^{ABB}_x$, at a certain time s after r and before t the probability distribution $p_{AB}(z,s)$, $z\in\mathbb{R}^n$, consists of two functions p^s_{ABA} and p^s_{ABB} $\in L_1(\mathbb{R}^n)$ with nonoverlapping supports. At that time the wavefunction $\psi_{AB}(z,s)$ is therefore the sum of two corresponding parts $\psi_{ABA}(z,s)$ and $\psi_{ABB}(z,s)$. The linearity of the Schrödinger equation (2.4) claims, then,

$$\psi_{AB}(x,t) = \psi_{ABA}(x,t) + \psi_{ABB}(x,t), \tag{3.7}$$

where

$$\psi_{ABA}(x,t) = p_{ABA}(x,t)^{1/2} \times\exp\Big[\frac{i}{\hbar}E^{ABA}_x\Big[\int_r^t L(X^{AB}_s, DX^{AB}_s, D_*X^{AB}_s)\,ds\Big]\Big], \tag{3.8}$$

$$\psi_{ABB}(x,t) = p_{ABB}(x,t)^{1/2} \times\exp\Big[\frac{i}{\hbar}E^{ABB}_x\Big[\int_r^t L(X^{AB}_s, DX^{AB}_s, D_*X^{AB}_s)\,ds\Big]\Big], \tag{3.9}$$

$p_{ABA}(x,t)$ and $p_{ABB}(x,t)$ are probability distributions of X^{AB}_t satisfying the initial conditions $p_{ABA}(z,s) = p^s_{ABA}(z)$ and $p_{ABB}(z,s) = p^s_{ABB}(z)$, respectively, and E^{ABA}_x and E^{ABB}_x are the expectation over Ω^{ABA}_x and Ω^{ABB}_x, respectively. By the symmetry consideration again, we know that ψ_{ABB} coincides with ψ_{ABA} shifted by a,

$$\begin{aligned}\psi_{ABB}(x,t) &= \psi_{ABA}(x-a,t)\\ &= p_{ABA}(x-a,t)^{1/2}\\ &\quad\times\exp\left[\frac{i}{\hbar}E^{ABA}_{x-a}\left[\int_r^t L(X^{AB}_s,DX^{AB}_s,D_*X^{AB}_s)\,ds\right]\right].\end{aligned}\tag{3.10}$$

The detecting pattern in this case is $p_{AB}(x,t) = |\psi_{AB}(x,t)|^2$ which can be computed through Eqs. (3.7), (3.8), and (3.10), obtaining

$$p_{AB} = p_{ABA} + p_{ABB} + 2(p_{ABA}p_{ABB})^{1/2}\cos\Theta, \tag{3.11}$$

where the phase difference Θ is given by

$$\begin{aligned}\Theta = \hbar^{-1}\Big[&E^{ABA}_x\Big[\int_r^t L(X^{AB}_s,DX^{AB}_s,D_*X^{AB}_s)\,ds\Big]\\ &- E^{ABA}_{x-a}\Big[\int_r^t L(X^{AB}_s,DX^{AB}_s,D_*X^{AB}_s)ds\Big]\Big].\end{aligned}\tag{3.12}$$

This gives the quantum mechanical interference. The phase difference Θ can be manipulated by the formula (2.13) and we obtain

$$\Theta = \tfrac{1}{2}\frac{m}{\hbar}(b(x,t) + b_*(x,t))\cdot a + o(a). \tag{3.13}$$

The present sample path analysis of the Nelson process thus shows that the quantum mechanical interference appears when the slits are both open. It tells us also the interesting fact that in any case the particle certainly goes through either the slit A or B, since the sample paths of the Nelson process all go through either A or B.

C. Gravitational effect in quantum mechanics

In classical mechanics the equivalence principle asserts that the dynamics of a particle is of geometric nature and does not depend on the mass. Indeed the classical F trajectory for a gravitational force $F = -m\nabla V_G$ satisfies Newton's equation of motion

$$\ddot{x} = -\nabla V_G(x), \tag{3.14}$$

in which the mass parameter does not appear. However, in quantum mechanics, the Schrödinger equation (2.4) contains essentially the mass parameter even if the gravitational potential mV_G is involved. Therefore it might be supposed that the quantum dynamics of the particle in the gravitational field is no longer a purely geometric one but depends on the mass of the particle. We will soon see, contrary to this supposition, that the quantum dynamics in the gravitational field is also of geometric nature as is the classical dynamics.

Let us consider two independent particles with different masses m and m'. Let X and X' be the Nelson processes in the class $\mathrm{N}(\mathscr{S})$ representing the quantum dynamics of those particles, respectively. Since the Euler–Nelson equations for X and X' are of the same form,

$$\tfrac{1}{2}(DD_*X + D_*DX) = -\nabla V_G(X), \tag{3.15}$$

we conclude that their sample path behaviors are identical and so their dynamics are. As the magnitude of their quadratic variations are different, i.e., $\hbar/2m$ and $\hbar/2m'$, however, the probability measures defined on the sample paths are different. This fact results in having the mass parameter in the Schrödinger equation. Notice that the mass enters in the Schrödinger equation only through a combination $(\hbar/m)$ in the case of gravitation. Even in quantum mechanics, the dynamics of the particle in the gravitational field is of geometric nature, but the quantum fluctuation, that is, the path probability measure, is not.

[1] E. Nelson, Phys. Rev. **150**, 1079 (1966).

[2] E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, 1967).

[3] E. Nelson, Bull. Am. Math. Soc. **84**, 121 (1978).

[4] E. Nelson, "Connection between Brownian Motion and Quantum Mechanics," talk, Einstein Symposium, Berlin (1979).

[5] K. Yasue, Ann. Phys. N. Y. **114**, 479 (1978).

[6] K. Yasue, J. Math. Phys. **19**, 1892 (1978).

[7] K. Yasue, Phys. Rev. Lett. **40**, 665 (1978).

[8] K. Yasue, Phys. Rev. D **18**, 532 (1978).

[9] E. Etim, Nuovo Cimento **51** A, 405 (1979).

[10] E. Etim, "Stochastic Quantization on a Riemannian Manifold," edited by A. Blaquière, F. Fer, and A. Marzollo in *Dynamical Systems and Microphysics*, (Springer, New York, 1980).

[11] E. Etim, "A Stochastic Description of Tunneling in Quantum Mechanics," in *Functional Integration*, edited by J.-P. Antoine and E. Tirapegui (Plenum, New York, 1980).

[12] G. Jona-Lasinio, "Stochastic Dynamics and the Semiclassical Limit of Quantum Mechanics, "*Talk, Bielefeld Encounters in Physics and Mathematics II* (Bielefeld, 1978).

[13] F. Guerra and P. Ruggiero, Phys. Rev. Lett. **31**, 1022 (1973).

[14] S. Albeverio and R. Høegh-Krohn, J. Math. Phys. **15**, 1745 (1974).

[15] T. Dankel, Jr., J. Math. Phys. **18**, 253 (1977).

[16] K. Yasue, Intern. J. Theor. Phys. **18**, 861 (1979).

[17] W. J. Lehr and J. L. Park, J. Math. Phys. **18**, 1235 (1977).

[18] F. Guerra and P. Ruggiero, Lett. Nuovo Cimento **23**, 529 (1978).

[19] T. Dankel, Jr., Arch. Rat. Mech. Anal. **37**, 192 (1970).

[20] D. S. Shucker, J. Funct. Anal. **38**, 146 (1980).

[21] G. C. Ghirardi, C. Omero, A. Rimini, and T. Weber, Riv. Nuovo Cimento **1**, no. 1 (1978).

[22] D. Dohrn and F. Guerra, Lett. Nuovo Cimento **22**, 121 (1978).

[23] K. Yasue, J. Math. Phys. **22**, 1010 (1981).

[24] R. Carmona, "Processus de Diffusion Gouverné par la Forme de Dirichlet de l'Operateur de Schrödinger," *Lecture Notes in Mathematics 721* (Springer, Berlin, 1978).

[25] M. Nagasawa, J. Math. Biol. **9**, 213 (1980).

[26] A. Blaquière, J. Optim. Theory Appl. **27**, 71 (1979).

[27] A. Blaquière, "Wave Mechanics as a Two-Player Game," in *Dynamical Systems and Microphysics*, edited by A. Blaquière, F. Fer, and A. Marzollo (Springer-Verlag, New York, 1980).

[28] K. Itô, "Extension of Stochastic Integrals," in *Proc. Intern. Symp. SDE, Kyoto 1976*, edited by K. Itô (Wiley, New York, 1978).

[29] K. Yasue, "Stochastic Calculus of Variations," J. Funct. Anal. **41**, 327 (1981).

[30] K. Yasue, Lett. Math. Phys. **4**, 357 (1980).

[31] J.-C. Zambrini, Lett. Math. Phys. **4**, 457 (1980).

[32] S. Albeverio, M. Fukushima, W. Karwowski, and L. Streit, "Capacity and Quantum Mechanical Tunneling," preprint, Universität Bielefeld (September 1980).

On Jacobi's decomposition of the motion of a heavy symmetrical top into the motions of two free triaxial tops

K. Yamada and S. Shieh
Department of Physics and Astronomy, The University of Tennessee, Knoxville, Tennessee 37996-1200

(Received 1 March 1982; accepted for publication 23 April 1982)

Jacobi discovered that the motion of a heavy symmetrical top can be decomposed into the motions of two torque-free triaxial tops. In this paper we investigate the connection between the three sets of the dynamical constants in the three top motions. The formulas connecting these constants are found to be projective transformations (fractional linear transformations).

PACS numbers: 03.20. + i

1. INTRODUCTION

The position of a rotating rigid body with one point fixed is completely determined by three independent variables, for example, three Euler's angles. Therefore, once the Euler's angles as functions of time are found, the further computation of the orthogonal matrix whose nine elements are the direction cosines of the moving body axes with respect to the fixed space axes, may seem to be an unnecessary tedious work without yielding any further new information in it. However, Jacobi found that the orthogonal matrix M, which describes the motion of a heavy symmetrical top can be decomposed into two orthogonal matrices M_1 and M_2, each of which represents the motion of a torque-free triaxial top.[1,2] Jacobi expressed the matrix elements of all the three matrices explicitly in terms of the theta functions and employed the addition theorem of the theta functions in establishing the relation

$$M = M_1\widetilde{M}_2, \tag{1.1}$$

where $\widetilde{M}_2$ is the transpose of M_2. The physical meaning of Eq. (1.1) is that the motion of a heavy symmetrical top can be decomposed into two motions of two torque-free triaxial tops. Let us consider the motion of a heavy symmetrical top M with the principal axes (x_2, y_2, z_2) and the motion of a free triaxial top M_1 with the principal axes (x_1, y_1, z_1), both of which are described with respect to a frame of reference (x ,y, z) fixed in space. Then the motion of the torque-free triaxial top observed from the fixed space is described by the orthogonal matrix M_1 and the same motion when observed from the frame of the heavy symmetrical top (i.e., the observer moving with the heavy symmetrical top) is described by the orthogonal matrix M_2. The relative motion M_2 turns out to be also representing the motion of a torque-free triaxial top. But not only the dynamical parameters but also the kinematical parameters (e.g., moments of inertia) of the two triaxial tops underlying the motions M_1 and M_2 are different.

In Sec. 2 we will give a summary of those aspects of the rigid-body motions that are needed for our study. The precise description of the position of a rigid body at any instant requires three transcendental constants: the modulus of the elliptic functions, the time scaling factor, and the additive constant in the argument of the elliptic functions. Klein and Sommerfeld call these constants transcendental in contrast to elementary physical constants which represent the energy and angular momenta of the motion.[3] Among the three top motions M, M_1, M_2 involved in Jacobi's decomposition theorem, the equality of the three time-scaling factors and that of the three moduli are obvious because elliptic functions with different moduli represent completely different functions and the different time-scaling factors would make the concordant repetition of the periodic motions of the three top motions impossible. Jacobi found a simple addition relation (including sum and difference) for the third type of transcendental constants (the additive constants).

The purpose of our study is to find some simple relations between the three sets of physical constants of the three top motions. The possibility of finding the simple relations is suggested by the following two observations: (1) The squared modulus k^2 of the elliptic functions is the cross ratio of the four physical constants and the constancy of the cross ratio under a projective transformation [fractional linear transformation $x' = (ax + b)/(cx + d)$][4] gives a hint that these three sets of physical constants may be connected by the projective transformation which will ensure the equality of the moduli of all the elliptic functions used in the description of the three top motions. (2) The formula for the addition theorem of the elliptic functions is fractional (the trigonometric functions are special elliptic functions with zero modulus and the denominator for the trigonometric addition theorem degenerates into unity).[5] The addition relation $p = q_1 \pm q_2$ discovered by Jacobi for the third type of transcendental constants when combined with the fractional addition theorem of the elliptic functions also hints at the projective relation between the three sets of physical constants.

We will show in Sec. 3 that the above two observations indeed lead to the remarkable result: the three sets of physical constants in Jacobi's decomposition theorem are connected to each other by simple projective transformations.

The relative magnitudes of the physical constants of a top motion must satisfy either increasing or decreasing sequential order in a particular way, and our result certainly meets this criterion which we see in Sec. 4.

2. THE PERTINENT RESULTS FROM THE RIGID-BODY DYNAMICS

A. A heavy symmetrical top

We denote the principal moments of inertia of a rigid body by (I_1, I_2, I_3). For the motion of a symmetrical top $(I_1 = I_2)$ with one point fixed under the action of the uniform

0022-2488/82/091584-05$02.50

gravitational force, we have

$$E = \tfrac{1}{2}I_2(\dot{\theta}^2 + \dot{\phi}^2 \sin^2\theta) + \tfrac{1}{2}I_3\omega_{z_2}^2 + Mgl\cos\theta, \tag{2.1}$$

where E is the total energy. θ, ϕ, and ψ are usual Euler's angles. ω_{z_2} is the angular velocity component along the axis of symmetry. M is the total mass and l is the length between the center of gravity and the fixed point. We rewrite (2.1) as

$$E' = E - \tfrac{1}{2}I_3\omega_{z_2}^2 = \tfrac{1}{2}I_1(\dot{\theta}^2 + \dot{\phi}^2\sin^2\theta) + Mgl\cos\theta. \tag{2.2}$$

Now we introduce

$$u = (2E'/I_1)^{1/2} \quad \text{and} \quad v = (2Mgl/I_1)^{1/2}, \tag{2.3}$$

where both u and v have the dimension of angular velocity. In terms of u and v, the three dimensionless physical constants of the motion of a heavy top will be

$$\alpha' = u/v. \tag{2.4}$$

a = the angular momentum component along the axis of symmetry divided by vI_1. (2.5)

b = the angular momentum component along the space z axis (antiparallel to the gravity) divided by vI_1. (2.6)

We will always divide any angular velocity by v to make it dimensionless. The expression for the dimensionless angular velocity $\dot{\theta}$ of the nutation is given by

$$\dot{\theta}^2\sin^2\theta = (1 - \cos^2\theta)(\alpha' - \cos\theta) - (b - a\cos\theta)^2. \tag{2.7}$$

We put $x = \cos\theta$. Then (2.7) becomes a cubic expression in x,

$$\dot{x}^2 = (1 - x^2)(\alpha' - x) - (b - ax)^2 = (x - x_1)(x - x_2)(x - x_3) = f(x), \tag{2.8}$$

where x_1, x_2, x_3 are the roots of the cubic equation $f(x) = 0$. It can be shown that the three roots are all real,

$$x_1 > 1 > x_2 > x_3 > -1, \tag{2.9}$$

and the physical range of $x(=\cos\theta)$ is

$$x_2 > x > x_3. \tag{2.10}$$

The cubic equation $f(x) = 0$ has an interesting structure in it: when a and b are interchanged in (2.8) and if α' is replaced by α which is given by

$$\alpha = a^2 - b^2 + \alpha', \tag{2.11}$$

then the three roots x_1 ,x_2 x_3 will still remain unchanged. In that case, Eq. (2.8) can be written

$$f(x) = (1 - x^2)(\alpha - x) - (a - bx)^2 = (x - x_1)(x - x_2)(x - x_3) \tag{2.12}$$

with the same set of the three roots. It turns out that this symmetric property of the cubic expression $f(x)$ plays an important role in our analysis on the decomposition theorem.

From either (2.8) or (2.12), by putting $x = \pm 1$, we obtain the following two expressions:

$$a - b = \pm[(x_1 - 1)(1 - x_2)(1 - x_3)]^{1/2}, \tag{2.13}$$

$$a + b = [(x_1 + 1)(1 + x_2)(1 + x_3)]^{1/2}. \tag{2.14}$$

Further, from (2.8) we obtain

$$b - a\alpha = \pm[(x_1 - \alpha')(\alpha' - x_2)(\alpha' - x_3)]^{1/2}, \tag{2.15}$$

$$ax_1 - b = \pm[(x_1^2 - 1)(x_1 - \alpha')]^{1/2}, \tag{2.16}$$

$$b - ax_2 = \pm[(1 - x_2^2)(\alpha' - x_2)]^{1/2}. \tag{2.17}$$

Also from (2.12), we get

$$a - b\alpha = \pm[(x_1 - \alpha)(\alpha - x_2)(\alpha - x_3)]^{1/2}, \tag{2.18}$$

$$bx_1 - a = \pm[(x_1^2 - 1)(x_1 - \alpha)]^{1/2}, \tag{2.19}$$

$$a - bx_2 = \pm[(1 - x_2^2)(\alpha - x_2)]^{1/2}. \tag{2.20}$$

Since all the quantities in the above expressions are real and we have the inequality condition (2.9), the expressions (2.16), (2.17), (2.19), and (2.20) will give

$$x_1 > \alpha' > x_2 > x_3, \qquad x_1 > \alpha > x_2 > x_3. \tag{2.21}$$

All the above expressions (2.13) through (2.21) will be needed later.

The period T of the heavy top motion is given by[1]

$$T = 2K/n, \tag{2.22}$$

when n, the time scaling factor, is defined by[1]

$$n = \tfrac{1}{2}[x_1 - x_3]^{1/2}. \tag{2.23}$$

K in (2.22) is the complete elliptic integral

$$K = \int_0^{\pi/2} \frac{dy}{(1 - k^2\sin^2 y)^{1/2}}, \tag{2.24}$$

where k is the modulus[1]

$$k^2 = (x_2 - x_3)/(x_1 - x_3). \tag{2.25}$$

In order to express the Euler's angles ϕ and ψ as explicit functions of time, Jacobi introduced the two constants q_1 and q_2 which are expressed in terms of the three roots through the elliptic function sn[1]

$$\operatorname{sn}(iq_1) = i\left[\frac{x_1 - 1}{1 - x_2}\right]^{1/2}, \quad \operatorname{sn}(iq_2) = i\left[\frac{x_1 - x_3}{1 + x_3}\right]^{1/2}. \tag{2.26}$$

Since $\mathrm{cn}^2 = 1 - \mathrm{sn}^2$ and $\mathrm{dn}^2 = 1 - k^2\,\mathrm{sn}^2$, (2.26) gives

$$\operatorname{cn}(iq_1) = \left[\frac{x_1 - x_2}{1 - x_2}\right]^{1/2}, \quad \operatorname{cn}(iq_2) = \left[\frac{1 + x_1}{1 + x_3}\right]^{1/2}, \tag{2.27}$$

$$\operatorname{dn}(iq_1) = \left[\frac{(x_1 - x_2)(1 - x_3)}{(x_1 - x_3)(1 - x_2)}\right]^{1/2},$$

$$\operatorname{dn}(iq_2) = \left[\frac{1 + x_2}{1 + x_3}\right]^{1/2}, \tag{2.28}$$

where (2.25) has been used for (2.28). In accordance with Klein and Sommerfeld, we call the constants n (the time-scaling factor), k (the modulus), $q_1$and q_2 introduced above "the transcendental constants,"[3] while the constants a, b, $\alpha(\alpha')$, x_1, x_2, and x_3 all of which appear in the cubic expressions (2.8) and (2.12) we call "the physical constants."

B. A torque-free triaxial top

We denote the principal moments of inertia of the first free triaxial top by (A_1, B_1, C_1). Then the four physical constants of the motion of the torque-free triaxial top will be[6]

$$\frac{l_1}{A_1}, \quad \frac{l_1}{B_1}, \quad \frac{l_1}{C_1}, \quad \frac{h_1}{l_1}, \tag{2.29}$$

where l_1 is the magnitude of the total angular momentum and h_1 is twice the total energy (the kinetic energy). Each of these has the dimension of angular velocity. In the decomposition theorem, since the motion of a heavy top is connected with the motions of two free tops, we divide each of the above four constants by v which is defined by (2.3) to make it dimensionless as we did in the case of the heavy top motion. Therefore it is always understood that the four physical constants in (2.29) have already been divided by v and they are dimensionless. For the sake of conveniency, we write the four physical constants as

$$s_0 = \frac{h_1}{l_1}, \quad s_1 = \frac{l_1}{C_1}, \quad s_2 = \frac{l_1}{A_1}, \quad s_3 = \frac{l_1}{B_1}. \tag{2.30}$$

The period T_1 of the motion is given by[6]

$$T_1 = 2K_1/n_1, \tag{2.31}$$

where n_1, the time-scaling factor, is defined by[6]

$$n_1 = [(s_0 - s_2)(s_1 - s_3)]^{1/2}. \tag{2.32}$$

K_1 in (2.31) is the complete elliptic integral

$$K_1 = \int_0^{\pi/2} \frac{dy}{(1 - k_1^2 \sin^2 y)^{1/2}}. \tag{2.33}$$

k_1 is the modulus given by[6]

$$k_1^2 = \frac{(s_3 - s_2)(s_1 - s_0)}{(s_0 - s_2)(s_1 - s_3)}. \tag{2.34}$$

Since two torque-free triaxial tops are involved in the decomposition theorem, we write the four physical constants for the motion of the second free top as

$$s_0' = \frac{h_2}{l_2}, \quad s_1' = \frac{l_2}{C_2}, \quad s_2' = \frac{l_2}{A_2}, \quad s_3' = \frac{l_2}{B_2}. \tag{2.35}$$

Like the case of the first free-top motion, the period T_2, the time-scaling factor n_2, the complete elliptic integral K_2, and the modulus k_2 are given by

$$T_2 = \frac{2K_2}{n_2}, \tag{2.36}$$

$$n_2 = [(s_0' - s_2')(s_1' - s_3')]^{1/2}, \tag{2.37}$$

$$k_2^2 = \frac{(s_3' - s_2')(s_1' - s_0')}{(s_0' - s_2')(s_1' - s_3')}. \tag{2.38}$$

In order to express the Euler's angle ϕ_1 of the first free-top motion as an explicit function of time, Jacobi[6] introduced a constant p_1

$$\operatorname{sn}(ip_1) = i\left[\frac{s_0 - s_2}{s_1 - s_0}\right]^{1/2}, \tag{2.39}$$

whence we get

$$\operatorname{cn}(ip_1) = \left[\frac{s_1 - s_2}{s_1 - s_0}\right]^{1/2}, \tag{2.40}$$

$$\operatorname{dn}(ip_1) = \left[\frac{s_1 - s_2}{s_1 - s_3}\right]^{1/2}, \tag{2.41}$$

where (2.34) has been used for (2.41).

Similarly, for the second free-top motion, a constant p_2 is defined by

$$\operatorname{sn}(ip_2) = i\left[\frac{s_0' - s_2'}{s_1' - s_0'}\right]^{1/2}, \tag{2.42}$$

whence we have

$$\operatorname{cn}(ip_2) = \left[\frac{s_1' - s_2'}{s_1' - s_0'}\right]^{1/2}, \tag{2.43}$$

$$\operatorname{dn}(ip_2) = \left[\frac{s_1' - s_2'}{s_1' - s_3'}\right]^{1/2}. \tag{2.44}$$

3. DERIVATION OF THE PROJECTIVE TRANSFORMATIONS CONNECTING THE THREE SETS OF PHYSICAL CONSTANTS

Our goal is to find the connecting formulas between the three sets of physical constants (x_1, x_2, x_3), (s_0, s_1, s_2, s_3), and (s_0', s_1', s_2', s_3').

Because of the concordance of the periodic motions of the three tops, the periods of the three motions must be the same and further the nature of all the elliptic functions describing the three top motions must also be the same, that is, the three moduli must be equal. Then we must have

$$k = k_1 = k_2, \tag{3.1}$$

$$n = n_1 = n_2. \tag{3.2}$$

The expressions of the three squared moduli,

$$k^2 = \frac{x_2 - x_3}{x_1 - x_3},$$

$$k_1^2 = \frac{(s_3 - s_2)(s_1 - s_0)}{(s_0 - s_2)(s_1 - s_3)},$$

$$k_2^2 = \frac{(s_3' - s_2')(s_1' - s_0')}{(s_0' - s_2')(s_1' - s_3')},$$

indicate that they are cross ratios of (x_1, x_2, x_3), (s_0, s_1, s_2, s_3), and (s_0', s_1', s_2', s_3'). There is a theorem which states that the cross ratio of four numbers is invariant under a projective transformation. Therefore, we try the projective transformation[4]

$$s = \frac{Ax + B}{Cx + D}, \tag{3.3}$$

where A, B, C, D are constants. The transformation (3.3) carries the four numbers (∞, x_1, x_2, x_3) into the four numbers (s_0, s_1, s_2, s_3) and automatically ensures the equality of the two squared moduli $k^2 = k_1^2$. Since $x = \infty$ is transformed into s_0, (3.3) can be rewritten as

$$s - s_0 = \frac{BC - AD}{C(Cx + D)}, \quad s_0 = \frac{A}{C}. \tag{3.4}$$

Now from (2.32), (2.34), and (2.39)–(2.41), we obtain[6]

$$\begin{aligned} s_1 - s_0 &= \pm n_1 \frac{i\,\operatorname{dn}(ip_1)}{\operatorname{sn}(ip_1)\operatorname{cn}(ip_1)}, \\ s_2 - s_0 &= \pm n_1 \frac{\operatorname{sn}(ip_1)\operatorname{dn}(ip_1)}{i\,\operatorname{cn}(ip_1)}, \\ s_3 - s_0 &= \pm n_1 \frac{k_1'^2 \operatorname{sn}(ip_1)}{i\,\operatorname{cn}(ip_1)\operatorname{dn}(ip_1)}, \end{aligned} \tag{3.5}$$

where $k_1'^2 = 1 - k_1^2$.

Jacobi found the addition relation between the transcendental constants p_1, q_1, q_2 all of which we have introduced in Sec. 2. It is[1]

$$p_1 = q_1 + q_2. \tag{3.6}$$

We apply the addition theorem of the elliptic function sn to (3.6)[5]:

$$\mathrm{sn}(ip_1) = \mathrm{sn}(iq_1 + iq_2) = \frac{\mathrm{sn}(iq_1)\mathrm{cn}(iq_2)\mathrm{dn}(iq_2) + \mathrm{sn}(iq_2)\mathrm{cn}(iq_1)\mathrm{dn}(iq_1)}{1 - k^2\,\mathrm{sn}^2(iq_1)\mathrm{sn}^2(iq_2)}. \tag{3.7}$$

We substitute (2.26) through (2.28) into (3.7) to obtain

$$\begin{aligned}\mathrm{sn}(ip_1) &= i\,\frac{[(x_1^2-1)(1-x_2^2)]^{1/2} + (x_1-x_2)(1-x_3^2)^{1/2}}{1 - x_1x_2 + x_1x_3 - x_2x_3}\\ &= i\,\frac{[(x_1-1)(1+x_2)(1+x_3)]^{1/2} + [(x_1+1)(1-x_2)(1-x_3)]^{1/2}}{[(x_1+1)(1-x_2)(1+x_3)]^{1/2} - [(x_1-1)(1+x_2)(1-x_3)]^{1/2}}\\ &= i\,\frac{[(x_1-1)/(x_1+1)]^{1/2}(a+b) + [(x_1+1)/(x_1-1)]^{1/2}(a-b)}{[(1-x_2)/(1+x_2)]^{1/2}(a+b) - [(1+x_2)/(1-x_2)]^{1/2}(a-b)}\\ &= i\left(\frac{1-x_2^2}{x_1^2-1}\right)^{1/2}\frac{bx_1-a}{a-bx_2}\end{aligned} \tag{3.8}$$

in which we have used (2.13) and (2.14). Substituting (2.19) and (2.20) into (3.8), we obtain

$$\mathrm{sn}(ip_1) = i\left[\frac{x_1-\alpha}{\alpha-x_2}\right]^{1/2}. \tag{3.9}$$

Since $\mathrm{cn}^2 = 1 - \mathrm{sn}^2$, $\mathrm{dn}^2 = 1 - k^2\,\mathrm{sn}^2$, and $k^2 = (x_2 - x_3)/(x_1 - x_3)$, (3.9) leads to

$$\mathrm{cn}(ip_1) = \left[\frac{x_1-x_2}{\alpha-x_2}\right]^{1/2}, \tag{3.10}$$

$$\mathrm{dn}(ip_1) = \left[\frac{(x_1-x_2)(\alpha-x_3)}{(x_1-x_3)(\alpha-x_2)}\right]^{1/2}. \tag{3.11}$$

From (2.23) and (3.2), we have

$$n_1 = n = \tfrac{1}{2}[x_1 - x_3]^{1/2}. \tag{3.12}$$

From (2.25) and (3.1), we have

$$k_1'^2 = k'^2 = 1 - k^2 = \frac{x_1-x_2}{x_1-x_3}. \tag{3.13}$$

Now we substitute (3.9)–(3.13) into (3.5) to obtain

$$s_i - s_0 = \frac{a-b\alpha}{2(\alpha-x_i)}, \quad i = 1,2,3, \tag{3.14}$$

where we have used (2.18). Comparing (3.14) with (3.4), we can put

$$\frac{BC-AD}{C(Cx+D)} = \frac{a-b\alpha}{2(\alpha-x)}, \quad s_0 = \frac{A}{C}.$$

Since this relation must hold for any x, we obtain

$$s_0 = A/C = b/2. \tag{3.15}$$

Thus (3.14) can be written

$$s_i = \frac{b}{2} + \frac{a-b\alpha}{2(\alpha-x_i)} = \frac{a-bx_i}{2(\alpha-x_i)}, \quad i = 0,1,2,3, \tag{3.16}$$

where $x_0 = \infty$ and it is transformed into $s_0 = b/2$. So the transformation formula between the two sets of four numbers s_i and x_i ($x_0 = \infty$) can be written, dropping the subscript i,

$$s = (a - bx)/2(\alpha - x), \tag{3.17}$$

which has the same form as (3.3) and we have obtained the desired projective transformation between s_i and x_i with $i = 0,1,2,3$.

For the second free top motion, that is, the relation between the physical constants s_i' and x_i, starting with the relation $p_2 = q_1 - q_2$, using (2.15)–(2.17) and employing the same method used for the derivations of (3.14) and (3.17), we obtain

$$s_i' - s_0' = \frac{b-a\alpha'}{2(\alpha'-x_i)}, \tag{3.18}$$

$$s_0' = \frac{a}{2}, \tag{3.19}$$

$$s' = \frac{b-ax}{2(\alpha'-x)}. \tag{3.20}$$

$x_0 = \infty$ is carried into s_0' by (3.20). If we recall (2.30) and (2.35), the above results (3.17) and (3.20) can be written

$$\begin{aligned}\frac{l_1}{A_1} &= \frac{a-bx_2}{2(\alpha-x_2)}, & \frac{l_2}{A_2} &= \frac{b-ax_2}{2(\alpha'-x_2)},\\ \frac{l_1}{B_1} &= \frac{a-bx_3}{2(\alpha-x_3)}, & \frac{l_2}{B_2} &= \frac{b-ax_3}{2(\alpha'-x_3)},\\ \frac{l_1}{C_1} &= \frac{a-bx_1}{2(\alpha-x_1)}, & \frac{l_2}{C_2} &= \frac{b-ax_1}{2(\alpha'-x_1)},\\ \frac{h_1}{l_1} &= \frac{b}{2}, & \frac{h_2}{l_2} &= \frac{a}{2}.\end{aligned}$$

Thus we have found that the two sets of physical constants of the two free-top motions are connected to the physical constants of the heavy top motion in the form of projective transformation.

4. DISCUSSION

In the motion of a free triaxial top, if we denote the principal moments of inertia by A, B, C, twice the kinetic energy by h, and the magnitude of the total angular momentum by l, it can be shown that for any possible motion of the top, the parameter h/l must lie between l/A and l/C for $A > B > C$ or $A < B < C$, otherwise no physical motion is possible.[6] Observing (3.14) and (3.18), we note that all the numerators in each set of the expressions are the same. Then if we use the inequalities in (2.21), we can easily show $s_2 > s_3 > s_0 > s_1$ for $a - b\alpha > 0$ and the opposite order of sequence for $a - b\alpha < 0$. $s_2' > s_3' > s_0' > s_1'$ for $b - a\alpha' > 0$ and the opposite order of sequence for $b - a\alpha' < 0$. If we use (2.30) and (2.35), the above argument can be written

$$\frac{l_1}{A_1} > \frac{l_1}{B_1} > \frac{h_1}{l_1} > \frac{l_1}{C_1} \quad \text{or the opposite sequence,}$$
$$\frac{l_2}{A_2} > \frac{l_2}{B_2} > \frac{h_2}{l_2} > \frac{l_2}{C_2} \quad \text{or the opposite sequence.} \tag{4.1}$$

This result (4.1) certainly meets the criterion mentioned earlier.

[1] C. G. J. Jacobi, *Werke* (Reimer, Berlin, 1882), Vol. 2, p. 493.

[2] W. D. MacMillan, *Dynamics of Rigid Bodies* (Dover, New York, 1960), Chap. VII, Sec. 117.

[3] F. Klein and A. Sommerfeld, *Über die Theorie des Kreisels* (Teubner, Leipzig, 1897), p. 428.

[4] R. A. Silverman, *Introductory Complex Analysis* (Dover, New York, 1972), Chap. 5, Sec. 25.

[5] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge U. P., Cambridge, 1943), p. 494.

[6] Reference 1, pp. 306–314 and pp. 427–430.

Vector fields generating invariants for classical dissipative systems[a]

F. Cantrijn[b),c)]
Department of Physics, Northeastern University, Boston, Massachusetts 02115

(Received 11 May 1981; accepted for publication 26 June 1981)

A class of vector fields is identified which (locally) generate first integrals of a dissipative system. The structure of these vector fields and of the corresponding invariants is studied. The relationship with a previously proposed generalization of Noether's theorem for nonconservative systems, is pointed out.

PACS numbers: 03.20. + i, 02.40. + m

I. INTRODUCTION

In this paper we will study a relationship between a class of vector fields, defined on the extended state space of a classical dissipative system, and the set of (local) first integrals of that system.

In calling a system dissipative (or nonconservative) we refer to the property that, in terms of a given Lagrangian $\mathscr{L}$, the equations of motion are of the form

$$\frac{d}{dt}\left(\frac{\partial \mathscr{L}}{\partial \dot{q}^i}\right) - \frac{\partial \mathscr{L}}{\partial q^i} = Q_i, \quad i = 1,...,n, \tag{1}$$

with Q_i some functions which may depend on all variables t, $q^i, \dot{q}^i$. As is well known,[1] such a description does not necessarily prevent the system from admitting a purely Lagrangian description in terms of some other Lagrangian $\mathscr{L}'$. The notion of dissipativity should therefore always be interpreted relative to some given Lagrangian.

Recently, several papers[2-5] have been devoted to the problem of finding transformations which, in some way or another, generate an invariant of a dissipative system. A general approach, based on d'Alembert's differential variational principle, has been presented in Ref. 2. More closely related to the present paper is a generalized version of Noether's theorem and its converse, which applies to general nonconservative systems (see Ref. 3). This result has been further extended to generalized mechanical systems (with $\mathscr{L}$ and Q_i depending on higher-order derivatives),[4] whereas the field-theoretical case has been treated in Ref. 5.

The method described in this paper arises mainly from the geometrical description of a dissipative system in terms of a nonclosed two-form of maximal rank. The existence of such a two-form immediately allows for the identification of a class of vector fields which generate first integrals of the given system in an unambiguous way. To a certain extent the present treatment will closely resemble the discussion of Noether symmetries of a Lagrangian system (with Lagrangian $\mathscr{L}$), regarded as symmetries of an exact contact structure $d\theta$, where θ represents the so-called Cartan-form associated with $\mathscr{L}$:

$$\theta = \mathscr{L}dt + \frac{\partial \mathscr{L}}{\partial \dot{q}^i}(dq^i - \dot{q}^i dt), \tag{2}$$

(see, e.g., Refs. 6 and 7). Unlike the case of Noether symmetry vector fields, however, the vector fields associated with invariants of a dissipative system in general fail to be dynamical symmetries of that system.

In Sec. 2 we recall some general concepts which we shall adopt in this paper concerning the description of a dissipative system. In Sec. 3, we exhibit a set $\mathscr{V}$ of vector fields on the extended state space, which can be related in a very precise way to the first integrals of the system under consideration. It is also shown that $\mathscr{V}$ remains invariant under a class of dynamical symmetries. Section 4 is devoted to the structure of vector fields $Y \in \mathscr{V}$, and a system of partial differential equations for the components of Y is derived. In Sec. 5, we analyze the structure of an invariant of a dissipative system in terms of the components of the corresponding generating vector field. Noether's theorem for nonconservative systems, as presented in Ref. 3, is briefly reviewed in Sec. 6 and its connection with the present approach is clarified. Before concluding with a few general remarks we discuss, in Sec. 7, a special class of dissipative systems for which one can immediately associate a dynamical symmetry with each $Y \in \mathscr{V}$.

The following notations will be used. The sets of smooth functions, vector fields and p-forms on a differentiable manifold N are denoted by $C^\infty(N)$, $\mathscr{X}(N)$ and $\Omega^p(N)$, respectively. The Lie derivative of a p-form β with respect to a vector field X is denoted by $L_X\beta$, while the inner product of X and β is written as $i_X\beta$ or, in case β is a one-form, as $\langle X,\beta \rangle$. Finally, for the Lie derivative of a function f with respect to X we also frequently use the notation $X(f)$.

II. DISSIPATIVE SYSTEMS

Let M be a real n-dimensional differentiable manifold, representing the configuration space of a mechanical system. Local coordinates on M will be denoted by $(q^1,...,q^n)$. Since we will primarily be dealing with time-dependent systems, it is convenient to introduce the extended state space $N = \mathbb{R} \times TM$. The natural coordinates on N are $(t, q^i, \dot{q}^i)$.

A dissipative system will be characterized by a function $\mathscr{L} \in C^\infty(N)$, the Lagrangian of the system, and a semi-basic one-form $\bar{\mu}$ on TM, which may be time dependent. While a general discussion on semi-basic forms can be found, e.g., in Ref. 8, it suffices for our purpose to point out that $\bar{\mu}$ will locally be of the form

[a] Work supported bv a NATO Research Fellowship Grant.

[b] Permanent address: Ryksuniversiteit Gent, Instituut voor Theoretische Mechanika, Krijgslaan 271-S9, B-9000 Gent, Belgium.

[c] Senior Research Assistant at the National Fund for Scientific Research (Belgium).

$$\tilde{\mu} = Q_i dq^i, \tag{3}$$

i.e., without terms in $d\dot{q}^i$, but with the functions Q_i depending on the $2n + 1$ variables $(t, q^j, \dot{q}^j)$. Regarding $\tilde{\mu}$ as a one-form on N in a natural way, we now introduce the following two-form on N:

$$\alpha = d\theta + \tilde{\mu} \wedge dt \tag{4}$$

with θ the Cartan-form (2) associated with $\mathscr{L}$. For convenience, we henceforth put $\mu = \tilde{\mu} \wedge dt$.

The further discussion will be restricted to those cases in which $\mathscr{L}$ is regular, i.e., the Hessian matrix $(\partial^2 \mathscr{L} / \partial \dot{q}^i \partial \dot{q}^j)$ is nowhere singular. It then easily follows that α is a two-form of maximal rank, namely $2n$. In particular, Ker $\alpha(p) = \{v \in T_p N : i_v \alpha(p) = 0\}$ defines a one-dimensional subspace of the tangent space $T_p N$ at $p \in N$. We furthermore assume that Ker $\alpha (= \bigcup_{p\in N}$ Ker $\alpha(p))$, as a line bundle over N, is trivial. This means that there exists a nowhere-vanishing vector field $X \in \mathscr{X}(N)$ which generates Ker $\alpha(p)$ at each $p \in N$. A straightforward computation (in local coordinates) reveals that $\langle X, dt \rangle \neq 0$. This in turn guarantees the existence of a vector field $\Delta \in \mathscr{X}(N)$ which is completely determined by

$$i_\Delta \alpha = 0, \tag{5a}$$

$$\langle \Delta, dt \rangle = 1. \tag{5b}$$

Using the local expressions (2) and (3) for θ and $\tilde{\mu}$, respectively, it can easily be verified that Δ is a second-order vector field which, in terms of the natural coordinates on N, is given by

$$\Delta = \frac{\partial}{\partial t} + \dot{q}^i \frac{\partial}{\partial q^i} + \Lambda^i \frac{\partial}{\partial \dot{q}^i}, \tag{6}$$

where the functions Λ^i are uniquely specified by the identities

$$\frac{\partial^2 \mathscr{L}}{\partial \dot{q}^i \partial \dot{q}^j} \Lambda^j + \frac{\partial^2 \mathscr{L}}{\partial \dot{q}^i \partial q^j} \dot{q}^j + \frac{\partial^2 \mathscr{L}}{\partial \dot{q}^i \partial t} - \frac{\partial \mathscr{L}}{\partial q^i} \equiv Q_i, \qquad i = 1,\ldots,n. \tag{7}$$

Consequently, the vector field Δ defined by (5a and 5b) is precisely the dynamical field associated with a dissipative system whose equations of motion are locally given by (1).

Before proceeding, we notice the following. Suppose the one-form $\tilde{\mu}$ is such that $d\mu \equiv d(\tilde{\mu} \wedge dt) = 0$. Poincare's lemma[9] then assures us that μ is locally exact. Since $\tilde{\mu}$ is a semibasic form, this leads to the following local expression for μ:

$$\mu = -d(V \circ \pi) \wedge dt,$$

where V is some function defined on (an open subset of) $\mathbb{R} \times M$ and $\pi: \mathbb{R} \times TM \to \mathbb{R} \times M$ denotes the natural projection. Hence, in this case the substitution $\mathscr{L} \to \mathscr{L}' = \mathscr{L} - V \circ \pi$ immediately provides us with a purely Lagrangian description of the system in terms of $\mathscr{L}'$ (i.e., without dissipative terms). Unless explicitly stated otherwise we henceforth restrict our attention to those cases whereby $d\alpha = d\mu \neq 0$, in this way narrowing somewhat the notion of a dissipative system. If $\mathscr{L}$ represents the usual physical Lagrangian of a mechanical system (i.e., $\mathscr{L}$ = kinetic energy − potential energy), the functions Q_i can then be interpreted as the components of forces that are not derivable from a potential.

In order to fix the terminology we now introduce a few definitions.

Definition 1: $Y \in \mathscr{X}(N)$ is a dynamical symmetry of Δ iff $[Y, \Delta] = h\Delta$ for some function $h \in C^\infty(N)$. Dynamical symmetries map integral curves of Δ into integral curves, up to a possible change of parametrization.

Definition 2: $Y \in \mathscr{X}(N)$ is a trivial symmetry of Δ iff $Y = h\Delta$ for some $h \in C^\infty(N)$. The set of trivial symmetries is precisely the subset of dynamical symmetries which consists of all smooth sections of Ker α.

Definition 3: $Y \in \mathscr{X}(N)$ is a conformal symmetry of α iff $L_Y \alpha = k\alpha$ for some $k \in C^\infty(N)$. In particular, Y is called a symmetry of α if $L_Y \alpha = 0$.

From these definitions, the definition (5) of Δ and the observation that all smooth sections of Ker α form a one-dimensional module over $C^\infty(N)$, we immediately derive:

Corollary 2.1: $Y \in \mathscr{X}(N)$ is a dynamical symmetry of Δ iff $i_{[Y,\Delta]}\alpha = 0$.

Corollary 2.2: Each conformal symmetry of α is a dynamical symmetry of Δ. (The proof of Corollary 2.2 is based on the formula[9] $i_{[Y,\Delta]}\alpha = L_Y i_\Delta \alpha - i_\Delta L_Y \alpha$.)

To close this section we now give some useful expressions for certain differential forms on N, in terms of a specific basis of one-forms.

Given a dissipative system with Lagrangian $\mathscr{L}$ and semibasic one-form $\tilde{\mu}$, a basis for the set of one-forms on a local coordinate neighborhood $U \subset N$ is provided by the forms:

$$dt,\ dq^i - \dot{q}^i dt,\ d\dot{q}^i - \Lambda^i dt, \quad (i = 1,\ldots,n) \tag{8}$$

whereby the functions Λ^i are defined by (7). For any function $F \in C^\infty(U)$ we then have

$$dF = \Delta(F) dt + \frac{\partial F}{\partial q^i}(dq^i - \dot{q}^i dt) + \frac{\partial F}{\partial \dot{q}^i}(d\dot{q}^i - \Lambda^i dt), \tag{9}$$

and the local expressions for μ and α become

$$\mu = Q_i (dq^i - \dot{q}^i dt) \wedge dt, \tag{10a}$$

$$\alpha = \frac{\partial^2 \mathscr{L}}{\partial \dot{q}^i \partial q^j}(dq^j - \dot{q}^j dt) \wedge (dq^i - \dot{q}^i dt) + \frac{\partial^2 \mathscr{L}}{\partial \dot{q}^i \partial \dot{q}^j}(d\dot{q}^i - \Lambda^i dt) \wedge (dq^j - \dot{q}^j dt). \tag{10b}$$

III. VECTOR FIELDS GENERATING FIRST INTEGRALS

For a given dissipative system with associated two-form α, we consider the set of vector fields Y defined on N, or on some open subset of N, which satisfy the condition

$$d(i_Y \alpha) = 0. \tag{11}$$

This set will be denoted by $\mathscr{V}$. Clearly, $\mathscr{V}$ contains all trivial symmetries of Δ. For each open subset $U \subset N$ (regarded as a submanifold of N) the corresponding subset of $\mathscr{V}$, consisting of those vector fields which are defined on U and satisfy (11), admits the structure of a linear space (over $\mathbb{R}$). However, due to the fact that $d\alpha \neq 0$, $\mathscr{V}$ will not be closed under the Lie bracket, i.e., the vector fields in $\mathscr{V}$ do not form a subalgebra of $\mathscr{X}(N)$. Taking into account the definition (4) of α, we find that $Y \in \mathscr{V}$ iff

$$L_Y\alpha = i_Y d\mu. \tag{12}$$

If μ were closed we would recover here the characterization of Noether symmetry vector fields for Lagrangian systems.[6,7] The following result is now straightforward:

Proposition 3.1: For each $Y \in \mathscr{Y}$ there exists, at least locally, a first integral G of Δ which is related to Y according to[10]

$$i_Y\alpha = dG. \tag{13}$$

Proof: By Poincare's lemma,[9] condition (11) implies that $i_Y\alpha$ is locally exact. Hence, a relation like (13) holds at least in a neighborhood of each point (where Y is defined). Using (5a) we then see that

$$0 = \langle \Delta, dG \rangle = \Delta(G). \qquad \square$$

Conversely, we also have

Proposition 3.2: For each first integral G of Δ, defined on some open subset $U \subset N$, there exists a vector field $Y \in \mathscr{X}(U)$ for which (13) holds (and hence, $Y \in \mathscr{Y}$). This result is an immediate consequence of the following lemma.

Lemma 3.3: Let β be a one-form defined on some open subset $U \subset N$. Then, $\langle \Delta, \beta \rangle = 0$ iff $\beta = i_Y\alpha$ for some $Y \in \mathscr{X}(U)$.

Proof: We prove the lemma for $U = N$. The sufficiency immediately follows from (5a).

To prove the necessity we first notice that α induces a vector bundle homomorphism $\rho: TN \rightarrow T^*N$, defined by $\rho(v) = i_v\alpha(p)$ for $v \in T_pN$. Simple algebraic considerations reveal that $\rho(T_PN) = \{\beta_p \in T_p^*N: \langle \Delta(p), \beta_p \rangle = 0\}$. It then follows that for a given $\beta \in \Omega^1(N)$, satisfying $\langle \Delta, \beta \rangle = 0$, the equation $i_Y\alpha = \beta$ admits a solution for Y in a neighborhood U_p of each point $p \in N$. In order to obtain a global solution, we use a partition of unity argument (see, e.g., Ref. 9, p. 122). Let $\{U_\lambda, f_\lambda\}$ be a partition of unity subordinate to the covering $\{U_p\}$ of N. In particular we have $\Sigma_\lambda f_\lambda(p) = 1$ at each point p. It follows from above that for each λ there exists a vector field $Y'_\lambda \in \mathscr{X}(U_\lambda)$ such that $i_{Y'_\lambda}(\alpha|_{U_\lambda}) = \beta|_{U_\lambda}$. Next, we define a vector field $Y_\lambda \in \mathscr{X}(N)$ by

$$Y_\lambda(p)\begin{cases} = f_\lambda(p)Y'_\lambda(p) & \text{for all } p \in U_\lambda, \\ = 0 & \text{for all } p \notin U_\lambda. \end{cases}$$

Putting $Y = \Sigma_\lambda Y_\lambda$, whereby the sum on the right-hand side reduces to a finite sum at each point, we finally have

$$i_Y\alpha = \sum_\lambda f_\lambda i_{Y'_\lambda}(\alpha|_{U_\lambda}) = \Big(\sum_\lambda f_\lambda \beta\Big) = \beta,$$

which completes the proof. $\square$

In order to prove Proposition 3.2, it now suffices to notice that G is a first integral of Δ iff $\langle \Delta, dG \rangle = 0$. A few remarks are in order here. First of all, it is clear from (13) that for a given $Y \in \mathscr{Y}$, the corresponding first integral G is locally determined up to a constant. Conversely, it follows from the maximal rank condition for α that for a given invariant G of Δ, the vector field Y satisfying (13) will be determined up to a trivial symmetry of Δ. Consequently, if we call two vector fields in $\mathscr{Y}$ equivalent if they differ by a multiple of Δ, it is seen that there exists, at least locally, a precise one-to-one correspondence between the resultant set of equivalence classes of vector fields in $\mathscr{Y}$ and the set of first integrals of Δ (whereby two first integrals are identified if they differ by a constant).

Moreover, it is interesting to notice that each invariant G of Δ is also an invariant of the corresponding vector field $Y \in \mathscr{Y}$. [This follows immediately from (13)].

Obviously, the invariants generated by trivial symmetries of Δ are constants. However, if $\mathscr{Y}$ contains a globally defined vector field $Y \in \mathscr{X}(N)$ which is not a trivial symmetry of Δ, and if moreover N would be such that its first cohomology class vanishes [i.e., $H^1(N) = 0$], then (13) would provide us with a global (nontrivial) first integral G. This follows from the fact that if $H^1(N) = 0$, then each closed one-form is globally exact.[11]

So far, the situation described here is quite analogous to the one encountered in the theory of Noether symmetries.[7] In the present case, however, the vector fields $Y \in \mathscr{Y}$ are in general not dynamical symmetries of Δ. This can be easily seen from

$$\begin{aligned} i_{[Y,\Delta]}\alpha &= L_Y i_\Delta \alpha - i_\Delta L_Y \alpha \\ &= -i_\Delta i_Y d\mu, \end{aligned} \tag{14}$$

whereby use has been made of (5a) and (12). Since the right-hand side does not vanish in general,[12] Corollary 2.1 tells us that Y can not be a dynamical symmetry of Δ.

A well-known property concerning symmetries and invariants of a dynamical system, is that the deformation L_ZG of an invariant G under a dynamical symmetry Z yields a new (not necessarily independent) invariant. This result, which is sometimes referred to as the related (first) integral theorem,[13] enters the present discussion as follows (taking into account Corollary 2.2):

Proposition 3.4: If $Z \in \mathscr{X}(N)$ is a symmetry of α, then it leaves $\mathscr{Y}$ invariant, i.e., $[Z,Y] \in \mathscr{Y}$ for each $Y \in \mathscr{Y}$. Moreover, if G is a first integral of Δ, generated by Y, then L_ZG is the corresponding first integral generated by $[Z,Y]$.

Proof: If $L_Z\alpha = 0$ and $Y \in \mathscr{Y}$, we immediately obtain:

$$d(i_{[Z,Y]}\alpha) = d(L_Z i_Y\alpha) = L_Z d(i_Y\alpha) = 0.$$

Hence, $[Z,Y] \in \mathscr{Y}$. Furthermore, we locally have [using (13)]:

$$i_{[Z,Y]}\alpha = L_Z i_Y \alpha = d(L_ZG),$$

which completes the proof. $\square$

We can extend this result slightly to conformal symmetries Z of α for which $L_Z\alpha = c\alpha$ for some constant c. We then also have that Z leaves $\mathscr{Y}$ invariant and the related first integral, generated by $[Z,Y]$, will be given by $L_ZG - cG$.

In Sec. 4 we will focus our attention on the local structure of a vector field $Y \in \mathscr{Y}$.

IV. STRUCTURE OF VECTOR FIELDS IN $\mathscr{Y}$

From (4) and (12) it follows that $Y \in \mathscr{Y}$ iff

$$L_Y d\theta = -d i_Y \mu.$$

Applying Poincare's lemma,[9] we find that locally

$$L_Y\theta = -i_Y\mu + dg, \tag{15}$$

for some smooth function g.

Within the local context to which we restrict ourselves henceforth, we may contend that (15) represents the basic equation for constructing a vector field $Y \in \mathscr{Y}$.

Suppose we are working in a local coordinate neighborhood and put

$$Y=\zeta\frac{\partial}{\partial t}+\xi^{i}\frac{\partial}{\partial q^{i}}+\eta^{i}\frac{\partial}{\partial \dot{q}^{i}}. \qquad (16)$$

Using (2) and (3) we find that Eq. (15) splits up into a set of partial differential equations for ξ^{i},η^{i},ζ and g

$$\mathscr{L}\frac{\partial\zeta}{\partial\dot{q}^{i}}+\frac{\partial\mathscr{L}}{\partial\dot{q}^{j}}\left(\frac{\partial\xi^{j}}{\partial\dot{q}^{i}}-\dot{q}^{j}\frac{\partial\zeta}{\partial\dot{q}^{i}}\right)=\frac{\partial g}{\partial\dot{q}^{i}}, \qquad (17a)$$

$$\mathscr{L}\frac{\partial\zeta}{\partial q^{i}}+\frac{\partial\mathscr{L}}{\partial\dot{q}^{j}}\left(\frac{\partial\xi^{j}}{\partial q^{i}}-\dot{q}^{j}\frac{\partial\zeta}{\partial q^{i}}\right)+\frac{\partial^{2}\mathscr{L}}{\partial\dot{q}^{i}\partial q^{j}}\xi^{j}$$
$$+\frac{\partial^{2}\mathscr{L}}{\partial\dot{q}^{i}\partial\dot{q}^{j}}\eta^{j}+\frac{\partial^{2}\mathscr{L}}{\partial\dot{q}^{i}\partial t}\zeta-Q_{i}\zeta=\frac{\partial g}{\partial q^{i}}, \qquad (17b)$$

$$\mathscr{L}\frac{\partial\zeta}{\partial t}+\frac{\partial\mathscr{L}}{\partial\dot{q}^{j}}\left(\frac{\partial\xi^{j}}{\partial t}-\dot{q}^{j}\frac{\partial\zeta}{\partial t}\right)+\frac{\partial\mathscr{L}}{\partial t}\zeta+\frac{\partial\mathscr{L}}{\partial q^{i}}\xi^{i}$$
$$-\dot{q}^{i}\left(\frac{\partial^{2}\mathscr{L}}{\partial\dot{q}^{i}\partial q^{j}}\xi^{j}+\frac{\partial^{2}\mathscr{L}}{\partial\dot{q}^{i}\partial\dot{q}^{j}}\eta^{j}+\frac{\partial^{2}\mathscr{L}}{\partial\dot{q}^{i}\partial t}\zeta\right)+Q_{i}\xi^{i}=\frac{\partial g}{\partial t}. \qquad (17c)$$

When $Q_i=0$, these equations are precisely the partial differential equations for the components of the infinitesimal generator of a one-parameter family of Noether symmetries.[7] Moreover, from (17) we can easily derive the so-called generalized Killing equations for Noether symmetries of nonconservative systems, as established in Ref. 3. Indeed, multiplying (17b) by $\dot{q}^{i}$, summing over i, and adding the resulting equation to (17c) we obtain, together with (17a), a system of $n+1$ partial differential equations in ξ^{i},ζ, and g, which precisely coincides with the system of equations derived in Ref. 3. We will investigate this relationship more thoroughly in Sec. 6.

In order to gain further insight in the structure of solutions of (17), we return to relation (14). With Δ and Y given by (6) and (16), respectively, we have in local coordinates:

$$[Y,\Delta]=-\Delta(\zeta)\frac{\partial}{\partial t}+(\eta^{i}-\Delta(\xi^{i}))\frac{\partial}{\partial q^{i}}$$
$$+(Y(\Lambda^{i})-\Delta(\eta^{i}))\frac{\partial}{\partial\dot{q}^{i}}.$$

Writing out (14) in terms of the basis of one-forms (8) and using expressions (10a and 10b), we get by equating the coefficients of $d\dot{q}^{i}-\Lambda^{i}dt$,

$$\eta^{i}=\Delta(\xi^{i})-\dot{q}^{i}\Delta(\zeta)+g^{ij}\frac{\partial Q_{k}}{\partial\dot{q}^{j}}(\xi^{k}-\dot{q}^{k}\zeta), \qquad (18)$$

whereby we have introduced the matrix $(g^{ij})=(\partial^{2}\mathscr{L}/\partial\dot{q}^{i}\partial\dot{q}^{j})^{-1}$. Consequently, the vertical components η^{i} of each $Y\in\mathscr{Y}$ are completely determined by the horizontal components ξ^{i} and the time component ζ. From (18) we can also learn that the flow generated by Y will map integral curves of Δ into curves on N which in general do not arise from a natural lifting (and extension) of curves on the configuration space M. For the latter to hold it would be necessary that $\eta^{i}=\Delta(\xi^{i})-\dot{q}^{i}\Delta(\zeta)$, (see, e.g., Ref. 7).

Next, we recall that for each $Y\in\mathscr{Y}$ the vector field $Y+h\Delta$ (with h an arbitary smooth function) again belongs to $\mathscr{Y}$ and moreover generates the same constant of the motion. Since Δ has time component 1 it is clear that, in order to find a vector field Y which generates a certain invariant of Δ, the time component ζ may, in principle, be chosen arbitrarily (e.g., $\zeta=0$). It is conceivable that in some cases this freedom of choice might enable one to reduce the complexity of the system of Eqs. (17).

V. INVARIANTS OF A DISSIPATIVE SYSTEM

Rewriting (15) we get

$$i_{Y}d\theta=-i_{Y}\mu+d(g-\langle Y,\theta\rangle).$$

With (4) this becomes

$$i_{Y}\alpha=d(g-\langle Y,\theta\rangle).$$

Comparing this with (13) we may conclude that, up to a constant, the invariant generated by Y is given by

$$G=g-\langle Y,\theta\rangle=g-\left[\mathscr{L}\zeta+\frac{\partial\mathscr{L}}{\partial\dot{q}^{i}}(\xi^{i}-\dot{q}^{i}\zeta)\right]. \qquad (19)$$

Hence, whenever $(\xi^{i},\eta^{i},\zeta,g)$ is a solution of (17), Eq. (19) immediately provides us with a (local) constant of the motion. (Notice the resemblance with the expression for an invariant of a Lagrangian system, generated through Noether's theorem.)

Conversely, let us assume we are given a constant of the motion G of a dissipative system. According to Proposition 3.2 there exists a vector field $Y\in\mathscr{Y}$ for which (13) holds. Using (9) and (10) and again representing Y by (16), we can write (13) in terms of the local basis of one-forms (8), as

$$\left(\frac{\partial^{2}\mathscr{L}}{\partial\dot{q}^{i}\partial q^{j}}-\frac{\partial^{2}\mathscr{L}}{\partial\dot{q}^{j}\partial q^{i}}\right)(\xi^{j}-\dot{q}^{j}\zeta)(dq^{i}-\dot{q}^{i}dt)$$
$$+\frac{\partial^{2}\mathscr{L}}{\partial\dot{q}^{i}\partial\dot{q}^{j}}(\eta^{j}-\Lambda^{j}\zeta)(dq^{i}-\dot{q}^{i}dt)$$
$$-\frac{\partial^{2}\mathscr{L}}{\partial\dot{q}^{i}\partial\dot{q}^{j}}(\xi^{j}-\dot{q}^{j}\zeta)(d\dot{q}^{i}-\Lambda^{i}dt)$$
$$=\frac{\partial G}{\partial q^{i}}(dq^{i}-\dot{q}^{i}dt)+\frac{\partial G}{\partial\dot{q}^{i}}(d\dot{q}^{i}-\Lambda^{i}dt). \qquad (20)$$

Hereby we have taken into account that $\Delta(G)=0$. Equating the coefficients of $d\dot{q}^{i}-\Lambda^{i}dt$ in (20), we then obtain

$$\xi^{i}-\dot{q}^{i}\zeta=-g^{ij}\frac{\partial G}{\partial\dot{q}^{j}}, \quad i=1,\ldots,n. \qquad (21)$$

In view of the observations made in the previous section, we may conclude that for a given constant of the motion G, the corresponding vector field Y is locally completely determined by (21) and (18), taking into account the liberty of choice we have for ζ.

As a control, a simple but rather tedious calculation will show that if ξ^{i} and η^{i} are given by (21) and (18), respectively, (e.g., with $\zeta=0$), the remaining conditions which result from (20) by equating the coefficients of $dq^{i}-\dot{q}^{i}dt$ will be satisfied identically.

We illustrate the above procedure for linking a vector field to a given invariant of a dissipative system, on the following example.

Consider a mechanical system with Lagrangian $\mathscr{L} = \frac{1}{2}\Sigma_{i=1}^{n}(\dot{q}^i)^2 - V(q)$, and "gyroscopic forces" $Q_i = Q_i(t,q,\dot{q})$, satisfying the condition $Q_i\dot{q}^i = 0$, (e.g., take $Q_i = \gamma_{ij}(t,q)\dot{q}^j$, with $\gamma_{ij} = -\gamma_{ji}$). The equations of motion are[14]

$$\ddot{q}^i + \frac{\partial V}{\partial q^i} = Q_i(t,q,\dot{q}). \tag{22}$$

As is well known, a first integral of (22) is provided by the energy $E = \frac{1}{2}\Sigma_{i=1}^{n}(\dot{q}^i)^2 + V(q)$. From (21) and (18) we then derive that the corresponding vector field Y (with $\zeta = 1$) is given by

$$Y = \frac{\partial}{\partial t} + Q_i\frac{\partial}{\partial \dot{q}^i}.$$

Note that, accidentally, Y here also represents a symmetry of the Lagrangian (i.e., $L_Y\mathscr{L} = 0$). All other vector fields, generating the energy integral, are obtained by adding to Y an arbitrary multiple of the given dynamical field

$$\Delta = \frac{\partial}{\partial t} + \dot{q}^i\frac{\partial}{\partial q^i} + \Big(-\frac{\partial V}{\partial q^i} + Q_i\Big)\frac{\partial}{\partial \dot{q}^i}.$$

In Sec. 6 we will compare the generalization of Noether's theorem for nonconservative systems, as established in Ref. 3, with the method and results described above.

VI. NOETHER'S THEOREM FOR NONCONSERVATIVE SYSTEMS

We first briefly summarize the main argument developed in Ref. 3.

Suppose we are given a dissipative system, described by Eqs. (1). One then considers an infinitesimal transformation of the form

$$\delta t = \epsilon\zeta(t,q,\dot{q}), \delta q^i = \epsilon\xi^i(t,q,\dot{q}).$$

$$\delta\dot{q}^i = \epsilon[\dot{\xi}^i - \dot{q}^i\dot{\zeta} + \psi^i(t,q,\dot{q})], \quad i = 1,\ldots,n, \tag{23}$$

with ϵ an infinitesimal parameter and where $\dot{\xi}^i$ and $\dot{\zeta}$ represent the formal total time derivatives of ξ^i and ζ, respectively, with

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + \dot{q}^i\frac{\partial}{\partial q^i} + \ddot{q}^i\frac{\partial}{\partial \dot{q}^i}.$$

By requiring the functions ζ, ξ^i, and ψ^i to satisfy the relation

$$\frac{\partial\mathscr{L}}{\partial\dot{q}^i}\psi^i = Q_i(\xi^i - \dot{q}^i\zeta), \tag{24}$$

it can be shown[3] that, whenever the corresponding transformation (23) leaves the form $\mathscr{L}dt$ gauge-variant, i.e.,

$$\delta(\mathscr{L}\,dt) = \epsilon\,dg, \tag{25}$$

(to the first order in ϵ) for some function $g(t,q,\dot{q})$, one can assign to it a constant of the motion of the given system. This constant of the motion, which can be expressed in terms of $\mathscr{L}$, ξ^i,ζ, and g, is precisely given by the expression on the right-hand side of (19). Conversely, with each constant of the motion one can associate an infinitesimal transformation (23) for which (24) and (25) hold.

Condition (25) leads to a set of $n + 1$ linear first-order partial differential equations in ξ^i,ζ, and g

$$\mathscr{L}\frac{\partial\zeta}{\partial\dot{q}^i} + \frac{\partial\mathscr{L}}{\partial\dot{q}^j}\Big(\frac{\partial\xi^j}{\partial\dot{q}^i} - \dot{q}^j\frac{\partial\zeta}{\partial\dot{q}^i}\Big) = \frac{\partial g}{\partial\dot{q}^i},$$

$$\mathscr{L}\Big(\frac{\partial\zeta}{\partial t} + \dot{q}^i\frac{\partial\zeta}{\partial q^i}\Big) + \frac{\partial\mathscr{L}}{\partial t}\zeta + \frac{\partial\mathscr{L}}{\partial q^i}\xi^i$$
$$+ \frac{\partial\mathscr{L}}{\partial\dot{q}^j}\Big(\frac{\partial\xi^j}{\partial t} + \dot{q}^i\frac{\partial\xi^j}{\partial q^i} - \dot{q}^i\dot{q}^j\frac{\partial\zeta}{\partial q^i} - \dot{q}^j\frac{\partial\zeta}{\partial t}\Big)$$
$$+ Q_i(\xi^i - \dot{q}^i\zeta) = \frac{\partial g}{\partial t} + \dot{q}^i\frac{\partial g}{\partial q^i},$$

the so-called generalized Killing equations. This sytem of equations, henceforth referred to by (*), is precisely the system one obtains from (17) by making a suitable combination of (17b) and (17c), as explained in Sec. 4.

Hence, whenever we have a solution (ζ,ξ^i,η^i) of (17), for some function g, the $n + 1$ functions (ζ,ξ^i) also satisfy (*). We now claim that the converse also holds if η^i is defined by (18). This equivalence between (17) and (*, 18), establishes the link between the two methods for associating vector fields (i.e., infinitesimal generators of certain transformations) with invariants of dissipative systems, always under the assumption that $\mathscr{L}$ is regular. The proof of the equivalence is completely analogous to the one presented in Ref. 7, where a similar equivalence has been pointed out between two versions of Noether's theorem for classical Lagrangian systems.

First of all it should be noticed, however, that the infinitesimal transformation (23) can not be represented by a vector field on the extended state-space N, because of the explicit $\ddot{q}$—dependence of $\delta\dot{q}^i$. (A rigorous interpretation of (23) in fact requires the introduction of a higher-order jet space.)[5] However, since the functions ψ^i do not appear in (*), nothing prevents us from assigning, in a purely formal way, to each $(n + 1)$-tuple (ζ,ξ^i) which satisfies (*), a local vector field on N of the form (16), with the functions η^i defined by (18). Using the expression (2) for the Cartan-form θ, one can then easily verify that (*) can be rewritten in a concise form as follows:

$$\Big\langle\frac{\partial}{\partial\dot{q}^j}, L_Y\theta - dg\Big\rangle = 0, \quad j = 1,\ldots,n, \tag{26a}$$

$$\langle\Delta,\mathrm{d}(\mathrm{g} - \langle\mathrm{Y},\theta\rangle)\rangle = 0. \tag{26b}$$

With Y a (local) vector field on N of the form (16), the proposed equivalence then becomes:

Proposition 6.1: Y satisfies (26a), (26b) with η^i defined by (18), iff $L_Y\theta = -i_Y\mu + dg$.

Proof: The sufficiency is trivial and in fact follows from previous considerations (see also Sec. 4).

Conversely, suppose (26a), (26b) and (18) hold. From (26a) it immediately follows that $\langle\partial/\partial\dot{q}^j, i_Y d\theta - dG\rangle = 0$, where $G = g - \langle Y,\theta\rangle$ is a first integral of Δ in view of (26b). This in turn leads to the relations

$$\xi^i - \dot{q}^i\zeta = -g^{ij}\frac{\partial G}{\partial\dot{q}^j}, \quad i = 1,\ldots,n. \tag{27}$$

On the other hand, we know from Proposition 3.2 that a vector field $\tilde{Y}\in\mathscr{Y}$ exists (with components $\tilde{\zeta},\tilde{\xi}^i,\tilde{\eta}^i$) which satisfies (13), i.e.,

$$i_{\tilde{Y}}(d\theta + \mu) = dG. \tag{28}$$

The idea is now to prove that $Y = \tilde{Y} - h\Delta$ for some function h. It then follows that Y belongs to $\mathscr{Y}$ and, hence, satisfies (15).

Since, by definition, μ contains no terms in $d\dot{q}^i$, we derive from (28)

$$\left\langle \frac{\partial}{\partial \dot{q}^i}, i_{\tilde{Y}} d\theta - dG \right\rangle = 0,$$

from which we obtain

$$\tilde{\xi}^i - \dot{q}^i \tilde{\zeta} = -g^{ij} \frac{\partial G}{\partial \dot{q}^j}, \quad i = 1,\dots,n.$$

Comparing this with (27) it is seen that $\tilde{\xi}^i - \dot{q}^i \tilde{\xi} = \xi^i - \dot{q}^i \zeta$. If we then define a function h by $\zeta = \tilde{\zeta} - h$, it follows that $\xi^i = \tilde{\xi}^i - \dot{q}^i h$.

Finally, since both η^i (by assumption) and $\tilde{\eta}^i$ (as shown in Sec. 4) satisfy (18) in terms of (ζ, ξ^i) and $(\tilde{\zeta}, \tilde{\xi}^i)$, respectively, an immediate computation gives

$$\begin{aligned}\eta^i &= \Delta(\tilde{\xi}^i - \dot{q}^i h) - \dot{q}^i \Delta(\tilde{\zeta} - h) \\ &\quad + g^{ij} \frac{\partial Q_k}{\partial \dot{q}^j} (\tilde{\xi}^k - \dot{q}^k \tilde{\zeta}) \\ &= \tilde{\eta}^i - h\Lambda^i,\end{aligned}$$

which completes the proof that $Y = \tilde{Y} - h\Delta$. □

From the equivalence established in the previous proposition it follows in particular that in practical applications, when looking for a vector field $Y \in \mathscr{Y}$, it suffices to solve the somewhat simpler system of equations (*) instead of (17).

As an illustration we now give an example which has also been treated in Ref. 3.

Example: Consider a mechanical system whose equations of motion are

$$\begin{aligned}&\ddot{q}_1 + kq_1 + 2\mu\dot{q}_1 - pq_2 = 0, \\ &\ddot{q}_2 + kq_2 + 2\mu\dot{q}_2 + pq_1 = 0,\end{aligned} \tag{29}$$

where k, μ, and p are constants and where, for convenience, we have denoted the generalized coordinates with lower indices. The Lagrangian $\mathscr{L}$ and dissipative forces Q_i are here given by:

$$\begin{aligned}\mathscr{L} &= \tfrac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2) - \tfrac{1}{2}k(q_1^2 + q_2^2), \\ Q_1 &= -2\mu\dot{q}_1 + pq_2, \\ Q_2 &= -2\mu\dot{q}_2 - pq_1.\end{aligned}$$

A solution of the generalized Killing equations (*) is provided by (see Ref. 3): $\zeta = 0, \xi^1 = e^{2\mu t}(\dot{q}_2 + \mu q_2)$;

$$\xi^2 = e^{2\mu t}(\dot{q}_1 + \mu q_1); \text{ with}$$

$$g = e^{2\mu t}[-kq_1q_2 + \dot{q}_1\dot{q}_2 + \tfrac{1}{2}p(q_2^2 - q_1^2)].$$

From (18) we obtain for the vertical components η^i of Y:

$$\begin{aligned}\eta^1 &= -e^{2\mu t}(\mu\dot{q}_2 + kq_2 + pq_1), \\ \eta^2 &= -e^{2\mu t}(\mu\dot{q}_1 + kq_1 - pq_2).\end{aligned}$$

The first integral of (29), generated by Y, becomes [see (19)]

$$\begin{aligned}G = &-e^{2\mu t}[\dot{q}_1\dot{q}_2 + \mu(q_2\dot{q}_1 + q_1\dot{q}_2) \\ &+ kq_1q_2 + \tfrac{1}{2}p(q_1^2 - q_2^2)].\end{aligned}$$

Although, as has been pointed out before, vector fields in $\mathscr{Y}$ in general do not represent symmetries of the given system, it is not excluded that in certain cases a prescription can be found for assigning a dynamical symmetry to each $Y \in \mathscr{Y}$. In the next section we will consider a special class of systems for which such a prescription is immediately at hand.

VII. A SPECIAL CASE

Suppose we are given a dissipative system for which the associated two-form α is such that the dynamical field Δ, defined by (5a) and (5b), satisfies the condition

$$L_\Delta \alpha = k\alpha, \quad k \in C^\infty(\mathbb{R}), \tag{30}$$

i.e., Δ is a conformal symmetry of α, with k at most a function of time only. Putting $l(t) = \exp(\int^t k(s)\, ds)$, it follows that

Proposition 7.1: For each $Y \in \mathscr{Y}$ the vector field lY will be a dynamical symmetry of Δ.

Proof: First of all we notice that

$$\Delta(l) = kl. \tag{31}$$

Taking into account (5a) and (11) we have

$$i_{[lY,\Delta]}\alpha = -i_\Delta L_{lY}\alpha = l i_Y i_\Delta d\alpha - \Delta(l) i_Y \alpha. \tag{32}$$

Combining (5a) and (30) we see that

$$i_\Delta d\alpha = k\alpha. \tag{33}$$

Substituting this into (32) and taking into account (31), we finally obtain

$$i_{[lY,\Delta]}\alpha = (k - \Delta(l)) i_Y \alpha = 0,$$

which completes the proof, according to Corollary 2.1. □

We now analyze condition (30) which, in view of (5a), is equivalent to (33). In a local coordinate neighborhood we have, in terms of the basis of one-forms (8), and using (10a)

$$\begin{aligned}d\alpha = d\mu &= dQ_i \wedge (dq^i - \dot{q}^i dt) \wedge dt \\ &= \frac{\partial Q_i}{\partial q^j}(dq^j - \dot{q}^j dt) \wedge (dq^i - \dot{q}^i dt) \wedge dt \\ &\quad + \frac{\partial Q_i}{\partial \dot{q}^j}(d\dot{q}^j - \Lambda^j dt) \wedge (dq^i - \dot{q}^i dt) \wedge dt.\end{aligned}$$

Herewith, (33) becomes

$$\begin{aligned}&\frac{\partial Q_i}{\partial q^j}(dq^j - \dot{q}^j dt) \wedge (dq^i - \dot{q}^i dt) \\ &\quad + \frac{\partial Q_i}{\partial \dot{q}^j}(d\dot{q}^j - \Lambda^j dt) \wedge (dq^i - \dot{q}^i dt) = k\alpha.\end{aligned}$$

With the expression (10b) for α, this leads to the following relations between Q_i and $\mathscr{L}$:

$$\frac{\partial Q_i}{\partial q^j} - \frac{\partial Q_j}{\partial q^i} = k\left(\frac{\partial^2 \mathscr{L}}{\partial \dot{q}^i \partial q^j} - \frac{\partial^2 \mathscr{L}}{\partial \dot{q}^j \partial q^i}\right),$$

$$\frac{\partial Q_i}{\partial \dot{q}^j} = k \frac{\partial^2 \mathscr{L}}{\partial \dot{q}^i \partial \dot{q}^j}$$

Hence, condition (30) will be satisfied iff the functions Q_i are of the form

$$Q_i = k\frac{\partial \mathscr{L}}{\partial \dot{q}^i} - \frac{\partial}{\partial q^i}(V \circ \pi),$$

for some function V defined on an open subset of $\mathbb{R} \times M$, and with $\pi: \mathbb{R} \times TM \to \mathbb{R} \times M$ the natural projection operator. It then follows that the given dynamical field Δ will be (locally) a Lagrangian vector field with Lagrangian $\mathscr{L}' = l'(\mathscr{L} - V\circ\pi)$, where $l' = l^{-1} = \exp(-\int^t k(s)\,ds)$.

If we introduce the Cartan-form

$$\theta' = \mathscr{L}'dt + \frac{\partial \mathscr{L}'}{\partial \dot{q}^i}(dq^i - \dot{q}^i dt),$$

we will therefore have $i_\Delta d\theta' = 0$. Moreover, it can be shown by a straightforward computation that, for each $Y\in\mathscr{Y}$, the vector field lY will satisfy the condition $L_{lY}d\theta' = 0$ or, equivalently,

$$d(i_{lY}d\theta') = 0.$$

Consequently, lY turns out to be a Noether symmetry vector field of the given system,[7] with respect to the new Lagrangian $\mathscr{L}'$. This also confirms the result mentioned in Proposition 7.1. A typical example of a system for which (30) holds, is provided by the damped harmonic oscillator:

$$\ddot{q} + \gamma\dot{q} + \omega^2 q = 0,$$

with

$$\mathscr{L} = \tfrac{1}{2}(\dot{q}^2 - \omega^2 q^2)$$

and

$$Q = -\gamma\dot{q} = -\gamma\frac{\partial \mathscr{L}}{\partial \dot{q}}.$$

VIII. FINAL REMARKS

It is by no means our intention to recommend the method described in this paper, above others, for the detection and construction of first integrals of dissipative systems. However, in studying these systems it might be useful, at least from a theoretical point of view, to gain further insight into the relationship between vector fields and invariants, as established in Sec. 3. In particular, it would, for instance, be nice to find criteria for the existence of a precise connection between the vector fields defined by (11), and symmetries of the systems under consideration (as illustrated by Proposition 7.1). For that purpose it would probably be interesting to look at special classes of dissipative systems, such as, e.g., systems with a Rayleigh-type dissipation[15] (a special case of which has been treated in Sec. 7).

Finally, it should also be noticed that the above treatment applies equally well to the Hamiltonian description of classical dissipative systems. One then simply has to replace α by $\tilde{\alpha}\in\Omega^2(\mathbb{R}\times T^*M)$, which in canonical coordinates is given by $\tilde{\alpha} = dp_i\wedge dq^i - dH\wedge dt + \tilde{Q}_i dq^i\wedge dt$, with H the Hamiltonian and $\tilde{Q}_i = \tilde{Q}_i(t,q,p)$ the phase-space components of the dissipative forces.

ACKNOWLEDGMENTS

We are particularly grateful to the Department of Physics at Northeastern University for the hospitality offered to us during the academic year 1980–1981. It is furthermore a pleasure to thank Professor E. J. Saletan, Professor G. Marmo, and Dr. W. Sarlet for many stimulating discussions. We also would like to thank Professor R. Mertens for his continuous interest in our work.

[1] R. De Ritis, G. Marmo, G. Platania, and P. Scudellaro "The Inverse Problem in Classical Mechanics: Dissipative Systems" (Preprint, Istituto di Fisica Teorica, Universita de Napoli, 1980). For an extensive treatment of the subject, see also: R. M. Santilli, *Foundations of Theoretical Physics, Vol. I; The Inverse Problem in Newtonian Mechanics* (Springer, Heidelberg, 1978).

[2] B. D. Vujanovic, "A Geometrical Approach to the Conservation Laws of Nonconservative Dynamical Systems," Tensor, N.S. **32**, 357 (1978).

[3] Dj. S. Djukic and B. D. Vujanovic, "Noether's Theory in Classical Nonconservative Mechanics," Acta Mech. **23**, 17 (1975).

[4] Dj. S. Djukic and A. M. Strauss, "Noether's Theory for Nonconservative Generalized Mechanical Systems," J. Phys. A **13**, 431 (1980).

[5] T. Nôno and F. Mimura, "Dynamical Symmetries, V," Bull. Fukuoka Univ. Ed. **28**, Part III, 7 (1979).

[6] R. Hermann, *Geometry, Physics and Systems* (Dekker, New York, 1973).

[7] W. Sarlet and F. Cantrijn, "Generalization of Noether's Theorem in Classical Mechanics," SIAM (Soc. Ind. Appl. Math.) Rev. (to appear 1981).

[8] C. Godbillion, *Géometrie Différentielle et Mécanique Analytique* (Hermann, Paris, 1969).

[9] R. Abraham and J. E. Marsden, *Foundations of Mechanics*, 2nd ed. (Benjamin/Cummings, Reading, MA, 1978).

[10] Here and in the following there is sometimes a slight abuse of notation, in that we do not explicitly indicate the appropriate restrictions of α (and Y) to the domains under consideration.

[11] V. Guillemin and S. Sternberg, *Geometric Asymptotics* (Am. Math Soc., Providence, 1977).

[12] There may of course be some exceptions, such as, for instance, if Y is a trivial symmetry of Δ.

[13] G. H. Katzin and J. Levine, "Related First Integral Theorem: A Method for Obtaining Conservation Laws of Dynamical Systems with Geodesic Trajectories in Riemannian Spaces Admitting Symmetries," J. Math. Phys. **9**, 8 (1968); G. H. Katzin and J. Levine, "Dynamical Symmetries and Constants of the Motion for Classical Particle Systems," J. Math. Phys., **15**, 1460 (1974).

[14] It should be noticed that this gyroscopic system is dissipative with respect to $\mathscr{L}$ according to the prescription given in the introduction. However, in the common physical meaning of the word, the system is clearly conservative (since the energy is preserved).

[15] L. A. Pars, *A Treatise on Analytical Dynamics* (Wiley, New York, 1965), p. 179. See also the first of Ref. 1.

Root parities and phase behavior in the slow-fluctuation technique

Marijke F. Augusteijn
Institute for Medicine and Mathematics, Ohio University, Athens, Ohio 45701
Ernst Breitenberger
Department of Physics, Ohio University, Athens, Ohio 45701

(Received 18 January 1982; accepted for publication 12 March 1982)

The slow-fluctuation technique for integrating autonomous, conservative, nonlinear, near-resonant oscillatory systems of many degrees of freedom requires ultimately no more than the study of a certain polynomial. It is shown that a paritylike property can be attributed to the roots of this polynomial, which proves helpful in even the most complex situations. It aids to classify the solutions of the equations of motion in terms of "representatives" which involve only one-half of the integration constants, the other half being rather unimportant physically, and it allows one to start up the representative motions from representative initial conditions. It also leads to a characterization of phase behavior which in particular describes not only the constant-amplitude motions but also their dynamical neighborhoods, and in many cases it explains gross features of the motion such as the occurrence of Lissajous-like patterns and orbit reversals.

PACS numbers: 03.20. + i, 46.10. + z, 02.30. + g

I. INTRODUCTION

In two previous papers, hereafter referred to as SF[1] and STAB,[2] we have developed the slow-fluctuation method in its primary mathematical aspects[1] as an integration and approximation technique for conservative, autonomous, nonlinear oscillatory systems of several degrees of freedom (d.f.), and in its application[2] to the dynamical stability of constant-amplitude (c-a) motions. The next task in the general development of the method should be the qualitative description and classification of motions with variable amplitude.

The method rests upon a complete integration of the equations of motion by quadratures involving ultimately only one polynomial, called $f(\bar p_1)$ in the previous notation.[1,2] Most qualitative properties of any particular solution of the equations should be derivable more or less directly from that polynomial. In the present paper, we show how an inconspicuous property of the roots of f can help in the classification of the solutions. We call it "parity." In many cases, various gross features of the system motion depend on nothing but root parities.

The sequence of quadratures in the solution process is such that we are led in general to relate the integration constants to given initial conditions in a particular way which singles out "representative" solutions depending on only one-half of all integration constants. The other constants are all additive; they are merely phase shifts and a shift of the zero of time, with little potential for complications. Thus each representative stands for an entire class of solutions which differ amongst each other only by these relatively unimportant shifts. The integration constants which determine the representatives are also those which determine the polynomial f; as a consequence the special initial conditions pertaining to the representatives are also connected with the parities of the roots of f.

In order to illustrate our general observations we shall repeatedly refer to details in two completed studies of specific nonlinear systems by means of the slow-fluctuation technique.[3,4]

The subject is crisscrossed by overlapping classification criteria, and studded with exceptions and special cases, in a rather dismaying way. Such is the variety of nonlinear processes. The concept of root parity, despite its simplicity, seems to possess practical usefulness even in awkward situations. Still, our presentation does not strive for utmost generality, but only for emphasis for what we have come to regard as fairly typical. Momentum-dependent couplings are not mentioned at all, for sheer brevity [root parity can readily be defined, but the analog to the all-important phase synchronization (2.10) becomes less incisive].

II. INITIAL CONDITIONS

We begin with the Hamiltonian SF (2.8):

$$\bar S(\bar p,\bar q) = \sum_{1}^{n} \omega_i \bar p_i + \bar B(\bar p) + \bar F(\bar p)\cos(g_1\bar q_1 + \cdots + g_n\bar q_n), \tag{2.1}$$

where $\bar B$ and $\bar F^2$ are polynomials, and the g_i are integers. This may be the exact Hamiltonian of some model system, or it may be the slow-fluctuation approximation to the Hamiltonian SF (2.4) of an oscillatory system having the internal resonance

$$g_1\omega_1 + \cdots + g_n\omega_n = \epsilon \tag{2.2}$$

(where $g_1 > 0$ by a numbering and sign convention). In the latter case the variables $\bar p,\bar q$ normally arise from original Cartesian variables p,q by the canonical transformation SF (2.1), (2.2); therefore, we always call $\bar p$ and $\bar q$ amplitudes and phases, respectively.

The amplitudes $\bar p_1,\ldots,\bar p_n$ arising from the transformation of a physical, oscillatory system are necessarily nonnegative. Still, there is no reason why the Hamiltonian (2.1) should not be studied without regard to the signs of the canonical momenta. We shall accordingly not introduce the restriction $\bar p_i \geqslant 0$ until later, in Sec. III C.

The system defined by (2.1) is completely integrable by quadratures, as described in SF. When there are only two

0022-2488/82/091596-14$02.50

d.f., the integration process is straightforward in $\bar p_1, \bar p_2, \bar q_1, \bar q_2$; see Refs. 3 and 4 for explicit examples. In three or more d.f., however, various relations become quite clumsy in the barred variables. For conciseness, double-barred variables are therefore introduced by the canonical transformation SF (3.1), (3.2):

$$g_1\bar q_1 + \cdots + g_n\bar q_n = \bar{\bar q}_1, \quad \bar p_1 = g_1\bar{\bar p}_1,$$
$$\bar q_i = \bar{\bar q}_i, \quad \bar p_i = g_i\bar{\bar p}_1 + \bar{\bar p}_i \quad \text{for } i = 2,...,n. \tag{2.3}$$

The Hamiltonian (2.1) then takes the form

$$\bar{\bar S}(\bar{\bar p},\bar{\bar q}) = \epsilon\bar{\bar p}_1 + \sum_2^n \omega_i\bar{\bar p}_i + \bar{\bar B}(\bar{\bar p}) + \bar{\bar F}(\bar{\bar p})\cos\bar{\bar q}_1, \tag{2.4}$$

and we use this in the following.

Initial conditions will consist of values of the $2n$ canonical variables prescribed at some time t_0. Since the barred, the double-barred, and the original variables are connected by 1-1 transformations, they are interconvertible without ambiguities, and we usually need not distinguish between different sets of initial values specified at t_0. The important task is to connect the $2n$ initial values, whichever set may be given, to the $2n$ integration constants with due regard to simplicity and transparency of the resultant classification of all solutions.

Since $\bar{\bar q}_2,...,\bar{\bar q}_n$ are cyclic in (2.4), the quadrature process begins with

$$\bar{\bar p}_i = \text{const} = \alpha_i \quad \text{for } i = 2,...,n. \tag{2.5}$$

These $n-1$ constants are clearly independent. The phase equations $\dot{\bar{\bar q}}_i = \partial\bar{\bar S}/\partial\bar{\bar p}_i$ yield n more, independent constants $\bar{\bar q}_i(t_0)$. In order to integrate these equations in their explicit forms SF (4.2), (4.3) one needs to know $\bar{\bar p}_1(t)$ explicitly. Using the amplitude conservation laws (2.5) the equation of motion $\dot{\bar{\bar p}}_1 = -\partial\bar{\bar S}/\partial\bar{\bar q}_1$ can be written in the form SF (3.8),

$$\dot{\bar{\bar p}}_1 = \bar{\bar F}(\bar{\bar p}_1,\alpha)\sin\bar{\bar q}_1, \tag{2.6}$$

from which $\bar{\bar q}_1$ can be eliminated by means of the obvious

$$\bar{\bar S} = \bar S = \text{const} = E \tag{2.7}$$

together with Eqs. (2.4) and (2.5); the result is SF (3.9),

$$\dot{\bar{\bar p}}{}_1^2 = f(\bar{\bar p}_1), \tag{2.8}$$

where f is a polynomial which contains $\alpha_2,...,\alpha_n$ and also E. The general integral of this equation is

$$t - t_0 = \int_{\bar{\bar p}_1(t_0)}^{\bar{\bar p}_1} [f(\bar{\bar p}_1)]^{-1/2}\, d\bar{\bar p}_1; \tag{2.9}$$

it conveniently brings in the initial time t_0, and also contains the initial value $\bar{\bar p}_1(t_0)$ in the role of another integration constant.

If we organize the quadrature process in this manner, the $2n$ integration constants $\bar{\bar p}_1(t_0)$, $\alpha_2,...,\alpha_n$, $\bar{\bar q}_1(t_0),...,\bar{\bar q}_n(t_0)$ are identical with the given initial values, and E merely plays the role of an auxiliary which is easily calculated from Eqs. (2.4) and (2.7) by insertion of the $n+1$ values $\bar{\bar p}_1(t_0),...,\bar{\bar q}_1(t_0)$. Moreover, the other $n-1$ initial values $\bar{\bar q}_2(t_0),...,\bar{\bar q}_n(t_0)$ are simply additive phase constants, corresponding in physical systems to simple shifts of carrier oscillations $\cos\bar q_i(t)$ under fixed amplitude modulation curves, and which we may well regard as physically trivial. Thus we have effectively obtained an $(n+1)$-way classification of solutions (excepting the $n-1$ phase shifts) directly in terms of initial values $\bar{\bar p}_1(t_0)$, $\alpha_2,...,\alpha_n$, $\bar{\bar q}_1(t_0)$. Nothing could be simpler, or better suited for numerical computations, yet an n-way classification can be devised which yields better insight and links up much more readily with all formal developments.

The appropriate change of viewpoint derives from the physical fact that the n phase functions $\bar q_i(t)$ are in general synchronized. We noted this in passing in SF Sec. IV, and gave detailed examples in two d.f. in Refs. 3 and 4. We now develop the idea in full.

Consider the system at some time $t = t_e$ such that $\bar{\bar p}_1$ is at an extremum, i.e., $\dot{\bar{\bar p}}_1(t_e) = 0$ holds. We include here all motions, even if at *constant* amplitude, or aperiodically modulated and reaching an amplitude extremum only for $t_e = \pm\infty$ (see Ref. 4, Sec. 3 F, for an example). Whichever the case, Eq. (2.8) implies that $f(\bar{\bar p}_1) = 0$ holds at $t = t_e$. In the earlier form (2.6) of this equation there are two distinct factors. It is entirely possible that the first of them, $\bar{\bar F}$, has a zero simultaneously with f. Now $\bar{\bar F}$ is in general not a polynomial, but $\bar{\bar F}^2$ always is. If the root of f at $t = t_e$ is a *multiple* root of $\bar{\bar F}^2$, then the explicit form of f in SF (3.9) [repeated in a simplified way below in Eq. (3.4)] shows that the root of f is also multiple; consequently the motion is a c-a Case (I), the phase behavior of which may require separate study, as we discuss exhaustively below in Sec. IV. If $\bar{\bar F}^2$ has only a single root at $t = t_e$, then the explicit form of $\bar{\bar F}$ in STAB (3.2) [repeated below in Eq. (3.5)] shows that $\bar{\bar p}_1$ goes through one of those exceptional low-amplitude conditions which require special treatment because the phase equations become singular, cf. SF Sec. VI and the complementary discussion in Sec. IV E below. If on the other hand $\bar{\bar F} \neq 0$ at $t = t_e$, then $\dot{\bar{\bar p}}_1$ can vanish if and only if the second factor $\sin\bar{\bar q}_1$ in Eq. (2.6) vanishes, so that $\bar{\bar q}_1(t_e)$ must be a multiple of π, as stated in SF (4.4). We summarize this trichotomy in the

Theorem: At any time t_e such that the amplitude modulation $\bar{\bar p}_1$ is at an extremum, either exceptional low-amplitude conditions hold with attendant phase singularities which may require separate study, or the motion is a c-a Case (I) whose phase behavior may also require separate study, or $\bar{\bar F} \neq 0$ holds and for the extremum it is necessary and sufficient that

$$\bar{\bar q}_1(t_e) = r\pi, \quad r \text{ integer}. \tag{2.10}$$

The vast majority of extrema clearly belongs to the third type, which is characterized by Eq. (2.10). The associated property that $\bar{\bar F} \neq 0$ at $t = t_e$ will repeatedly be used in later sections; in passing, note that it also covers all c-a motions which are Case (II) but not simultaneously Case (I), see SF Sec. V. In the remainder of this section, we deal only with this type, except where explicitly stated otherwise.

According to the transformation (2.3), the meaning of the condition (2.10) is that the n single-barred phase functions $\bar q_i(t)$ are at any amplitude modulation extremum time t_e synchronized to each other in the linear combination $\bar{\bar q}_1$. The consideration of an arbitrary initial time t_0 against this background of synchronization times t_e leads us naturally to deal first of all with the special case $t_0 = t_e$. We may then regard a motion with $t_0 = t_e$ as the *representative* of a class of motions which is generated by time shifts $t_0 - t_e$, with t_0

arbitrary, treating t_0 like an integration constant. Indeed, a representative motion with $t_0 = t_e$ has its phase constants tied by the synchronization condition (2.10), so that one integration constant is pre-empted by the need to ensure that $\bar{\bar{p}}_1$ is at an extremum, i.e., $\bar{\bar{p}}_1(t_e) = R$, a root of the polynomial f. The class will thus consist of all motions which differ only by t_0 and have the same values of all *other* integration constants. It remains to be seen how these others are most conveniently chosen.

In order to standardize the representative solutions, we may take $t_0 = t_e = 0$ without loss of generality. If in order to shorten the notation we set

$$\begin{aligned}&\bar{q}_1(0) = \delta_1,\\ &\bar{q}_i(0) = \bar{\bar{q}}_i(0) = \delta_i \quad \text{for } i = 2,\ldots,n,\end{aligned} \tag{2.11}$$

we may write the constraint (2.10) as, say,

$$\delta_1 = r\pi/g_1 - (g_2\delta_2 + \cdots + g_n\delta_n)/g_1, \tag{2.12}$$

with δ_1 dependent on the other $n - 1$ phase constants whose freedom of choice continues unrestricted. Since these remaining phase shifts appear physically not very important, we may disregard them to start with, and adopt the special initial conditions

$$\delta_1 = r\pi/g_1, \quad \delta_2 = \cdots = \delta_n = 0 \tag{2.13}$$

which give rise to what we call for brevity a "$\delta = 0$ state" of motion. Again, each such state can serve as the *representative* of an $(n - 1)$-fold class of motions which are generated by $n - 1$ shifts of carrier oscillations $\cos \bar{q}_i(t)$ under fixed amplitude curves, subject only to the constraint (2.12). Note that this class is similar to, but not quite identical with, the phase shift class left over by the previous classification in terms of $n + 1$ initial values at an arbitrary t_0.

So far, the constant E still plays the role of an auxiliary which must be calculated from the initial values. The procedure is unlogical, for E arises in the elimination process before the $\bar{\bar{p}}_1(t_0)$ in Eq. (2.9) is introduced, and before phase relations play any role. It is proper, therefore, to replace $\bar{\bar{p}}_1(0)$ by E, and accordingly to characterize the representative motions by the parameters $\alpha_2,\ldots,\alpha_n, E$. Now we have essentially the mentioned n-way classification, with each representative solution standing for an n-fold class of solutions which is generated by a time shift of the entire motion and $n - 1$ independent phase shifts of carrier oscillations, subject to (2.10).

Since E does in general not depend monotonically on $\bar{\bar{p}}_1$, we pay for the formal clarity of the new classification by a certain complication in determining inversely the proper $\bar{\bar{p}}_1(0)$ from the given $\alpha_2,\ldots,\alpha_n, E$. The polynomial f calculated from these values can have *several* suitable pairs of roots, and after one pair has been chosen, say R' and R'', either R' or R'' may still be taken as the $\bar{\bar{p}}_1(0)$ of a representative motion. Nor is it immediately clear what value of r should be associated with the root which is chosen. Thus, for any given set $\alpha_2,\ldots,\alpha_n, E$ there may be *several* (distinct) representatives associated with *different* pairs of roots and/or *different* values of r.

It is helpful to characterize these differences in terms of "representative initial conditions" which set up the representative motions. We deal with this largely practical matter in Sec. III E.

The remaining question is technical: Given *arbitrary* initial conditions at some time t_0, how do we find the explicit solution if we know only $\delta = 0$ states calculated beforehand for $t_0 = t_e = 0$, with $\delta_1 = r\pi/g_1$ and possibly several values of r, of which the right one is not at once obvious? We give the answer in the form of a complete procedure.

Since we are always free to choose the zero of the time scale, we may regard t_0 at first as indeterminate to the extent of a zero adjustment. As a preparatory step, if $\alpha_2,\ldots,\alpha_n$ are not given directly, calculate them from the initial values together with the conservation laws (2.5). Then calculate E from Eq. (2.4). Likewise calculate $\bar{\bar{p}}_1(t_0)$, if not given directly. If applicable, check the value of $\bar{F}$ for an exceptional low-amplitude condition and treat the motion accordingly, or else go on. Calculate f and find its roots. If $\bar{\bar{p}}_1(t_0)$ should be identical with a multiple root of f, treat as Case (I) or (II) c-a motion and stop there. Otherwise select the two roots R' and R'' which enclose $\bar{\bar{p}}_1(t_0)$. Go to Eq. (2.6) and determine by means of the initial values whether $\dot{\bar{\bar{p}}}_1(t_0) \gtrless 0$; then call R' that one of the roots R', R'' from which $\bar{\bar{p}}_1$ evolves *away* at t_0. If R' is single, calculate first by obvious adaptation of Eq. (2.9),

$$t_0 = \int_{R'}^{\bar{\bar{p}}_1(t_0)} [f(\bar{\bar{p}}_1)]^{-1/2}\, d\bar{\bar{p}}_1, \tag{2.14}$$

with the given initial value as the upper limit and the sign of the square root $+$ or $-$ according as $\bar{\bar{p}}_1$ increases or decreases from R' onwards; in this manner the shift t_0 is defined uniquely so as to be non-negative and less than half a modulation period in duration. Next, for every $i = 2,\ldots,n$ calculate δ_i from

$$\bar{\bar{q}}_i(t_0) = \delta_i + \{[\bar{\bar{q}}_i(t)]_{\text{repr}}\}_{t = t_0}, \tag{2.15}$$

where on the left stand the given initial values and on the right the given phase functions calculated for $t_e = 0$, for R', and for zero phase constants, but taken at the instant t_0 calculated from Eq. (2.14). Lastly, determine r from

$$\begin{aligned}\bar{\bar{q}}_1(t_0) = {}& r\pi/g_1 - (g_2\delta_2 + \cdots + g_n\delta_n)/g_1\\ &+ \{[\bar{\bar{q}}_1(t)]_{\text{repr}}\}_{t - t_0},\end{aligned} \tag{2.16}$$

where the notation is as in Eq. (2.15) (with the phase function calculated for a *zero* phase constant), and the δ_i have the values calculated above. Evidently the representative solution, if phase-shifted by the calculated $\delta_2,\ldots,\delta_n$ and by δ_1 as determined from Eq. (2.12), will at the calculated instant $t = t_0$ fulfill the given initial conditions. Finally, adjust the zero of the time scale such that t_0 takes a prescribed value, if any. If R' is multiple, modifications are necessary which are fairly obvious and can be omitted. The procedure is awkward but feasible in principle; we presented it here in order to demonstrate that the classification of solutions by means of representatives leaves no gaps.

III. ROOT PARITY

In the full classification of all representative solutions and initial conditions one needs the numerical values of the integer r generated by the synchronization condition (2.10), but for many other purposes it suffices to consider only the parity of r. For instance, in the study of Case (II) c-a motions the distinction between even and odd types is enough for a general theory; see STAB. We now show that a paritylike property can be associated directly with every root of the polynomial f.

A. Definition of root parity

By means of the conservation laws (2.5) and (2.7) the Hamiltonian (2.4) can be written as a hybrid conservation law,

$$E - \sum_2^n \omega_i \alpha_i - \epsilon \bar{\bar{p}}_1 - \bar{\bar{B}}(\bar{\bar{p}}_1,\alpha) = \bar{\bar{P}}(\bar{\bar{p}}_1,\alpha,E)$$
$$= \bar{\bar{F}}(\bar{\bar{p}}_1,\alpha)\cos \bar{\bar{q}}_1. \quad (3.1)$$

With the explicit notation $\bar{\bar{B}} = \Sigma\, b_j \bar{\bar{p}}_1{}^j$ from STAB (3.5) we can write the polynomial $\bar{\bar{P}}$ as

$$\bar{\bar{P}}(\bar{\bar{p}}_1,\alpha,E) = c_1 - (\epsilon + b_1)\bar{\bar{p}}_1 - \cdots - b_m \bar{\bar{p}}_1{}^m, \quad (3.2)$$

where

$$c_1 = E - \sum_2^n \omega_i \alpha_i - b_0;$$

this new constant c_1 is the only one which contains the quasienergy E.

When the motion is at an amplitude extremum $\bar{\bar{p}}_1 = R$, with $\bar{\bar{F}}(R,\alpha) \neq 0$, then $\bar{\bar{q}}_1 = r\pi$ holds, the cosine in Eq. (3.1) equals $+1$ or -1, and

$$\bar{\bar{P}}(R,\alpha,E) = \pm \bar{\bar{F}}(R,\alpha), \quad r_{\text{odd}}^{\text{even}} \quad (3.3)$$

follows. The existence of this twofold algebraic constraint is the physical origin of the importance of the parity of r. By a slight change of approach the idea can yet be generalized.

If $\bar{\bar{q}}_1$ is eliminated between Eqs. (2.6) and (3.1) by squaring and adding, the equation of motion results in its final form (2.8) with

$$f(\bar{\bar{p}}_1) = \bar{\bar{F}}^2 - \bar{\bar{P}}^2 = (\bar{\bar{F}} + \bar{\bar{P}})(\bar{\bar{F}} - \bar{\bar{P}}). \quad (3.4)$$

It is seen that at any root of f either one or the other relation (3.3) holds, or both in case $\bar{\bar{F}} = 0$ (and only then). We are thus led to the exhaustive

Definition: A root of f with $\bar{\bar{F}} \neq 0$ is called even if $\bar{\bar{P}} = +\bar{\bar{F}}$ and odd if $\bar{\bar{P}} = -\bar{\bar{F}}$. A root with $\bar{\bar{F}} = 0$ is called skew.

This applies to all roots including the negative and the complex ones which are not necessarily associated with physical motions. The new term "skew" is a generalization which also helps in the classification of solutions; according to the theorem of the preceding section, a motion at a skew root is either "exceptional low-amplitude" or Case (I), and conversely every Case (I) [even if it is simultaneously Case (II)] takes place at a multiple skew root.

An obvious question is: Given f, how many roots are even, odd, and skew, respectively? There is no general answer because the possibilities are unlimited, but at least one partial result can be stated in worthwhile generality. It follows directly from the factorization (3.4) as the

Theorem: If $\bar{\bar{F}}$ is a polynomial, and if $\bar{\bar{F}}$ and $\bar{\bar{P}}$ are of different degree, then half the roots are even and half are odd (if each root is counted according to its multiplicity, and if a skew root is counted as both even and odd with equal multiplicities).

If $\bar{\bar{F}}$ and $\bar{\bar{P}}$ happen to have the same degree this would not always be true because cancellation could take place and leave $\bar{\bar{F}} + \bar{\bar{P}}$ and $\bar{\bar{F}} - \bar{\bar{P}}$ at different degrees. Still, even in that case no skew roots of odd multiplicity will be possible, for instance.

If $\bar{\bar{F}}$ is not a polynomial, everything seems possible, beginning with odd-order skew roots. For example, in STAB Fig. 2 the polynomial f is cubic, point B represents a third-order odd root, and the $\bar{\bar{p}}_1$ axis represents a single skew root, as is easily shown from the details given there.

B. Conservation of root parity

Since the parity of a particular root is a discrete attribute we should expect it to be conserved under small variations of the parameters $\alpha_2,...,\alpha_n$, E; a change of parity should only be possible at the ends of certain parameter ranges where singular conditions arise, or where a confluence of several roots with a kind of barter of parities takes place. This is indeed so, but the possibilities proliferate as system complexity grows. We therefore discuss only some principal cases.

In the assumed Hamiltonian (2.1), $\bar{B}(\bar{p})$ and $[\bar{F}(\bar{p})]^2$ are given as polynomials. From these we derive the $\bar{\bar{P}}$ and the $\bar{\bar{F}}$ of Eq. (3.1) through the substitutions (2.3), (2.5), and (2.7) together with the extraction of square roots. These operations are continuous in every respect, hence $\bar{\bar{P}}$ and $\bar{\bar{F}}$ are both continuous functions of $\bar{\bar{p}}_1$ and of the parameters $\alpha_2,...,\alpha_n,E$.

If we choose a continuous curve in a complex $\bar{\bar{p}}_1$ plane and map it into a complex w plane by means of $\pm\bar{\bar{F}}$ or $\bar{\bar{P}}$ at a fixed set of values $\alpha_2,...,\alpha_n,E$, the map curves will be continuous. Moreover, if we restrict the chosen curves to the interior of a suitably bounded domain of the $\bar{\bar{p}}_1$ plane, the map will be 1–1 and the entire plane can be covered by such domains.

The roots of f are also continuous in $\alpha_2,...,\alpha_n,E$, a fact of which much was made in STAB. If we vary these parameter values through a continuous n-dimensional sequence, a root R will trace out a continuous curve in the $\bar{\bar{p}}_1$ plane. Call it a root curve. Two root curves may intersect or touch. A root may also abruptly appear or disappear at certain parameter values, but only if at these values the degree of f changes. This is an exceptional case which would always need special study, cf. STAB Sec. III A, so we exclude it by a suitable restriction on the parameter variations.

Now choose an arc of some root curve and map it triply into the w plane by means of $+\bar{\bar{F}}$, $-\bar{\bar{F}}$, and $\bar{\bar{P}}$ at a set of parameter values which is associated with a particular $\bar{\bar{p}}_1 = R$ lying on the chosen curve. R will be even or odd if the map of either $+\bar{\bar{F}}$ or $-\bar{\bar{F}}$, respectively, intersects with $\bar{\bar{P}}$,' and conversely; it will be skew if and only if all three maps intersect at one point (which must be the origin of the w plane). Under a change of parameter values to another set associated with an $R\,'$ nearby on the same root curve, intersections off the origin in the w plane cannot jump between the $+\bar{\bar{F}}$ and $-\bar{\bar{F}}$ maps, because only continuous deformations of the maps are possible. There may also be an intersection (or contact) of the given root curve with another one; by continuity, there will be corresponding multiple intersections of the $+\bar{\bar{F}}$ maps or the $-\bar{\bar{F}}$ maps with the $\bar{\bar{P}}$ maps. From this all-round continuity alone we conclude the

Conservation Theorem: Throughout some open neighborhood of a set of values $\alpha_2,...,\alpha_n,E$, a root of $f(\bar{\bar{p}}_1)$ of even or odd parity keeps this parity, or splits into, or coalesces with, roots of this same parity.

Thus, a change of parity between even and odd, or a mixing of parities, can only occur through an intermediary skew stage.

C. Sign conventions

An arbitrary sign reversal of $\bar{\bar{F}}$ would turn even roots of f into odd, and vice versa. Thus, when the system (2.1) is defined in the abstract, with $\bar{F}^2$ given as a polynomial, the square root leading to the eventual $\bar{\bar{F}}$ needs to have its sign fixed by a convention in order to fix the parities. On the other hand, when the system is defined in terms of physical properties, then $\bar{F}$ and with it $\bar{\bar{F}}$ will usually have their signs unmistakably defined within the range of amplitude values permitted in the system, but there may be doubt as to the continuation of $\bar{\bar{F}}$ beyond this range, and hence, a sign convention may again be needed if we want to discuss the parities of roots which are not associated with physical motions.

The explicit form of $\bar{\bar{F}}$ is always as given in SF (3.2):

$$\bar{\bar{F}}(\bar{\bar{p}}_1,\alpha)=C(g_1\bar{\bar{p}}_1)^{l_1/2}(g_2\bar{\bar{p}}_1+\alpha_2)^{l_2/2} \cdots(g_n\bar{\bar{p}}_1+\alpha_n)^{l_n/2}\bar{\bar{Q}}(\bar{\bar{p}}_1,\alpha), \qquad (3.5)$$

where C is a system constant and $C\neq 0$, $\bar{\bar{Q}}$ is a polynomial, and the l_i are positive integers or zero, with $l_1\neq 0$ by numbering convention. If for physical reasons, say, $\bar{\bar{p}}_1\geqslant 0$ is prescribed, do we continue $\bar{\bar{F}}$ to negative $\bar{\bar{p}}_1$ values, say, by writing $|\bar{\bar{p}}_1|$ everywhere in (3.5)? The same doubt might exist in regard of the continuation of $\bar{\bar{P}}$, and hence, even of f.

The basic conservation laws (2.5) become, after the substitutions (2.3),

$$g_i\bar{\bar{p}}_1+\alpha_i=\bar{p}_i, \quad i=2,...,n. \qquad (3.6)$$

[In conjunction with $g_1\bar{\bar{p}}_1=\bar{p}_1$, this is what makes the elimination from (2.1) to (2.4) possible, and what causes the structure of the factors in (3.5).] It is seen that a zero of $\bar{\bar{F}}$ implies the vanishing of one or more $\bar{p}_i$, or a root of $\bar{\bar{Q}}$, or both. The observation is as general as it is trivial, but it leads to a much sharper distinction when the given system is a physical one of coupled oscillators having non-negative amplitudes.

If $\bar{p}_i\geqslant 0$ for $i=1,...,n$ is thus prescribed, then also $\bar{\bar{p}}_1=\bar{p}_1/g_1\geqslant 0$ holds and Eqs. (3.6) imply that

$$-g_i\bar{\bar{p}}_1\leqslant\alpha_i, \quad i=2,...,n. \qquad (3.7)$$

This means an upper or lower bound to $\bar{\bar{p}}_1$ according as g_i is negative or positive. In STAB Sec. III C we called the (closed or right-infinite) interval from the least upper bound to the largest lower bound (including $\bar{\bar{p}}_1\geqslant 0$) the "domain of $\bar{\bar{p}}_1$". No physically possible motion can take place outside the domain. If now a $\bar{p}_i$ (with $g_i\neq 0$) vanishes, it follows from Eq. (3.6) that the bound (3.7) is actually reached by this $\bar{p}_i$; the vanishing therefore occurs exactly at an endpoint of the domain, and $\bar{\bar{F}}$ vanishes at this endpoint, too. On the other hand, $\bar{\bar{F}}$ can vanish *inside* the domain only if the equality sign in (3.7) does *not* hold for any i; then a root of $\bar{\bar{Q}}$ must be responsible. This can happen only in systems which have a (nonconstant) $\bar{\bar{Q}}$-polynomial. For reference purposes we describe the main features of the situation in the

Theorem: A skew root of f inside the domain of $\bar{\bar{p}}_1$ is due to a root of $\bar{\bar{Q}}$ and has even multiplicity.

Of course, a root of $\bar{\bar{Q}}$ could also by coincidence occur at a domain endpoint. If so, one of the bounds (3.7) must be reached at the same time so that the resultant zero of $\bar{\bar{F}}$ is of a higher order than the root of $\bar{\bar{Q}}$.

At any rate, except for this occasional possibility of a root of $\bar{\bar{Q}}$ inside the domain, skew roots can occur only at the domain endpoints, together with a vanishing of one or more $\bar{p}_i$, and beyond the domain. Accordingly, changes of root parity by passage through a skew stage will in general take place at domain endpoints. It becomes all the more important to be definite about the signs of $\bar{\bar{F}}$, etc., on *both* sides of the endpoints. We shall henceforth employ the

Physical Oscillator Convention. $\bar{P}$, $\bar{Q}$, and $\bar{F}^2$ being given as polynomials over the domain of $\bar{\bar{p}}_1$, they will be continued as polynomials beyond the domain. Even-l_i factors in $\bar{\bar{F}}$ will likewise be continued as polynomials, but odd-l_i factors will be continued as positive square roots.

The immutable sign of the square roots has the consequence that the rule $\sqrt{x}\times\sqrt{x}=x$ does not hold where x can be negative, i.e., precisely outside the domain within which all the arguments in the parentheses of Eq. (3.5) are non-negative; we must write $\sqrt{x}\times\sqrt{x}=\sqrt{x^2}=|x|$ instead. For example, when there is a confluence of two zeros of $\bar{\bar{F}}$ which are both of order $\frac{1}{2}$ (and equivalent to "exceptional low-amplitude conditions"), the graph of $\bar{\bar{F}}$ will have a kink. Such a confluence can only occur in special systems, as in the case of the "three interacting waves" of STAB Fig. 6b, where the looped curve happens to be (apart from a numerical factor) the graph of $\pm\bar{\bar{F}}$ with the upper and lower halves corresponding to the two signs, and kinks at the confluence $\alpha_2=\alpha_3$.

This sign convention is entirely natural and simple. In systems with negative momenta, different conventions are possible, of course, and could conceivably appear more natural.

D. Change of root parity

We now discuss a few typical instances of parity change. We assume the physical oscillator convention and speak only of real roots, but the generalization to complex ones will be obvious where appropriate.

Let Γ be real and $\bar{\bar{F}}(\Gamma,\alpha)=0$. Such a Γ is generally not a root of f because it need not be a root of $\bar{\bar{P}}$. However, from the explicit expression (3.2) it is seen that E is an additive constant in $\bar{\bar{P}}$, hence there exists exactly one $E=E_\Gamma$ such that $\bar{\bar{P}}(\Gamma,\alpha,E_\Gamma)=0$. This E_Γ does not have to be a physically possible system energy; it is merely a (real) parameter value such that $\bar{\bar{P}}$ and $\bar{\bar{F}}$ at the given set of values $\alpha_2,...,\alpha_n,E_\Gamma$ vanish for the same $\bar{\bar{p}}_1=\Gamma$, which is therefore a skew root of f. For brevity, we now write the polynomial (3.2) in Taylor series form

$$\bar{\bar{P}}=(E-E_\Gamma)+K_1(\bar{\bar{p}}_1-\Gamma)+\cdots. \qquad (3.8)$$

Assume first for discussion that $\bar{\bar{F}}$ vanishes near Γ as

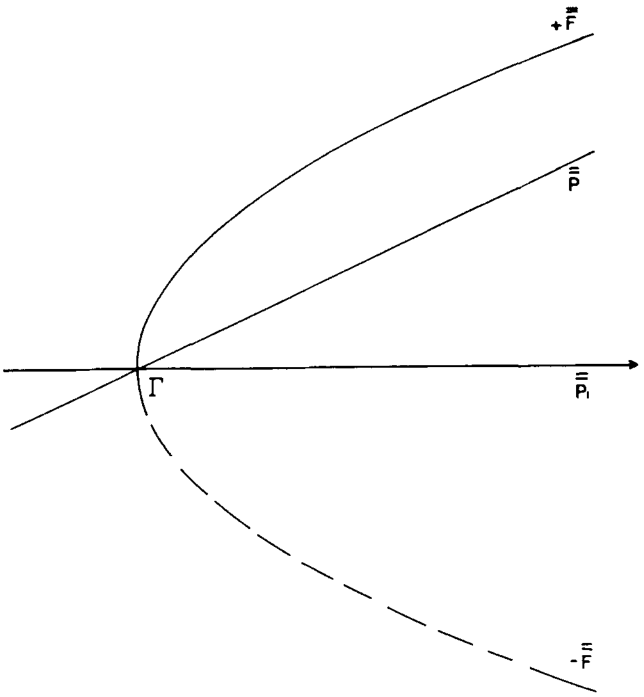


FIG. 1. The formation of a single skew root in accordance with Eqs. (3.8) and (3.9). As drawn, Γ is a lower endpoint of the domain of $\bar{p}_1$; $K_1 > 0$ has been chosen for definiteness, but it makes no essential difference if $K_1 < 0$ or even $K_1 = 0$.

$$\bar{\bar{F}} = K_0(\bar{p}_1 - \Gamma)^{1/2}, \qquad (3.9)$$

with some $K_0 > 0$. Then Γ is a single root of $\bar{\bar{F}}^2$ and a double root of $\bar{\bar{P}}^2$ for $E = E_\Gamma$, hence it is a single, skew root of f for the parameter values $\alpha_2,...,\alpha_n,E_\Gamma$. If E is increased slightly over E_Γ, the graph of $\bar{\bar{P}}$ will intersect the graph of $+\bar{\bar{F}}$ to the right of Γ at some $\bar{p}_1$ which is then an even, single root of f, see Fig. 1; decrease of E yields an intersection with the graph of $-\bar{\bar{F}}$ and therefore an odd root. Thus, as E is varied, the root parity changes between even and odd by a passage through skew.

Equation (3.5) shows that a single skew root can only be of the form $\Gamma = 0$ or $\Gamma = -\alpha_i/g_i$ for some i: it is therefore real. According to the Theorem of Sec. III C it cannot lie inside the domain of $\bar{p}_1$. Under a variation of the α_i as well as of E the single root of f evolving out of Γ and back must also be real, for a complex root can only approach Γ jointly with its conjugate. The only way this argument could be vitiated is that under the parameter change Γ itself coalesces with some other root and becomes multiple, but by continuity this is only possible for sufficiently large changes. In fair generality we may therefore state the

Theorem: A single skew root Γ is real and does not lie inside the domain of $\bar{p}_1$. For sufficiently small parameter changes a root of definite parity evolving out of Γ is real and single, and changes parity when passing through Γ.

Next, assume instead of (3.9) that near Γ we have

$$\bar{\bar{F}} = K_0(\bar{p}_1 - \Gamma), \qquad (3.10)$$

with some $K_0 > 0$, and take also $K_1 > 0$ but $K_1 \neq K_0$. Instead of Fig. 1 we then have the two graphs of Fig. 2, according as $K_0 \gtrless K_1$. Here Γ is a double, skew root of f for $E = E_\Gamma$, and if the graph of $\bar{\bar{P}}$ is shifted up or down by a change of E the skew root is seen to split into an even and an odd one. Characteristically, the two new, single roots of f straddle the skew one in Fig. 2(b) but lie to one side of it in Fig. 2(a). If we write the inequality $K_0 \gtrless K_1$ as

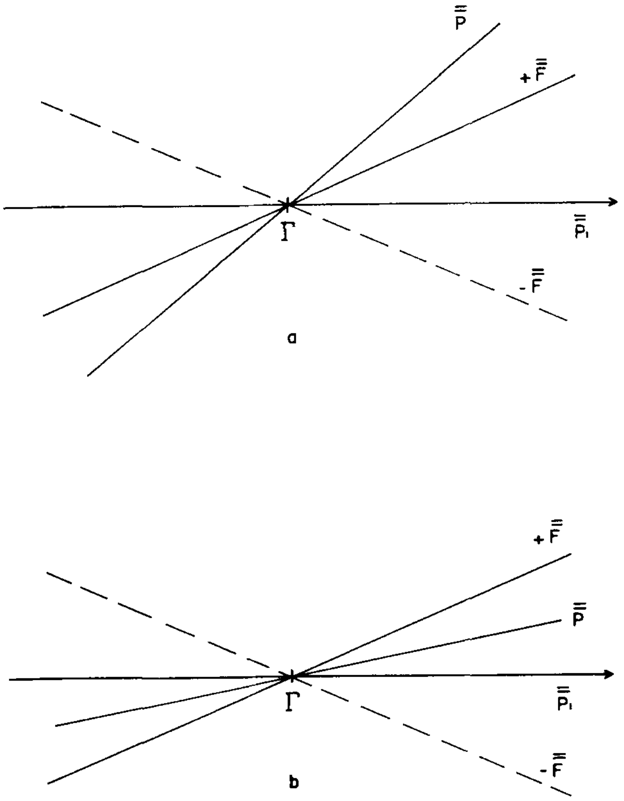


FIG. 2. The formation of the simplest type of double skew root in accordance with Eqs. (3.8) and (3.10), when both $\bar{\bar{P}}$ and $\bar{\bar{F}}$ exhibit locally linear behavior (with different slopes). Depending on the steepness of the graph of $\bar{\bar{P}}$, the c-a motion at Γ will be (a) orbitally stable, or (b) unstable. Γ may be an upper or lower domain endpoint, or may lie inside the domain of $\bar{p}_1$.

$$\partial\bar{\bar{F}}/\partial\bar{p}_1 \gtrless \partial\bar{\bar{P}}/\partial\bar{p}_1 = -(\epsilon + \partial\bar{\bar{B}}/\partial\bar{p}_1), \qquad (3.11)$$

with all the derivatives taken at the value $\bar{p}_1 = \Gamma$, the criterion STAB (3.12) shows that the Case (I) c-a motion at the double root $R = \Gamma$ (if it is physically possible) will be orbitally stable or unstable, respectively, in the two situations. The skew root can here be regarded as the confluence of two roots of opposite parity, and can only split into such an even–odd pair. Reference 4 contains clear examples.

Third, assume instead of (3.9) that

$$\bar{\bar{F}} = K_0(\bar{p}_1 - \Gamma)^{3/2}, \qquad (3.12)$$

with some $K_0 > 0$, and also assume $K_1 > 0$, see Fig. 3. With Γ now a triple root of $\bar{\bar{F}}^2$ but still a double root of $\bar{\bar{P}}^2$, it is again a double, skew root of f. Under a decrease of E the skew root splits into a real, even–odd pair, much as in the orbitally stable situation of Fig. 2(a), but an increase of E evidently yields a pair of complex roots which again must be of opposite parity, since of the two factors on the right-hand side of Eq. (3.4) neither can have a double zero in the vicinity of Γ, by continuity.

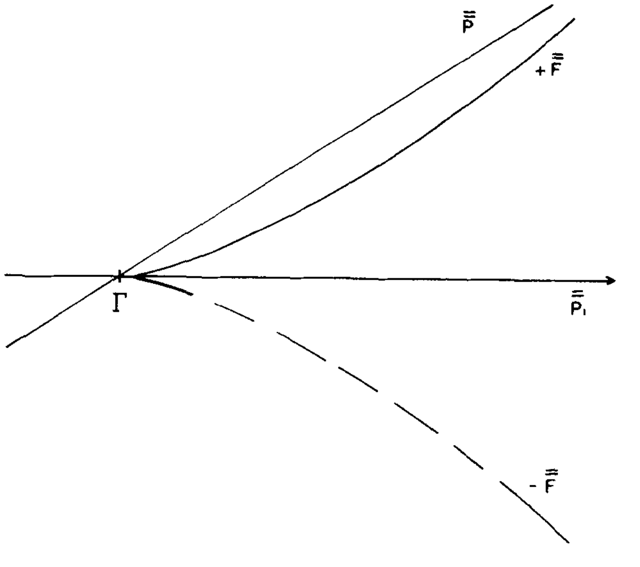


FIG. 3. The formation of a double skew root in accordance with Eqs. (3.8) and (3.12). As drawn, Γ is a lower endpoint of the domain of $\bar{p}_1$; $K_1 > 0$ has been chosen for definiteness, but it makes no essential difference if $K_1 < 0$, and the c-a motion at Γ is always orbitally stable, a Case (I) with $\bar{p}_1 \equiv \Gamma$.

The row of examples (3.9)–(3.12) can readily be continued. We may also in the Taylor expansion (3.8) admit $K_1 = 0$, and so on, to explore the behavior of complicated higher roots. The principle of the argument is always the same, and no further instances are needed.

Double skew roots deserve particular attention because of their frequent occurrence. If real, their properties are described by Figs. 2(a) and 2(b) and obvious generalizations of these. However, a first-order zero of $\bar{\bar{F}}$, which is the least we must have, can also arise in certain systems as the confluence of two zeros of order $\frac{1}{2}$ (i.e., exceptional low-amplitude conditions are reached simultaneously by exactly two d.f.). If it is also a double skew root, the latter can split into two single roots which may be of any parity, or skew. Figure 4 explains the situation. Considering all the possibilities, we have the

Theorem: A real, double, skew root can only split into, or result from the confluence of, two single roots of opposite parities; except if it occurs as a confluence of two exceptional low-amplitude conditions, when it can split into two single roots which may be of any parity, or skew.

A word of caution: These results hold for variations of the parameters *only*. Under other changes parity may be significantly affected. In particular, parity can change under the (degenerate) coordinate transformations which often are allowed in systems with intrinsic symmetries. Thus, Ref. 4 contains clear examples of straight-line c-a motions which convert between Cases (I) and (II) under 45° rotations, with conversion of double roots between skew and even.

E. Representative initial conditions

The representative $\delta = 0$ motions which start out with $\bar{p}_1$ at one of the roots of f depend on the root parity by Eq. (2.13). Since an entire range of 2π must be available for δ_1, $2g_1$ values of r will be admissible. For instance, if $g_1 = 1$, then no more than the values $r = r' = 0,1$ (say) come into question, but if $g_1 = 2$, then $r = r' = 0,1,2,3$ (say) are admissible.

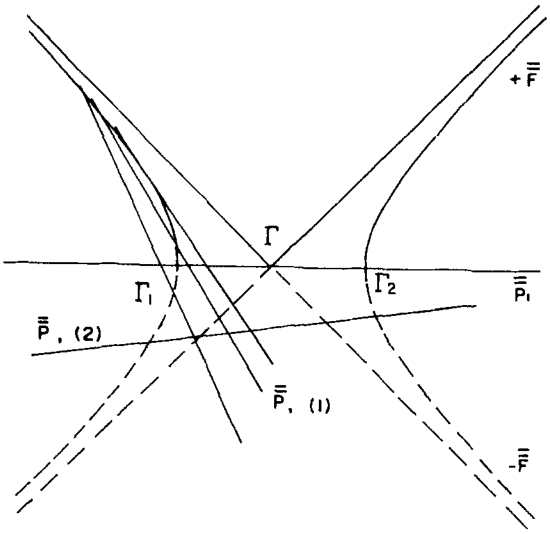


FIG. 4. Schematic representation of the confluence of two single skew roots Γ_1, Γ_2 into a double skew root Γ at which $\bar{\bar{F}}$ graphs as a kinked straight line. The motions in the neighborhood of the c-a motion at Γ can be of entirely different types. Assuming that $\bar{\bar{P}}$ graphs as a straight line and that Γ_1 is an upper endpoint of the domain of $\bar{p}_1$: (1) shows three possibilities of neighboring motions in the orbitally stable case, one even Case (II) c-a, one even–even and one even–odd modulation; (2) shows one unstable possibility (slope of $\bar{\bar{P}}$ less than the slope of the kink at Γ).

And so on. Thus the representative initial conditions which set up the $\delta = 0$ states come with a multiplicity which derives from the multiplicity of values r'. Furthermore, this multiplicity depends on the g-coefficient of the phase shift for which we solve Eq. (2.10). In Eq. (2.12) we solved for δ_1, but there is no need for that, we might solve for any δ_j if only $g_j \neq 0$. (In fact, in both Refs. 3 and 4 the choice $j = 2$ was made for simple reasons of convenience.) Evidently, in a given system, for different choices of j (or what may amount to the same, for different numberings of the d.f.) the representative initial conditions, and consequently the sets of representative solutions, can be substantially different. Usually it will be best to choose a d.f. with the smallest available g_j to be assigned that phase shift which depends on the others. As there can be no universal rule we assume for the following that a firm convention has been made, and that the specific numerical values r' of r have been selected.

The initial conditions are most easily understood in a physical oscillatory system with all $\bar{p}_i \geqslant 0$ when the motion is translated back into coordinate oscillations by means of the canonical SF (2.1),

$$q_i = (2\bar{p}_i/m_i\omega_i)^{1/2}\cos\bar{q}_i,$$
$$p_i = -(2m_i\omega_i\,\bar{p}_i)^{1/2}\sin\bar{q}_i. \qquad (3.13)$$

Using Eqs. (2.3), (2.5), (2.11), and (2.13), together with the notation $\bar{p}_1(0) = R'$, we have first

$$q_i(0) = [2(g_i R' + \alpha_i)/m_i\omega_i]^{1/2},$$
$$p_i(0) = 0, \quad i = 2,\ldots,n. \qquad (3.14)$$

The $n-1$ d.f. with the independent phase constants $\delta_i = 0$ are therefore always released from rest at $t = 0$; to be precise: From canonical rest in the p,q phase space, see Ref. 3, Sec. XI, for an example of the distinction.

On the other hand,

$$q_1(0) = (2g_1R\,'/m_1\omega_1)^{1/2}\cos(r'\pi/g_1),$$
$$p_1(0) = -(2m_1\omega_1 g_1R\,')^{1/2}\sin(r'\pi/g_1) \qquad (3.15)$$

depends on g_1 and the choices of r'. Let $g_1 = 1$; then $p_1(0) = 0$ and we again have release from rest, while $q_1(0) > 0$ or < 0 according as r' is even or odd, so that parity distinguishes merely between release over the two halves of the q_1 axis. Reference 3 is a type case. If $g_1 = 2$ and we choose $r' = 0,1,2,3$ as above, then the even $r' = 0,2$ are seen to correspond to release from rest over the two axis halves, but the odd $r' = 1,3$ yield $q_1(0) = 0$, with $p_1(0) < 0$ and > 0, respectively, a condition which in the type case of Ref. 4 we called "transverse launch." If $g_1 = 3$, release from rest is still obtained for $r' = 0,3$ but other values require transverse launch at certain coordinates $q_1(0) \neq 0$. And so on.

Regardless of the various conventions made, these links of root parity with physical features of the system motion are seen to rest mainly upon the g coefficients. That is, they arise from the nature of the *resonance* in the Hamiltonian (2.1) rather than from the physical *coupling* terms $\bar{B}$ and $\bar{F}$. The corresponding patterns of motion can claim physical interest as well as classification usefulness.

Still, the coupling terms determine the structure of $f(\bar{\bar{p}}_1)$ and hence, the number and location of its real, non-negative roots. When several such pairs of roots $R\,'$, $R\,''$ coexist, release from rest in one region of space will lead to a root of one pair, release in another region to a root of another pair, and similar for the other g-dependent initial conditions. Reference 4, Fig. 2, is a type case. A general discussion is better not attempted because of too much variety in the possible couplings.

IV. ROOT PARITY AND THE PHASE FUNCTIONS

The evolution of $\bar{\bar{q}}_1(t)$ between two roots of $f(\bar{\bar{p}}_1)$ must depend sensibly on whether they have equal or opposite parities. This aspect of the synchronization (2.10) can be exploited variously to put the concept of root parity to practical use.

A. The main result

Lemma: At a root $\bar{\bar{p}}_1 = R$ of $f(\bar{\bar{p}}_1)$ with $\bar{\bar{F}}(R,\alpha) \neq 0$ the slope of f satisfies

$$f' = \pm 2\bar{\bar{F}}\dot{\bar{\bar{q}}}_1, \quad r^{\text{even}}_{\text{odd}}. \qquad (4.1)$$

Proof: At a root of definite parity, the cosine in Eqs. (2.4) and (3.1) is stationary at ± 1. From the definition of f in Eq. (3.4) we have in general

$$\frac{df}{d\bar{\bar{p}}_1} = 2\bar{\bar{F}}\frac{d\bar{\bar{F}}}{d\bar{\bar{p}}_1} - 2\bar{\bar{P}}\frac{d\bar{\bar{P}}}{d\bar{\bar{p}}_1};$$

if we specialize to $\bar{\bar{p}}_1 = R$ we can first eliminate both $\bar{\bar{P}}$ and its derivative by means of (3.1) to yield

$$f' = 2\bar{\bar{F}}\left[\frac{d\bar{\bar{F}}}{d\bar{\bar{p}}_1} \pm \left(\epsilon + \frac{d\bar{\bar{B}}}{d\bar{\bar{p}}_1}\right)\right],$$

and here the bracket is seen from the equation of motion $\dot{\bar{\bar{q}}}_1 = \partial\bar{\bar{S}}/\partial\bar{\bar{p}}_1$ to be $\pm$ the value of $\dot{\bar{\bar{q}}}_1$ at R. Q.E.D.

Corollary: At a root of definite parity, if $\bar{\bar{q}}_1$ is stationary then the root is multiple.

Consider now an amplitude modulation between two *single* roots $R\,'$, $R\,''$ of definite parities (even or odd). The phase $\bar{\bar{q}}_1$ starts out at $R\,'$, say, with the value $r\pi$, and necessarily with $\dot{\bar{\bar{q}}}_1 \neq 0$. As $\bar{\bar{p}}_1$ evolves towards $R\,''$, but before $R\,''$ is reached, $\bar{\bar{q}}_1$ cannot change by as much as $+\pi$ or $-\pi$, or return to the value $r\pi$, because according to the equation of motion (2.6) that would imply an extremum of $\bar{\bar{p}}_1$ between $R\,'$ and $R\,''$, contrary to hypothesis. When $R\,''$ is reached there are consequently two possibilities.

Either $\bar{\bar{q}}_1$ is back at its initial value $r\pi$; then $R\,''$ is of the same parity as $R\,'$. Also, the sign of $\dot{\bar{\bar{q}}}_1$ must be opposite to what it was at $R\,'$, otherwise $\bar{\bar{q}}_1$ would again have a zero between $R\,'$ and $R\,''$. In essence, $\bar{\bar{q}}_1$ is sinelike, with zeros at the modulation extrema and nowhere else.

Or we have $\bar{\bar{q}}_1 = (r \pm 1)\pi$ at $R\,''$; then the root parities are opposite. The rate of change $\dot{\bar{\bar{q}}}_1$ must have the same sign at $R\,''$ as at $R\,'$, otherwise $\bar{\bar{q}}_1$ would also go through the value $(r+1)\pi$ or $(r-1)\pi$ somewhere in between. Thus $\bar{\bar{q}}_1$ could be described as an unbroken straight line which rises or descends by π per half-period of modulation, with a superimposed periodic function having simple or multiple zeros at the modulation extrema, and possibly others; the sum of the two is not necessarily monotonic, although the straight line is. In short summary we have the

Theorem: In amplitude modulation between simple roots with definite parities, the combined phase obeys

$$0 < |\bar{\bar{q}}_1(t) - \bar{\bar{q}}_1(t_e)| < \pi \qquad (4.2)$$

in every open interval between successive modulation extrema. If the parities are equal, r remains constant; $\bar{\bar{q}}_1$ is purely periodic with simple zeros at the extrema, and has no other zeros. If the parities are opposite, r goes through consecutive integers; $\dot{\bar{\bar{q}}}_1$ at the extrema is positive or negative throughout according as r increases or decreases, but $\bar{\bar{q}}_1$ is not necessarily monotonic in between two extrema.

As an immediate application, consider the representative initial condition for motion between roots of equal parity. Since r remains constant, one single type of condition suffices, e.g., release from rest, with certain regions of space corresponding to the two roots of the pair. Quite different for roots of opposite parity; the pertaining monotonic change of r presents a cycle of $2g_1$ possible, representative initial conditions. One of these will have to be chosen by convention, for definiteness. In fact, a $\delta = 0$ state will not have the phases back to $\bar{q}_i(t_e) = \delta_i$ at the next extremum $t_e \neq 0$ because of the lack of synchronization between modulation and carrier oscillations; in other words, the representative solutions are rather different for different choices of zero time.

Note that, occasional graphic language notwithstanding, this entire section is independent of the Physical Oscillator Convention.

B. Orbit patterns in configuration space (roots of definite parity)

For ease of discussion, let us take a physical system of oscillators and return into the space of the configuration coordinates q_i by means of the transformation (3.11). Assume two d.f.; assume both normal frequencies positive, and a near-resonance $g_1\omega_1 + g_2\omega_2 = \epsilon$. Since $g_1 > 0$ by convention, it is necessary that $g_2 < 0$; we write $g_2 = -g_2'$ with $g_2' > 0$. Now consider only motions between (or at) roots of definite parity; if, in particular, a representative $\delta = 0$ motion at $t = t_e = 0$ is set up, the constraint equation (2.12) yields with $\delta_2 = 0$ and the specific choice $r = r'$,

$$\delta_1 = r'\pi/g_1 \tag{4.3}$$

for the initial phase difference.

Our aim is to compare the orbit in the q_1,q_2 plane to Lissajous figures. As this term is commonly understood, it applies to the *closed* paths which a complex vector $q_1 + iq_2$ will map out if both real and imaginary part are purely harmonic with constant amplitudes, while the two frequencies are commensurate, say $n_1\Omega$ and $n_2\Omega$, where n_1, n_2 are integers which we take to be relatively prime to avoid ambiguities. For each frequency ratio n_1:n_2 there is an infinity of such figures, depending on the initial phase difference between the two harmonic motions,

$$q_1 = A_1 \cos(n_1\Omega t + d_1), \quad q_2 = A_2 \cos(n_2\Omega t + d_2). \tag{4.4}$$

Note that the arguments of the two trigonometric functions satisfy the almost trivial identity

$$n_2(n_1\Omega t + d_1) = n_1(n_2\Omega t + d_2) + (n_2 d_1 - n_1 d_2) \tag{4.5}$$

at all times.

Our system motion is

$$q_1 = A_1 \cos \bar{q}_1, \quad q_2 = A_2 \cos \bar{q}_2. \tag{4.6}$$

We may compare it to (4.4) provided A_1, A_2 are constant; this will be the case, at least to a good approximation, for some time interval around a modulation extremum time $t = t_e$ when the amplitudes are stationary. If in these circumstances the motion were to be an exact Lissajous figure, it is necessary first of all that the two phase functions satisfy an identity

$$n_2\bar{q}_1 - n_1\bar{q}_2 = n_2 d_1 - n_1 d_2 \tag{4.7}$$

in analogy to (4.5), at all times during some subinterval of the said interval.

Equation (4.7) resembles the synchronization condition (2.10),

$$\bar{\bar{q}}_1(t_e) = g_1\bar{q}_1(t_e) - g_2'\bar{q}_2(t_e) = r\pi, \quad r \text{ integer}. \tag{4.8}$$

Here g_1 and g_2' are not necessarily relatively prime, in contrast to n_1 and n_2, but if we divide (4.8) by d, the greatest common divisor of g_1 and g_2', then the two relations can be equated term-by-term with the result

$$n_2 = g_1/d, \quad n_1 = g_2'/d, \quad n_2 d_1 - n_1 d_2 = r\pi/d. \tag{4.9}$$

The last of these three identifications shows how the two phase functions in Eqs. (4.4) must be synchronized in the particular Lissajous figures which may arise from the motion (4.6). The two phase constants d_1, d_2 are tightly connected; for instance, in a $\delta = 0$ state with $d_2 = 0$, we see from (4.9) that $d_1 = r\pi/dn_2$ holds at $t = t_e = 0$, consistent with (4.3). Given this link between d_1 and d_2, we shall henceforth regard our Lissajous figures as determined, not by the phase difference $d_1 - d_2$ in Eqs. (4.4), but by the value $r\pi/d$ of the linear combination $n_2 d_1 - n_1 d_2$, which is a kind of "weighted relative phase" and unambiguously determines the simple relative phase $d_1 - d_2$ as soon as one of d_1, d_2 is given in any manner whatever.

For an exact Lissajous figure to emerge it is furthermore necessary that $\bar{q}_1, \bar{q}_2$ have proportional rates of growth. Hence, if the instantaneous frequencies happen to be exactly commensurate at the modulation extremum under consideration,

$$\dot{\bar{q}}_1(t_e):\dot{\bar{q}}_2(t_e) = n_1:n_2, \tag{4.10}$$

then the motion must begin as a Lissajous figure having the frequencies and phases specified by the relations (4.9); in particular, the frequency ratio will be n_1:$n_2 = g_2'$:g_1. Because of $\bar{\bar{q}}_1 = g_1\bar{q}_1 - g_2'\bar{q}_2$, the condition (4.10) will be met *exactly* if $\dot{\bar{\bar{q}}} = 0$; according to the Corollary of the preceding section we are then at a multiple root, and so we have simply a Case (II) c-a motion, and an orbit of permanent Lissajous shape. At a single root, according to the same corollary, $\dot{\bar{\bar{q}}}_1(t_e) \neq 0$ holds so that the frequencies cannot be exactly commensurate at the ratio g_2':g$_1$, nor can they remain at any constant ratio for long. Still, the ratio g_2':g_1 will hold approximately for some time, and the orbit will still resemble a Lissajous figure for a while. We may speak loosely of a "Lissajous pattern," meaning a curve which is not necessarily closed or exactly periodic, but still lies in the neighborhood of a specific Lissajous figure for some time. As time passes, the pattern may be lost rapidly, depending on the evolution of amplitudes and phases. At the next modulation extremum, however, the amplitudes are again stationary and another phase relation (4.8) is in effect; hence in general the orbit alternates between Lissajous patterns which are most clearly recognizable near the modulation extrema. Depending on conditions, a pattern may also persist, of course; typical are the motions of an elastic pendulum in the vicinity of a cup or cap c-a motion which never lose their approximate cup or cap shape.[3] On the other hand, in between extrema, a well-formed pattern can hardly arise, both amplitudes and periods being variable.

Suppose now that we have amplitude modulation between two roots of equal parity. Then r remains the same throughout, $r = r'$, and with it the weighted relative phase in (4.9) remains the same, whether or not g_1 and g_2' are relatively prime. The pattern near a half-period $t \simeq T/2$ of the modulation must consequently be the same as near $t \simeq 0$, only it will be traced at a different amplitude ratio.

With modulation between roots of different parities, suppose for definiteness that we start at the even root and with $r' = 0$. At $t = T/2$ we must have $\bar{\bar{q}}_1(T/2) = \pm\pi$, and the Lissajous pattern will obviously be different. Continuing, we shall have $\bar{\bar{q}}_1(T) = \pm 2\pi$, but does the orbit return to the original pattern it had around $t \simeq 0$? Noting that $\bar{q}_1$ and $\bar{q}_2$ are determined only mod 2π, we must conclude that the weighted relative phase $n_2 d_1 - n_1 d_2$, too, is determined only mod 2π and not modulo some *lesser* multiple of π. If it changes by 2π, then the combined phase $\bar{\bar{q}}_1$ goes through a corresponding change by $2d\pi$, but if conversely $\bar{\bar{q}}_1$ changes by 2π, then $n_2 d_1 - n_1 d_2$ changes only by $2\pi/d$. We must, therefore, distinguish $d = 1$ and $d \neq 1$.

If g_1 and g_2' are relatively prime, the pattern at $t \simeq T$ will be the same as at $t \simeq 0$ because the weighted relative phase has remained unchanged (mod 2π). At $t \simeq 3T/2$ the pattern will be the same as at $t \simeq T/2$, and so on. Thus there is a pair of different, typical, alternating patterns for each even–odd modulation.

Quite different if $d \neq 1$. At $t \approx T$, $\bar{q}_1$ has changed by 2π, but $n_2 d_1 - n_1 d_2$ only by $2\pi/d$ so that the pattern cannot be the same as at $t \approx 0$. Similar for the odd patterns occurring at $T/2$, $3T/2$ and so on. Thus now there is an alternation of $2d$ patterns, until $n_2 d_1 - n_1 d_2$ has run through a change by a full 2π. The modulation period is still T, but the recurrence time of the orbit patterns will be $2dT$.

How many distinct patterns are possible in a given system will depend on the pair of integers g_1, g_2', i.e., on the resonance but not on the coupling. If $g_1 = 1$ so that only two values $r' = 0,1$ are possible, there are obviously just one even and one odd pattern. In the elastic pendulum,[3] these are simply cup- and caplike, respectively. If both g_1, g_2' are greater than 1 but relatively prime, there is still only one even and one odd pattern; the different possibilities for r' then correspond to different points on the pattern where the motion is started up at $t = 0$. If $d \neq 1$, there are more patterns, however. For example, in the simplest case $g_1 = g_2' = 2$, with $r = 0,1,2,3$ possible, there are two even patterns in the form of straight-line motions like the arms of a St. Andrew's cross, and two odd patterns in the form of the same ellipse described in opposite senses. Reference 4 is an example (with an added symmetry degeneracy leading to a $\times$ cross and a circle).

Many Lissajous figures are not closed loops but curved line segments with motion reversals at the two ends. At such endpoints the phases $\bar{q}_1$ and $\bar{q}_2$ go simultaneously through multiples of π. Release from rest necessarily takes place at an orbit endpoint, whatever the pertaining value of r, cf. Sec. III E. Of course, the phases of a $\delta = 0$ state, even if it has been set up by release from rest, do not remain precisely synchronized to the amplitude modulation later on, so that at a later amplitude extremum the orbit will only exceptionally backtrack precisely on itself, but whenever both phases go nearly through multiples of π near an extremum, there will be a cusplike or looplike return motion with, characteristically, a reversal of the sense of motion around the origin. Such reversals are possible because the restoring force in the system is in general not exactly central. Angular momentum about the origin is then not conserved and can indeed change so much that the angular velocity reverses as described. Reference 4 contains typical examples.

When a system has more than two d.f. there are generally no closed orbits. Arguments similar to the above can still be valid for two-dimensional projections of the orbit, however, given that c-a motions usually still exist and will induce two-dimensional patterns.

C. Approach to stable c-a motion at a double root

We return to an arbitrary system of any number of d.f.

When initial conditions are changed such that the modulation range between two single roots R_1 and R_2 tends to zero, the motion gradually approaches an orbitally stable c-a motion of Case (I) or (II) at a double root. Between the roots, the polynomial f can be approximated by a parabola opening downward, say

$$f(\bar{p}_1) = K(\bar{p}_1 - R_1)(R_2 - \bar{p}_1), \quad K > 0.$$

The integral in the formula SF (3.11) for the modulation period then becomes elementary and yields at once

$$T = 2\pi/K. \tag{4.11}$$

In the limit $R_1 \to R_2$, K becomes in essence the curvature of f at the top, and this is finite.

The limit process is thus straightforward for two simple roots of *equal* parity: Going towards a Case (II), the modulation does not become infinitely slower; it only narrows to zero while $\bar{q}_1(t)$ tends to the constant value $r'\pi$ (and in a physical oscillatory system of two d.f. the orbit would tend to a stable Lissajous figure).

For roots of *opposite* parity, $\bar{\bar{q}}_1(t)$ must still change by π in every interval $T/2$, with T remaining finite. It is not to be seen immediately how such a steady variation of the combined phase can go with constant amplitudes. The approach to a Case (I) motion thus requires special attention. It will be clarified in the following section in Theorems 3–6.

It is also possible that a Case (I) motion at a double skew root be reached as the limit of motion between a single root of *definite* parity and a single *skew* root. Equation (4.11) still applies, but a *single* skew root marks Lipschitz-singular conditions of exceptional low amplitude which demand special study; we must defer this exceptional case until Sec. IV E.

D. Motion at and near a multiple skew root

For ease of discussion we generally assume that the given system is one of physical oscillators, and we use the sign convention of Sec. III C.

1. Phase equations

For c-a motion $\bar{\bar{p}}_1 \equiv R$ at a multiple root R of f, the $n-1$ phase equations SF (4.1) become

$$\dot{\bar{q}}_i = u_{i0} + u_{i1} \cos \bar{\bar{q}}_1, \quad i = 2,\ldots,n, \tag{4.12}$$

where

$$u_{i0} = \omega_i + \left.\frac{\partial \bar{\bar{B}}(\bar{\bar{p}}_1,\alpha)}{\partial \alpha_i}\right|_{\bar{\bar{p}}_1 = R}, \tag{4.13}$$

$$u_{i1} = \left.\frac{\partial \bar{\bar{F}}(\bar{\bar{p}}_1,\alpha)}{\partial \alpha_i}\right|_{\bar{\bar{p}}_1 = R}, \tag{4.14}$$

with all u_{i0}, u_{i1} therefore constant in time. Likewise the equation of motion for the combined phase becomes

$$\dot{\bar{\bar{q}}}_1 = \partial \bar{\bar{S}}/\partial \bar{\bar{p}}_1 = v_0 + v_1 \cos \bar{\bar{q}}_1, \tag{4.15}$$

where

$$v_0 = \epsilon + \left.\frac{\partial \bar{\bar{B}}(\bar{\bar{p}}_1,\alpha)}{\partial \bar{\bar{p}}_1}\right|_{\text{at } R}, \tag{4.16}$$

$$v_1 = \left.\frac{\partial \bar{\bar{F}}(\bar{\bar{p}}_1,\alpha)}{\partial \bar{\bar{p}}_1}\right|_{\text{at } R}, \tag{4.17}$$

with v_0, v_1 constant.

Any nonresonant d.f. with $g_i = 0$ (but $i \neq 1$ because of the numbering convention $g_1 > 0$) also have the phase equation (4.12), but do not take part in the combined phase $\bar{\bar{q}}_1$; therefore, their integrated phases $\bar{q}_i$ cannot be tied to the resonant phases by some synchronization condition. Their amplitudes are necessarily constant, regardless of phase variability; see SF Sec. IV. If such an amplitude should be zero, it could also nullify the entire nonlinear coupling in some systems. It is to be understood that the language in the following excludes the possibility that $\bar{\bar{F}} \equiv 0$ for all $\bar{\bar{p}}_1$.

In a Case (II), the cosine is constant $\equiv \pm 1$; the phases are then seen to be linear functions of time, with all n periods $\dot{\bar{q}}_i$ constant and given by SF (5.7) together with SF (5.8), which is merely the synchronization condition (2.10) in disguise. The periods of any participating nonresonating d.f. are also constant, only unrelated.

Case (I), with $\bar{\bar{F}} \equiv 0$ as a function of time, is quite different. The (resonant) phases are not necessarily synchronized, as the discussion leading up to the theorem with the condition (2.10) clearly shows. The phase behavior can, in fact, be understood only with proper regard to both root parity and orbital stability of the motion. In SF we did not yet possess the requisite concepts; thus, in some places the preliminary discussion of Case (I) [but not of Case (II)] given there needs to be amended, as will be mentioned explicitly after Theorem 6 below.

The possibility of frequency variations in Case (I) c-a motion evidently depends on the vanishing, or otherwise, of the coefficients v_1 in Eq. (4.15) and u_{i1} in Eqs. (4.12). It so happens that these coefficients can vanish only in patterns which depend markedly on the multiplicity of R as a root of $\bar{\bar{F}}^2$ (which can be higher than its multiplicity as a root of f, depending on the multiplicity as a root of $\bar{\bar{P}}$). Thus the phase behavior of a Case (I) turns out to depend sensitively on the type of coupling which operates in the given system.

2. Higher-order roots

Assume first that R is a root of $\bar{\bar{F}}^2$ of the third or higher order. A glance at the explicit Eq. (3.5) shows that in this case, either one l_i equals at least 3, or $\bar{\bar{Q}}$ has a root of at least the second order at R, or a confluence of zeros of sufficiently high order takes place. In any event, a derivative of $\bar{\bar{F}}$ with respect to either $\bar{\bar{p}}_1$ or any one of the α_i must still vanish at R; hence

Theorem 1: If R, a root of f, is a root of $\bar{\bar{F}}^2$ of at least the third order, then $v_1 = 0$ and $u_{i1} = 0$ for all $i = 2,...,n$ (including the nonresonant d.f.).

It follows from Eqs. (4.12) that in the c-a motion at such an R the d.f. $q_2,...,q_n$ move harmonically with the frequencies u_{i0}, which Eq. (4.13) shows to be identical with SF (5.3). One or more amplitudes may vanish, of course. The combined phase (4.15) obeys

$$\dot{\bar{\bar{q}}}_1 = g_1\dot{\bar{q}}_1 + \cdots + g_n\dot{\bar{q}}_n = v_0, \tag{4.18}$$

where the definition (2.3) has been used. Thus q_1 also moves harmonically, with its frequency $\dot{\bar{q}}_1$ determined from Eq. (4.18) by the value of v_0, as stated in equivalent terms in SF Sec. V. Since $v_0 \neq 0$ in general, integration of Eq. (4.18) will normally not yield a result equivalent to Eq. (2.10); there is no such synchronization now, except, of course, if the Case (I) is simultaneously a Case (II), with R being a root of f of order at least three, in which case it is seen from Eqs. (4.16) and SF (5.5) that indeed $v_0 = 0$ holds, and also vice versa. In summary

Theorem 2: If R, a root of f, is a root of $\bar{\bar{F}}^2$ of at least the third order, then the c-a motion at R is purely harmonic in all d.f., the frequencies of the resonant d.f. are always related by Eq. (4.18), but phase synchronization as in Eq. (2.10) is obtained if and only if the motion is simultaneously a Case (II).

This theorem completely describes all Case (I) c-a motions at roots of f of order three and higher, including the behavior of nonresonant d.f. Note, however, that it also describes motions at those skew roots of f which are only double but are still at least triple roots of $\bar{\bar{F}}^2$. Such motions are necessarily orbitally stable, according to STAB (3.12); see also Fig. 3 above. Their special interest is that they may be Liapunov-unstable in a higher approximation only, as we showed in STAB Sec. IV by an argument which rested decisively on the c-a frequencies's being constant (and equal to the linearized ω_i because also $\bar{\bar{B}} \equiv 0$).

We have not attempted in Theorem 2 to characterize the phase behavior of motions in the *neighborhood* of the c-a motion at R, because there are too many possibilities of higher-order roots and of ways of approaching them. Each case will have to be studied on its own terms. Lemma (4.1) and Theorems 5 and 6 below point to suitable procedures. The only conceivable mathematical difficulty is an exceptional low-amplitude condition, and we deal with the main features of that in Sec. IV E.

3. Lowest-order roots

We are now left to treat double roots of $\bar{\bar{F}}^2$, which can only be double (skew) roots of f. Referring again to Eq. (3.5), it is seen that a first-order zero of $\bar{\bar{F}}$ at R can arise because $l_i = 2$ for one particular i, say $i = j$, or because $\bar{\bar{Q}}$ has a simple zero at R. However, this is not all.

In certain systems, according to the Theorem at the end of Sec. III D, there also exists the further possibility of a confluence of two exceptional low-amplitude conditions, i.e., a confluence of two zeros of order $\frac{1}{2}$ in Eq. (3.5). If this is a root of f, the corresponding c-a motion again needs special study. Its dynamical neighborhood, as is demonstrated by Fig. 4, can in general contain a medley of different motions. First, there are orbitally unstable motions with two amplitudes having their lower bounds at zero, and therefore with Lipschitz conditions not guaranteed. Then there are stable motions of the three kinds listed in the previous section. They may take place between two roots of equal parity, including limiting Case (II) motions when the two roots coalesce, and then they have a constant or almost constant combined phase $\bar{\bar{q}}_1(t)$. They may also take place between two roots of opposite parity, and then they have a combined phase which changes by π in each half-period of the modulation. They may even occur between a single root of definite parity and a single skew root, with the combined phase behaving as will be discussed in Sec. IV E. To some extent these complications are academic precisely because of the low amplitudes: When two amplitudes tend to zero, then in the limit of the c-a motion with both of them identically zero the behavior of the combined phase becomes irrelevant if only the phases of the other (nonvanishing) d.f. behave in an unexceptional manner. Still, one may wish to understand the *approach* to the c-a motion in some detail. Any particular approach is equivalent to some (continuous) path in the space of the parameters $\alpha_2,...,\alpha_n, E$. It stands to reason that not every geometrically possible path is necessarily allowed in any specific, given system which may possess its own, specific, restrictive properties. Thus, no general statement seems possible, and we leave these cases to be studied *ad hoc* when they arise.

To return to the other double roots of $\bar{\bar{F}}^2$, suppose to begin with that there is one resonant d.f. (having $g_j \neq 0$) with

$$(g_j \bar{\bar{p}}_1 + \alpha_j)^1 = 0 \quad \text{for } \bar{\bar{p}}_1 = R = -\alpha_j/g_j \tag{4.19}$$

if $j \neq 1$, or simply $R = 0$ if $j = 1$. As in Sec. III C, using Eq. (3.6), we conclude that $\bar{p}_j = 0$. Since R is to be a double root of f, we even have $\bar{p}_j \equiv 0$. Furthermore, from the inequality (3.7) we conclude that R is an endpoint of the domain of $\bar{\bar{p}}_1$ (upper or lower according as $g_j < 0$ or > 0). Vice versa, if R lies at an endpoint, then (4.19) must hold for some resonant d.f. with $j \neq 1$, or $R = 0$ with $j = 0$.

We need all the derivatives (4.14) and (4.17). Note that $\bar{\bar{Q}}$ in Eq. (3.5) cannot vanish at R because the given zero (4.19) is single, by hypothesis. Now if $j \neq 1$, it is seen directly from the definition (3.5) that $u_{j1} \neq 0$ while $u_{i1} = 0$ for all resonant d.f. with $i \neq j$; if we write $\bar{\bar{F}} = \bar{\bar{p}}_1{}^{1/2}(g_j \bar{\bar{p}}_1 + \alpha_j) G(\bar{\bar{p}}_1, \alpha)$, we find by differentiation with respect to $\bar{\bar{p}}_1$ that simultaneously $v_1 \neq 0$ holds. If $j = 1$, then $R = 0$, and if we write $\bar{\bar{F}} = \bar{\bar{p}}_1 G(\bar{\bar{p}}_1, \alpha)$, we find that all resonant $u_{i1} = 0$ while again $v_1 \neq 0$. In either case, the nonresonant d.f. with $g_i = 0$ have $u_{i1} = 0$.

On the other hand, if R lies *inside* the domain, none of the resonant d.f. can have a zero amplitude and we must have $\bar{\bar{Q}}(R, \alpha) = 0$. Conversely, a zero of $\bar{\bar{Q}}$, if it is a *simple* zero of $\bar{\bar{F}}$, cannot lie at an endpoint of the domain because additionally either (4.19) or $R = 0$ would be required, in order to reach equality in one of the bounds (3.7). As for the derivatives of $\bar{\bar{F}}$ at R, begin with v_1. If in Eq. (3.5) we differentiate factor by factor with respect to $\bar{\bar{p}}_1$, at R each resultant term containing $\bar{\bar{Q}}$ vanishes, and only the term with $d\bar{\bar{Q}}/d\bar{\bar{p}}_1$ needs to be considered. This can vanish if and *only* if $\bar{\bar{Q}}$ has a double zero at R, contrary to hypothesis; thus always $v_1 \neq 0$. As for $u_{21}, \ldots, u_{n1}$, any or all of them may vanish or may differ from zero, for resonant and nonresonant d.f. alike; it all depends on the given $\bar{\bar{F}}$, especially the structure of $\bar{\bar{Q}}$, and hence on the interplay of different resonant terms in the nonlinear interaction governing the given system.

In summary

Theorem 3: Let R be a root of f which is a double root of $\bar{\bar{F}}^2$ and does not result from the confluence of two exceptional low-amplitude conditions. If R lies at an endpoint of the domain of $\bar{\bar{p}}_1$, then in the c-a motion at R exactly one resonant d.f. is at rest, say $\bar{p}_j \equiv 0$, while $v_1 \neq 0$ and also $u_{j1} \neq 0$ if $j \neq 1$; all other $u_{i1} = 0$. If R lies inside the domain, then in the c-a motion at R no resonant d.f. can be at rest; $v_1 \neq 0$ holds always but $u_{21}, \ldots, u_{n1}$ may or may not differ from zero, depending on the given $\bar{\bar{F}}$.

4. Phase integration, general

By Theorems 1 and 3 the behavior of the coefficients $v_1, u_{21}, \ldots, u_{n1}$ is recognized to be unexpectedly complex. We proceed to integrate the phase equations explicitly for the cases covered in Theorem 3.

The integration of Eq. (4.15) is slightly complicated by the existence of a singular integral

$$\cos \bar{\bar{q}}_1 = -v_0/v_1 = \text{const.} \tag{4.20}$$

Stability also enters the picture. From the criterion STAB (3.12) it is seen that c-a motion at the given R will be orbitally stable iff

$$|v_0| > |v_1|, \tag{4.21}$$

and unstable for the opposite sign (with equality impossible at a *double* root R). Since in the physical system the cosine of a phase cannot be larger than 1 in amount, the singular integral cannot apply in the stable case, but will have to be considered if c-a motion at R is orbitally unstable.

In the stable case (4.21) the (elementary) integration of Eq. (4.15) yields

$$\bar{\bar{q}}_1(t) = 2 \arctan \left\{ \frac{v_0 + v_1}{(v_0^2 - v_1^2)^{1/2}} \tan \left[\tfrac{1}{2}(v_0^2 - v_1^2)^{1/2}(t - t_0) \right] \right\}. \tag{4.22}$$

This is evidently periodic with full period $T = 2\pi(v_0^2 - v_1^2)^{-1/2}$ [which is finite, and must be equivalent to the value (4.11)]. It changes monotonically from zero at $t = t_0$ to $+\pi$ or $-\pi$ half a period later; from there, monotonic change continues because under the condition (4.21) the derivative in Eq. (4.15) cannot change sign, and so $\bar{\bar{q}}_1$ exhibits precisely the behavior required by the Theorem of Sec. IV A for the combined phase of a motion between two single roots of opposite parity.

In the unstable case, condition (4.21) with the sign reversed, integration of Eq. (4.15) yields

$$\bar{\bar{q}}_1(t) = 2 \arctan \left\{ \frac{v_1 + v_0}{(v_1^2 - v_0^2)^{1/2}} \tanh \left[\tfrac{1}{2}(v_1^2 - v_0^2)^{1/2}(t - t_0) \right] \right\}. \tag{4.23}$$

This is a monotonic function with asymptotes at

$$\tan \tfrac{1}{2} \bar{\bar{q}}_1(\pm \infty) = \pm \frac{v_1 + v_0}{(v_1^2 - v_0^2)^{1/2}}. \tag{4.24}$$

By means of the trigonometric formula $\cos x = [1 - \tan^2(x/2)]/[1 + \tan^2(x/2)]$, (4.24) can be converted to

$$\cos \bar{\bar{q}}_1(\pm \infty) = -v_0/v_1, \tag{4.25}$$

which coincides with the singular integral (4.20). Which of the two integrals applies at an unstable R, (4.23) or the singular (4.20), will depend on initial conditions.

With unstable c-a motions one has in addition the conceptual difficulty of never quite knowing how to set them up operationally. At best, they should be regarded as limiting cases of neighboring motions. For any unstable R there must exist some other root R' such that $f(\bar{\bar{p}}_1) > 0$ for $\bar{\bar{p}}_1$ lying between R and R'; cf. STAB Fig. 1. Consider now a motion with an initial amplitude in this range, and with the amplitude evolving towards R.

When $\bar{\bar{p}}_1(t) \to R$, the desired c-a motion will result after infinite time. During the approach to R the coefficients v_0, v_1 in the phase equation (4.15) are not yet rigorously constant so that the singular integral (4.20) is not valid, but to a first approximation the combined phase should be represented by the integral (4.23) during some time interval and with some t_0. After a while, the approximation will have to be renewed, of course, but the process clearly tends to the singular limit (4.25).

It is still unrealistic to try and set up a modulation evolving towards an *exact double* root R. The faintest inaccuracy in meeting the initial conditions will split R into two single roots R_1 and R_2, say, with R_1 lying between R_2 and the above R'. The modulation will then evolve towards R_1, very slowly, with the combined phase again representable by (4.23) and therefore again evolving toward the limit (4.25). Close to R_1, relatively rapid phase change will take over in accordance with Lemma (4.1), and consistent with the synchronization (2.10) required as $\bar{\bar{p}}_1$ passes through R_1. This phase change will proceed monotonically towards an asymptote of (4.23), which then remains valid approximately for some time as the amplitude still remains close to R_1 during its subsequent evolution towards R'. Except for this time interval around $f(R_1) = 0$ the phase is again approximated by the singular value (4.25), as long as the motion remains in the dynamical neighborhood of the c-a motion. For reference, we summarize the salient feature of this neighborhood thus:

Theorem 4: Let R be a root of f which is a double root of $\bar{\bar{F}}^2$ and does not result from the confluence of two exceptional low-amplitude conditions. If the c-a motion at R is orbitally unstable, every motion in its phase space neighborhood tends towards the singular behavior (4.25) of the combined phase, for $\bar{\bar{p}}_1$ either approaching R, or approaching one of the single roots from which R is being formed by confluence, or developing away from such a root, except for relatively short time stretches while $\bar{\bar{p}}_1$ is passing through such a root.

5. Neighborhoods of motions at lowest-order roots

We now turn to a root of the type covered by Theorems 3 and 4, and assume first that it lies at an endpoint of the domain of $\bar{\bar{p}}_1$. Without loss of generality we may assume that it is $R = 0$, and hence, that q_1 (or $\bar{\bar{p}}_1$) is the one d.f. that must remain at rest. Indeed, if the resonant d.f. at rest is some other q_j (or $\bar{p}_j$) with $j \neq 1$, and which also has $u_{j\,1} \neq 0$ according to Theorem 3, then we may simply renumber the d.f. so that this q_j becomes the new number one, and in the process the old $u_{j\,1}$ becomes the new v_1 whereas all the new $u_{i\,1}$ must vanish because no resonant d.f. q_i (or $\bar{p}_i$) vanishes besides the new q_1 (or $\bar{\bar{p}}_1$).

It now becomes clear how a stable Case (I) c-a motion with $\bar{\bar{p}}_1 \equiv 0$ can arise through the coalescence of two single roots of opposite parity: All d.f. apart from q_1 approach harmonic motion at constant frequencies $u_{i\,0}$, while $\bar{q}_1$, owing to $v_1 \neq 0$, takes on the entire variability required to make the combined phase $\bar{\bar{q}}_1$ vary by 2π per period as it should. The resultant $q_1(t)$ is thoroughly anharmonic but becomes progressively irrelevant as its amplitude $\bar{p}_1 = g_1 \bar{\bar{p}}_1$ shrinks to zero. In the limit of root coalescence the phase synchronization (2.10) also becomes irrelevant, again on account of the vanishing amplitude $\bar{\bar{p}}_1$, and leaves the remaining $n - 1$ d.f. to follow their harmonic motions with the frequencies $u_{i\,0}$ unrelated, and with arbitrary phase constants, as stated in equivalent terms in SF (5.3).

In the unstable case, the approach to the c-a motion is subject to Theorem 4. When the motion is close (in phase space) to the exact c-a motion, it is again only $q_1(t)$, which can have a notably variable phase, and only during the relatively short time stretches while the modulation $\bar{\bar{p}}_1$ goes through a minimum at a single root R_1; since R_1 is by hypothesis very close to an R_2 (of opposite parity) with which it is to coalesce into $R = 0$, this phase behavior again becomes irrelevant in the unstable limit $R_1 = R_2$. The other d.f. again move harmonically as long as $\bar{\bar{p}}_1$ remains close enough to zero. In summary:

Theorem 5: Let R be a root of f which is a double root of $\bar{\bar{F}}^2$ and lies at an endpoint of the domain of $\bar{\bar{p}}_1$ but does not result from the confluence of two exceptional low amplitude conditions. The c-a motion at R, with exactly one resonant d.f. vanishing, is purely harmonic in all remaining d.f. In the course of the approach to the c-a motion the phase function of the vanishing resonant d.f. assumes in the stable case the entire variability required to account for the behavior of the combined phase (4.22), whereas in the unstable case it tends to harmonic behavior and remains approximately synchronized to the other resonant d.f. by the condition (4.25), except for relatively short stretches of time while the pertaining amplitude modulation goes through a minimum close to zero with attendant phase synchronization (2.10); in either case this phase becomes irrelevant when the c-a limit is reached, and then all other phase constants are arbitrary.

These results go beyond the statements made in SF inasmuch as the *approach* to the exact c-a motion is clarified.

Secondly, let R lie inside the domain of $\bar{\bar{p}}_1$. Since in the c-a motion at R according to Theorem 3 no resonant d.f. can be at rest, and always $v_1 \neq 0$, at least one resonant d.f. has an anharmonic phase function, and so have all d.f. with $u_{i\,1} \neq 0$, resonant or not. We can eliminate the cosine between Eqs. (4.12) and (4.15) to yield

$$u_{i\,1}\dot{\bar{\bar{q}}}_1 - v_1\dot{\bar{q}}_i = u_{i\,1}v_0 - u_{i\,0}v_1 = -w_i,$$

and integrate to

$$v_1\bar{q}_i = u_{i\,1}\bar{\bar{q}}_1 + w_i t + \text{const}, \tag{4.26}$$

where $\bar{\bar{q}}_1$ must be of the form (4.22) or (4.23). Thus the phase functions of all d.f. with $u_{i\,1} \neq 0$, and possibly of q_1, have essentially the same time dependence; in the stable case they are anharmonic with the superimposed period T of the function (4.22), whereas in the unstable case the most important feature is again the tendency towards almost harmonic behavior described by Theorem 4. In short:

Theorem 6: Let R be a root of f which is a double root of $\bar{\bar{F}}^2$ and lies inside the domain of $\bar{\bar{p}}_1$ (and is therefore a single root of $\bar{\bar{Q}}$). In the c-a motion at R at least one resonant d.f. has an anharmonic phase function, and so have all d.f. with $u_{i\,1} \neq 0$, by Eq. (4.26); the remaining d.f. move harmonically. In the course of the approach to the c-a motion none of the anharmonic phases becomes irrelevant, but in the unstable case approximately harmonic behavior occurs in accordance with Theorem 4.

These "double skew root inside the domain" motions are thoroughly exceptional. We overlooked their existence in SF. They are the one and only exception to the statement made there that *all* c-a motions are purely harmonic, and to the implication that in Case (I) the phase functions are *always* unrelated.

A motion with constant amplitudes but a nonlinear phase evolution is, of course, not stationary in the customary sense. Suppose for the sake of illustration that we have only two d.f., that we have one of these double skew roots inside the domain, and that also $u_{21} = 0$. Then at this root q_2 moves harmonically, but because of $\bar{\bar{q}}_1 = g_1\bar{q}_1 + g_2\bar{q}_2$ the phase $\bar{q}_1$ consists of a linear part together with either a periodic part (4.22) or a monotonic but nonlinear part (4.23). We may paraphrase: q_1 moves at a constant frequency but with a (periodically or monotonically) changing phase constant. Thus, in the Lissajous picture the orbit will start as a particular pattern which gradually evolves through other patterns as the phase difference $d_1 - d_2$ in Eqs. (4.4) runs through its (periodic or monotonic) evolution. When such a motion is observed in a real system it need not stand out amongst the usual amplitude-modulated motions and may go unrecognized despite its mathematically different character.[5]

E. Phase behavior at a single skew root

A single skew root is a single root of $\bar{\bar{F}}^2$ and therefore entails that in Eq. (3.5) one of the l_i equals 1. It lies at an endpoint of the domain of $\bar{p}_1$; see Sec. III C. Hence, when the modulation $\bar{p}_1$ reaches this root, the d.f. in question has its amplitude going through a zero. In SF Sec. VI we discussed in detail what happens in a real oscillatory system with a nonlinear coupling involving one d.f. to the first power, when the amplitude of that d.f. drops to zero. The main result was that the phase went through a 180° jump and in addition changed quite rapidly in the vicinity of the amplitude zero. For completeness we now demonstrate the existence of an analogous phase jump in all systems with Hamiltonians of the type (2.1).

Thus, let $\bar{\bar{F}}$ behave as in Eq. (3.9) at a Γ which is also a root of $\bar{\bar{P}}$ (of at least the first order). From the conservation law (3.1) we conclude that

$$\cos\bar{\bar{q}}_1 = \bar{\bar{P}}/\bar{\bar{F}} = \bar{\bar{F}}(\bar{\bar{P}}/\bar{\bar{F}}^2) \tag{4.27}$$

as long as $\bar{\bar{F}} \neq 0$. If now $\bar{\bar{p}}_1 \rightarrow \Gamma$, then $\bar{\bar{P}}$ approaches a root of at least the same order as the root of $\bar{\bar{F}}^2$, whence $\bar{\bar{P}}/\bar{\bar{F}}^2$ remains bounded and

$$\lim \cos\bar{\bar{q}}_1 = 0 \quad \text{as } \bar{\bar{p}}_1 \rightarrow \Gamma.$$

It follows that

$$\lim \sin\bar{\bar{q}}_1 = \pm 1 \quad \text{as } \bar{\bar{p}}_1 \rightarrow \Gamma. \tag{4.28}$$

Γ is an extremum of the modulation $\bar{\bar{p}}_1$. The derivative $\dot{\bar{\bar{p}}}_1(t)$ must therefore have opposite signs before and after the passage through Γ, but in Eq. (2.6) the factor $\bar{\bar{F}}$ does not change sign, hence $\sin\bar{\bar{q}}_1$ must change sign, and it can only do so in accordance with (4.28) if $\bar{\bar{q}}_1$ jumps by π. Thus, $\sin\bar{\bar{q}}_1(t_e -) = -1$ and $\sin\bar{\bar{q}}_1(t_e +) = +1$ or the other way around, depending on the sign of K_0, and on Γ being a minimum or maximum of $\bar{\bar{p}}_1$.

The equation of motion for $\bar{\bar{q}}_1(t)$ is of the form (4.15), and if we at once eliminate the cosine by means of (3.1) it becomes

$$\dot{\bar{\bar{q}}} = \epsilon + \frac{\partial\bar{\bar{B}}}{\partial\bar{\bar{p}}_1} + \frac{\partial\bar{\bar{F}}}{\partial\bar{\bar{p}}_1}\frac{\bar{\bar{P}}}{\bar{\bar{F}}}.$$

Here the derivative of $\bar{\bar{F}}$ has a square-root singularity at Γ, and formally no Lipschitz condition holds, but in analogy to the manipulation in (4.27) we can write

$$\dot{\bar{\bar{q}}}_1 = \epsilon + \frac{\partial\bar{\bar{B}}}{\partial\bar{\bar{p}}_1} + \frac{1}{2}\frac{\partial\bar{\bar{F}}^2}{\partial\bar{\bar{p}}_1}\frac{\bar{\bar{P}}}{\bar{\bar{F}}^2} \tag{4.29}$$

and conclude that $\dot{\bar{\bar{q}}}_1$ remains safely bounded as $\bar{\bar{p}}_1 \rightarrow \Gamma$, because both $\bar{\bar{B}}$ and $\bar{\bar{F}}^2$ are polynomials and $\bar{\bar{P}}/\bar{\bar{F}}^2$ is bounded. Except for the 180° jump, $\bar{\bar{q}}_1$ is therefore entirely regular; however, its transient behavior right and left of the singularity depends strongly on the coupling terms in the given Hamiltonian, as Eq. (4.29) indicates.

Under these circumstances, the phase behavior in each given case will require separate study. The elastic pendulum is typical; for a detailed comparison of phase transients in the real system and its slow-fluctuation approximation, see Sec. X of Ref. 3.

[1] M. F. Augusteijn and E. Breitenberger, J. Math. Phys. **21**, 462 (1980).

[2] M. F. Augusteijn and E. Breitenberger, J. Math. Phys. **22**, 51 (1981).

[3] E. Breitenberger and R. D. Mueller, J. Math. Phys. **22**, 1196 (1981).

[4] M. F. Augusteijn and E. Breitenberger, "Bifurcation in a complex-valued wave-field model" (to appear).

[5] We have found one case (in connection with "$N = 6$ phase anisotropy" waves) which we expect to publish in due course.

Structure results for the Segal quantization of Fermi systems

Franco Gallone and Antonio Sparzani
Istituto di Scienze Fisiche dell'Universitá, via Celoria 16, I-20133 Milano, Italy
Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Milano, Italy

(Received 17 December 1980; accepted for publication 19 June 1981)

In Segal's approach to linear Fermi quantum systems, a one particle picture with linear symmetries (e.g., with a linear dynamics) can be quantized very straightforwardly when a complexification is given for the (real linear) one particle picture. We examine how the symmetries that are embodied in the one particle picture can determine the structure of the family of the possible complexifications. Among other results, we prove that if the symmetries can be represented in a suitably irreducible way then the complexification is essentially unique. Also, when the one particle space is a generalization of the one defined by the Dirac equation, we prove that there are many complexifications, and inequivalent too as they generate inequivalent representations of the canonical anticommutation relations; however, we find two criteria that single out the "physical" complexification. We use the general results we prove to discuss a few familiar models.

PACS numbers: 03.65.Bz, 03.65.Ca, 11.10.Cd, 02.20. + b

1. INTRODUCTION

In Segal's approach to linear Fermi quantum systems, to quantize a system means to represent the coordinates and the conjugate momenta of a "one particle" or "classical phase space" picture in a C^*-algebra (sometimes called the Clifford algebra of the system) in such a way that the canonical anticommutation relations hold.[1] The one particle picture that underlies this approach can be represented by a triple (M,S,T), where

(a) the pair (M,S) is an even finite or infinite dimensional real Hilbert space, i.e., M is an even finite or infinite vector space and S is an inner product on it,

(b) T is a group homomorphism from a group G into the group of the orthogonal transformations of (M,S).

The picture described by (M,S,T) is quantized when a representation of the canonical anticommutation relations (c.a.r.) over (M,S) is given, i.e., a real linear injection R from M into the self-adjoint part of the Clifford algebra (which need not be a concrete algebra) such that $\forall m,m' \in M$,

$$[R(m), R(m')]_+ \equiv R(m)R(m') + R(m')R(m) = S(m,m').$$

Indeed, once R is given T also is quantized, since for each element g of G a unique automorphism τ_g of the Clifford algebra exists such that $\forall m \in M$

$$\tau_g(R(m)) = R(T(g)m).$$

Although the quantization of linear Fermi systems can be discussed in this purely algebraic framework,[2] we examine here the well known straightforward quantization procedure called "second" or "Segal" quantization,[3] in which the Clifford algebra is realized as an algebra of operators on a (complex) Hilbert space and the automorphisms τ_g are implemented by unitary operators on this space. Indeed, we examine in this paper "how many" Segal quantizations exist for a given one particle picture. For a discussion about the one particle picture as a kinematical description in which a symmetry group is also defined (which can embody a linear dynamical evolution), see Ref. 4. For a discussion of Segal quantization versus general algebraic quantization, see Ref. 5. Actually, Bose systems are examined in Refs. 4 and 5. However, the discussion to be found there can be easily adapted to the Fermi case treated in the present paper by replacing symplectic spaces with real Hilbert spaces, symplectic transformations with orthogonal transformations, Weyl systems with representations of the canonical anticommutation relations, Weyl algebras with Clifford algebras, and symmetric Fock spaces with antisymmetric Fock spaces.

Segal quantization of (M,S,T) can be performed when there exists a complex Hilbert space structure on M that embodies the real Hilbert space structure already existing on M and such that $T(g)$ is a complex unitary operator for each element g of G (of course, this "one particle" complex Hilbert space is *not* the complex Hilbert space where the second quantized quantities are defined). The existence of such a complex Hilbert space structure is equivalent to the existence of a linear operator J on the real Hilbert space M satisfying

(C1) J is an orthogonal operator on M, i.e., $\forall m,m' \in M$

$$S(Jm,Jm') = S(m,m'),$$

(C2) $J^2 = -\mathbb{1}$ (the identity operator on M),

(C3) $\forall g \in G$, $[T(g),J]_- \equiv T(g)J - JT(g) = 0$.

In fact, if the complex Hilbert space exists, J is simply multiplication by the imaginary unit. If conversely J is given, a complex Hilbert space M^J with the required properties is constructed defining on M a complex scalar multiplication as

$$\forall \alpha \in \mathbb{C}, \forall m \in M, \ \alpha m \equiv ((\mathrm{Re}\alpha)\mathbb{1} + (\mathrm{Im}\alpha)J)m$$

and a complex inner product as

$$\forall m,m' \in M, \quad (m|m')_J \equiv S(m,m') + iS(Jm,m').$$

We call an operator J on M with the properties (C1), (C2), and (C3) listed above a complexification operator of (M,S,T) and denote the set of such operators by $C(M,S,T)$. Different elements of $C(M,S,T)$ lead to different Segal quantizations of

0022-2488/82/091610-07$02.50

(M,S,T). Notice that in this very precise sense, a variety of "Fock quantizations" may exist, corresponding to the variety of elements of $C(M,S,T)$; each of them is a Fock quantization in its own right and is determined by an element of $C(M,S,T)$. They are comparable, though, as quantizations over the *same real* Hilbert space, and some criteria exist which may be used to check whether the Hilbert space representations of the c.a.r. over (M,S) they contain are unitarily equivalent. As we will see, in the cases we are concerned with the representations of the c.a.r. that arise are all unitarily inequivalent.

In Sec. 2 we show how the structure of $C(M,S,T)$, when this set is not empty, can be determined by the structure of T. In particular, we find the condition for T that makes $C(M,S,T)$ contain a unique (up to the sign) complexification operator, if any. When a choice among many admissible complexification operators is possible, it will also be shown that all the corresponding (Fock) representations of the c.a.r. are unitarily inequivalent, thus making meaningful the problem of singling out one of them. It is worth mentioning that Weinless too proved a uniqueness condition for complexification operators, which is, however, quite different from ours; indeed, he proved that if G is a Lie group and T is strongly continuous, then in $C(M,S,T)$ there is at most one complexification operator with respect to which the self-adjoint generator of a fixed (but generic) one-parameter subgroup of T is positive (Lemma 1.6 in Ref. 2). A few more facts about $C(M,S,T)$ are worth mentioning. First, at most uniqueness up to the sign can hold, because if an operator J is an element of $C(M,S,T)$, so is $-J$. Second, when $G = \mathbb{R}$ and T is strongly continuous, a condition for $C(M,S,T)$ not to be empty was found by Weinless (Theorem 5.5 of Ref. 2). Finally, the discussion of the one degree of freedom case, which is interesting for Bose systems (see Ref. 4), is not really instructive for Fermi systems. In fact, for $M = \mathbb{R}^2$ there is just one (up to the sign) operator that satisfies conditions (C1) and (C2) above.

In Sec. 3 we use the results obtained in Sec. 2 to discuss nonfamiliar features of a few familiar models.

2. STRUCTURE RESULTS

In this section we show what operators are contained in $C(M,S,T)$ in the case that $C(M,S,T)$ is nonempty and T satisfies particular conditions. Of the four results we prove, the first one is a uniqueness condition for the complexification operators and the third one bears a uniqueness condition as an easy consequence. The statements of the first three results have the following common pattern. If J is a complexification operator and T has some properties as a unitary representation of G in M^J, then $C(M,S,T)$ has some structure determined by J.

Before the lemmas and the results, it may be useful to notice explicitely that, if J and K are two complexification operators, a (real) linear operator B on M is a (complex) linear operator from the complex Hilbert space M^J into the complex Hilbert space M^K if $BJ = KB$; it is antilinear if $BJ = -KB$. In particular, J being a complexification operator, a (real) linear operator in M is a (complex) linear (or antilinear) operator in M^J if it commutes (or anticommutes) with J. As to the notation, remember that M^J denotes M "complexified" by a complexification operator J. Also, we denote by T^J the (complex) unitary representation of G that is defined by T in the complex Hilbert space, M^J, namely $\forall g \in G$

$$T^J(g) = T(g),$$

where the difference between the left hand side and the right hand side is just the difference between M^J and M.

Lemma 1: Let J and K be elements of $C(M,S,T)$ and let the (complex) unitary representation T^J be (complex) irreducible. The following statements hold true.

(a) A real number ϑ exists such that $[J,K]_+ = \vartheta\mathbb{1}$.

(b) $\forall m \in M, \|[J,K]_- m\|^2 = (4-\vartheta^2)\|m\|^2$, where ϑ is the number that appears in (a).

(c) If $[J,K]_- = 0$, then $K = \pm J$.

Proof: (a) Because of property (C1) of the complexification operators, $JK + KJ$ is a bounded real linear operator on M. It is also complex linear on M^J because $[JK + KJ, J]_- = 0$ as a result of (C2). Therefore, $JK + KJ$ is a bounded complex linear operator on M^J which commutes with T^J, because of (C3). Thus, by Schur's theorem, a complex number ϑ exists such that $JK + KJ = \vartheta\mathbb{1}$. Moreover $\forall m \in M$

$$\mathrm{Im}((JK+KJ)m|m)_J = -S(Km,m) - S(KJm,Jm) = 0,$$

since

$$S(Km,m) = S(K^2m,Km) = -S(m,Km) = -S(Km,m),$$

where use is made of (C1) and (C2). Hence ϑ is real. (b) A straightforward calculation, in which use is made of (C1), (C2), and of the result (a) proved above, shows this equality. (c) This result is an easily proved consequence of (a) above and of (C2).

To state our first result, we need a remark which is used also in the proof of Lemma 2. Indeed we observe that if a (complex) unitary representation of a group is irreducible, then its antiunitary commutant either is empty (if the representation is not self-contragredient, i.e., it is not antiunitarily equivalent to itself) or contains just antiunitary operators that differ from one another by a phase factor. This is a direct consequence of Schur's theorem. Therefore, all the squares of the antiunitary operators that commute with the representation have the same value, which can be called the square of the antiunitary commutant of the representation.

Result 1: Let J be an element of $C(M,S,T)$. If the (complex) unitary representation T^J is (complex) irreducible and either of the following two conditions holds, (1) T^J is not self-contragredient, (2) the square of the antiunitary commutant of T^J is *not* $-\mathbb{1}$, then $C(M,S,T)$ contains only the operators $\pm J$.

Proof: We show that, if in $C(M,S,T)$ there is an operator $K \neq \pm J$, then an antiunitary operator A exists on M^J that commutes with T and such that $A^2 = -\mathbb{1}$, which contradicts both (1) and (2). In fact, if such a K exists, take

$$A \equiv (4-\vartheta^2)^{-1/2}[J,K]_-,$$

which is a well defined (real) linear operator in M because of (b) and (c) of Lemma 1. Using (a) of Lemma 1 and (C2) we can see that $A^2 = -\mathbb{1}$. This directly implies that the range of A is M. Moreover A is isometric because of (b) of Lemma 1. Finally, A is an antilinear operator in M^J since $[A, J]_+ = 0$ follows from (C2). Therefore A is an antiunitary operator on M^J and commutes with T because of (C3). This ends the proof.

Notice that—for an irreducible unitary representation—the case not covered by the result above is that the representation is self-contragredient through an antiunitary operator whose square is $-\mathbb{1}$. Such are, for instance, the continuous unitary irreducible representations of SU (2). Result 2 shows that in this case there is not (up to the sign) uniqueness for the complexification operator, in contrast with the situation dealt with by Result 1. We prove now a lemma that we need in the proofs of both Results 2 and 3.

Lemma 2: Let J be an element of $C(M,S,T)$ and A an antiunitary operator on M^J that commutes with T and such that $A^2 = -\mathbb{1}$. If α, β, γ are real numbers such that $\alpha^2 + \beta^2 \gamma^2 = 1$, the (real) linear operator $J_{(\alpha,\beta,\gamma)}$ defined by

$$J_{(\alpha,\beta,\gamma)} \equiv \alpha J + \beta A + \gamma JA$$

is an element of $C(M,S,T)$. Moreover, the (complex) unitary representations T^J in the complex Hilbert space M^J and $T^{J_{(\alpha,\beta,\gamma)}}$ in the complex Hilbert space $M^{J_{(\alpha,\beta,\gamma)}}$ are unitarily equivalent if $\alpha \neq -1$, antiunitarily equivalent if $\alpha = -1$.

Proof: Notice that A, as a (real) linear operator on M, is an element of $C(M,S,T)$ and the relation $[J,A]_+ = 0$ holds. A straightforward calculation in which use is made of this anticommutation relation and the properties (C1) and (C2) of J and A shows that $\forall m,m' \in M$

$$S(J_{(\alpha,\beta,\gamma)}m, J_{(\alpha,\beta,\gamma)}m') = S(m,m'),$$

i.e., the condition (C1) holds for $J_{(\alpha,\beta,\gamma)}$. Direct computation, in which use is made of $[J,A]_+ = 0$ and the property (C2) of J and A, also shows that $J_{(\alpha,\beta,\gamma)}{}^2 = -\mathbb{1}$, i.e., the condition (C2) holds for $J_{(\alpha,\beta,\gamma)}$. Finally, (C3) obviously holds for $J_{(\alpha,\beta,\gamma)}$ since it holds for both J and A. This ends the proof of the first statement of the lemma. For the equivalence between T^J and $T^{J_{(\alpha,\beta,\gamma)}}$, consider first the case $\alpha = -1$; this means $J_{(\alpha,\beta,\gamma)} = -J$ and the identity map on M is an antiunitary operator from M^J onto M^{-J} that commutes with T, i.e., that transforms T^J into T^{-J}. Taking now $\alpha \neq -1$, define the (real) linear operator V on M as

$$V = (2(1+\alpha))^{-1/2}(\mathbb{1} - J_{(\alpha,\beta,\gamma)}J).$$

The range of the operator V is M, since the equality

$$(\mathbb{1} - J_{(\alpha,\beta,\gamma)}J)(\mathbb{1} + 2\alpha + J_{(\alpha,\beta,\gamma)}J) = 2(1+\alpha)$$

can be shown by direct computation, making use of $[J,A]_+ = 0$ and of property (C2) of J and A; further, using these properties of J and A and also (C1), it is possible to see that V is an orthogonal operator on M, and therefore it is isometric; finally, V is (complex) linear from M^J onto $M^{J_{(\alpha,\beta,\gamma)}}$ since

$$(\mathbb{1} - J_{(\alpha,\beta,\gamma)}J)J = J_{(\alpha,\beta,\gamma)}(\mathbb{1} - J_{(\alpha,\beta,\gamma)}J)$$

holds as a consequence of (C2) for both $J_{(\alpha,\beta,\gamma)}$ and J. Therefore V is a unitary operator from M^J onto $M^{J_{(\alpha,\beta,\gamma)}}$ and commutes with T since both $J_{(\alpha,\beta,\gamma)}$ and J do. This completes the proof of the lemma.

Result 2: Let J be an element of $C(M,S,T)$. If the (complex) unitary representation T^J is (complex) irreducible and self-contragredient through an antiunitary operator A on M^J such that $A^2 = -\mathbb{1}$, then these two properties hold for T^K for any $K \in C(M, S, T)$. Under these conditions, the set $C(M,S,T)$ coincides with the set of the operators

$$J_{(\alpha,\beta,\gamma)} \equiv \alpha J + \beta A + \gamma JA,$$

for α,β,γ real numbers such that $\alpha^2 + \beta^2 + \gamma^2 = 1$.

Proof: The only thing we must prove is that for any $K \in C(M,S,T)$ there are real numbers α,β,γ with the property $\alpha^2 + \beta^2 + \gamma^2 = 1$ such that $K = J_{(\alpha,\beta,\gamma)}$. When this is proved, the result follows as an easy corollary of Lemma 2. Take then $K \in C(M,S,T)$. If $K = \pm J$, then of course $(\alpha,\beta,\gamma) = (\pm 1,0,0)$. If $K \neq \pm J$, as in the proof of Result1 we can consider the operator $(4 - \vartheta^2)^{-1/2}[J, K]_-$, which is a (complex) antiunitary operator on M^J that commutes with T. Owing to the uniqueness up to a phase factor of an operator with these properties, which stems from the irreducibility of T^J, two real numbers ζ_1 and ζ_2 with the property $\zeta_1^2 + \zeta_2^2 = 1$ exist such that

$$[J,K]_- = (4 - \vartheta^2)^{1/2}(\zeta_1\mathbb{1} + \zeta_2 J)A.$$

Summing this equality with the equality $[J,K]_+ = \vartheta\mathbb{1}$ (see Lemma 1a) and using the property (C2) of J we obtain

$$-2K = \vartheta J + (4 - \vartheta^2)^{1/2}(\zeta_1 J - \zeta_2\mathbb{1})A.$$

Therefore $K = J_{(\alpha,\beta,\gamma)}$, with $\alpha = -\frac{1}{2}\vartheta, \beta = \frac{1}{2}\zeta_2(4 - \zeta^2)^{1/2}$, $\gamma = -\frac{1}{2}\zeta_1(4 - \zeta^2)^{1/2}$. This proves the result.

It is worth pointing out explicitly that, if we chose instead of A another operator in the antiunitary commutant of T^J, the family of operators $J_{(\alpha,\beta,\gamma)}$ would not be affected. So it must be for the statement of Result 1 to make sense and so it is because another operator would differ from A by just a phase factor. Also, notice that the conditions of both Results 1 and 2 are conditions for (M,S,T) and not really conditions that hold for a particular complexification operator only. This is explicitly stated in Result 2 and immediately seen in Result 1 (since T^J and T^{-J} are antiunitarily equivalent through the identity map on M).

We prove now our last structure result for $C(M,S,T)$. In contrast with Results 1 and 2, the representation T^J is not requested to be irreducible here. Indeed, this result is suited for discussing the one particle picture that is defined by the Dirac equation.

Result 3: Let an element J of $C(M,S,T)$ exist such that the (complex) unitary representation T^J decomposes into a (complex orthogonal) direct sum of two (complex) unitary irreducible representations. Suppose these two representations are mutually antiunitarily equivalent and unitarily inequivalent. Then there is a unique (up to a phase factor) antiunitary operator A on M^J that commutes with T and such that $A^2 = -\mathbb{1}$. Moreover, the operators

$$J_{(\alpha,\beta,\gamma)} \equiv \alpha J + \beta A + \gamma JA,$$

with α,β,γ real numbers such that $\alpha^2 + \beta^2 + \gamma^2 = 1$, are elements of $C(M,S,T)$ and the only operators of $C(M,S,T)$ that are not of this form are $\pm J_0$, with

$$J_0 \equiv J(P_1 - P_2)$$

if P_1 and P_2 denote the (complex) projections from M^J onto the supports of the two components of T^J. The unitary representations $T^{J_{(\alpha,\beta,\gamma)}}$ are unitarily equivalent to T^J, while T^{J_0} (respectively, T^{-J_0}) is the (complex orthogonal) direct sum of two copies of the component of T^J relative to P_1 (respectively, P_2).

Proof: Denote by M_1^J and M_2^J the ranges of the projections P_1 and P_2, respectively, i.e., the two mutually orthogonal (complex) subspaces of M^J that are invariant with respect to T^J, and by T_1^J and T_2^J the restrictions of T^J to M_1^J and M_2^J, respectively. If B is an antiunitary isomorphism from the (complex) Hilbert space M_1^J onto the (complex) Hilbert space M_2^J such that $\forall g \in G$

$$B\,T_1^J(g) = T_2^J(g)B,$$

then

$$A \equiv BP_1 - B^{-1}P_2$$

can be quite easily shown to be an antiunitary operator on M^J that commutes with T and such that $A^2 = -1$. Moreover, A is the unique (up to a phase factor) antiunitary operator on M^J that has these two properties; in fact, if V is another such operator, then VA^{-1} is a unitary operator on M^J that commutes with T; therefore, since T^J is a multiplicity free unitary representation, two complex numbers ρ_1 and ρ_2 of modulus one exist such that

$$VA^{-1} = \rho_1 P_1 + \rho_2 P_2;$$ [6]

thus we have

$$\begin{aligned} -1 = V^2 &= ((\rho_1 P_1 + \rho_2 P_2)(BP_1 - B^{-1}P_2))^2 \\ &= -(\rho_1\bar{\rho}_2 P_1 + \rho_2\bar{\rho}_1 P_2) \end{aligned}$$

(bar means complex conjugation), whence $\rho_1\bar{\rho}_2 = 1$; therefore $\rho_1 = \rho_2$ and this shows that V and A differ by a phase factor only. Thus, A is an operator on M that satisfies the conditions of Lemma 2, and therefore the operators $J_{(\alpha,\beta,\gamma)}$ are elements of $C(M,S,T)$ for all the real numbers α, β, γ such that $\alpha^2 + \beta^2 + \gamma^2 = 1$. The uniqueness up to a phase factor of A shows that if we defined the operators $J_{(\alpha,\beta,\gamma)}$ through a different antiunitary operator that meets the conditions of Lemma 2, we would obtain the same family of complexification operators. Also, from Lemma 2 we know that the unitary representations $T^{J_{(\alpha,\beta,\gamma)}}$ are unitarily equivalent to T^J. Notice that this holds also for $\alpha = -1$; in fact, for $\alpha = -1$, $T^{J_{(\alpha,\beta,\gamma)}}$ is known to be antiunitarily equivalent to T^J, but in the present discussion T^J is also known to be self-contragredient (through the antiunitary operator A).

Consider now the operator J_0 on M. Taking into account that if for two vectors m,m' of M the equality $(m|m')_J = 0$ holds, then $S(m,m') = 0$ holds as well, and also using the property (C1) of J, we see that $\forall m,m' \in M$

$$\begin{aligned} S(J_0 m, J_0 m') &= S((P_1 - P_2)m, (P_1 - P_2)m') \\ &= S(P_1 m, P_1 m') + S(P_2 m, P_2 m') \\ &= S((P_1 + P_2)m, (P_1 + P_2)m') = S(m,m'), \end{aligned}$$

i.e., J_0 satisfies the condition (C1). Observing now that J commutes with P_1 and P_2 because M_1^J and M_2^J are complex subspaces of M^J, it is very easy to show that J_0 satisfies the condition (C2) since so does J. Finally, J_0 satisfies the condition (C3) because so does J and because P_1 and P_2 commute with T as M_1^J and M_2^J are invariant with respect to T^J. Thus, $\pm J_0$ are elements of $C(M,S,T)$. To examine now the unitary representation T^{J_0} in the (complex) Hilbert space M^{J_0}, note that P_1 and P_2 commute also with J_0 and therefore M_1^J and M_2^J are (complex) subspaces also of the (complex) Hilbert space M^{J_0}. We denote them by $M_1^{J_0}$ and $M_2^{J_0}$ when we consider them endowed with the (complex) Hilbert space structure of M_{J_0}. Clearly, $M_1^{J_0}$ and $M_2^{J_0}$ are invariant with respect to T^{J_0} since M_1^J and M_2^J are invariant with respect to T^J. Also, $M_1^{J_0}$ and $M_2^{J_0}$ are easily seen to be mutually orthogonal (complex) subspaces of M^{J_0}, since the restrictions of J to M_1^J and of J_0 to $M_1^{J_0}$ coincide while the restriction of J to M_2^J is the opposite of the restriction of J_0 to $M_2^{J_0}$. For the same reason, the component of T^{J_0} relative to $M_1^{J_0}$ is unitarily equivalent (through the identity mapping on $M_1^J = M_1^{J_0}$) to T_1^J, while the component of T^{J_0} relative to $M_2^{J_0}$ is antiunitarily equivalent (through the identity mapping on $M_2^J = M_2^{J_0}$) to T_2^J. Therefore both the components of T^{J_0} are unitarily equivalent to T_1^J, since T_2^J is antiunitarily equivalent to T_1^J, and this proves what had to be proved for T^{J_0}. The same holds for T^{-J_0}, replacing T_1^J and T_2^J.

To complete the proof, we now have only to show that if K is an element of $C(M,S,T)$, then K is one of the operators $J_{(\alpha,\beta,\gamma)}$, $\pm J_0$ defined above. For such an operator K, $[J,K]_-$ is a bounded operator on M that anticommutes with J, and therefore $[J,K]_- A$ is a bounded operator on M that commutes with J, i.e., a (complex) linear bounded operator on M^J. Moreover, since $[J,K]_- A$ commutes with T^J, it must be a complex linear combination of P_1 and P_2 because T^J is a multiplicity free unitary representation. This means that there are complex numbers λ_1, λ_2 such that

$$[J,K]_- A = \lambda_1 P_1 + \lambda_2 P_2,$$ [6]

namely

$$[J,K]_- = \lambda_1 B^{-1}P_2 - \lambda_2 BP_1.$$

A direct computation, where use is made of the properties (C1) and (C2) of J and K, shows that $\forall m \in M^J$

$$([J,K]_- m|m)_J = 0,$$

and therefore

$$\begin{aligned} 0 &= \bar{\lambda}_1(B^{-1}P_2 m|m)_J - \bar{\lambda}_2(BP_1 m|m)_J \\ &= \bar{\lambda}_1(B^{-1}P_2 m|P_1 m)_J - \bar{\lambda}_2(BP_1 m|P_2 m)_J; \end{aligned}$$

this holds only if $\forall m_1 \in M_1^J$, $m_2 \in M_2^J$

$$(\bar{\lambda}_1 - \bar{\lambda}_2)(Bm_1|m_2)_J = 0,$$

which implies $\lambda_1 = \lambda_2$ since B is an antiunitary isomorphism from M_1^J onto M_2^J. Therefore, there is a complex number λ such that

$$[J,K]_- = \lambda(B^{-1}P_2 - BP_1).$$

Observe now that $[J,K]_+$ is a bounded operator on M that commutes with J, i.e., a bounded (complex) linear operator on M^J; moreover it is self-adjoint because

$$\mathrm{Im}([J,K]_+ m|m)_J = 0,$$

as a direct computation shows; therefore, since $[J,K]_+$ commutes with T^J and this unitary representation is multiplicity free, there are two real numbers μ_1, μ_2 such that

$$[J,K]_+ = \mu_1 P_1 + \mu_2 P_2.$$

Summing the two equalities obtained above for the commutator and the anticommutator of J and K, and using the property (C2) of J, we obtain

$$-2K = \lambda J(B^{-1}P_2 - BP_1) + J(\mu_1 P_1 + \mu_2 P_2); \qquad \text{(R)}$$

squaring this equality and again using the property (C2) of J, we obtain

$$\begin{aligned}-4\mathbb{1} = (&-\mu_1^2 + (\lambda\mu_1 - \lambda\mu_2)B - |\lambda|^2)P_1 \\ &+ (-\mu_2^2 + (\lambda\mu_2 - \lambda\mu_1)B^{-1} - |\lambda|^2)P_2,\end{aligned}$$

whence $-4P_1 = (-\mu_1^2 - |\lambda|^2)P_1 + \lambda(\mu_1 - \mu_2)BP_1$
and $-4P_2 = (-\mu_2^2 - |\lambda|^2)P_2 + \lambda(\mu_2 - \mu_1)B^{-1}P_2$.
Since the range of BP_1 is M_2^J and the range of $B^{-1}P_2$ is M_1^J, these last two equalities imply

$$4 = \mu_1^2 + |\lambda|^2 = \mu_2^2 + |\lambda|^2 \quad \text{and}\ \lambda(\mu_1 - \mu_2) = 0.$$

If $\lambda = 0$, then either $\mu_1 = \mu_2 = \mp 2$ or $\mu_1 = -\mu_2 = \mp 2$ and —as can be seen inserting these values in (R) above — either $K = \pm J$ or $K = \pm J_0$. If $\lambda \neq 0$, then $\mu_1 = \mu_2$, and — setting $\alpha \equiv -\mu_1/2, \beta \equiv \mathrm{Im}\lambda/2, \gamma \equiv -\mathrm{Re}\lambda/2$ and inserting these values in (R) above — we get

$$K = \alpha J + \beta A + \gamma JA;$$

the real numbers α,β,γ satisfy the relation

$$\alpha^2 + \beta^2 + \gamma^2 = \tfrac{1}{4}(\mu_1^2 + |\lambda|^2) = 1.$$

This shows that if K is not either $\pm J_0$, then it must be one of the complexification operators $J_{(\alpha,\beta,\gamma)}$. Thus the proof is concluded.

Observe that for the two antiunitarily equivalent components into which T^J decomposes in the statement of Result 3, to assume they are unitarily inequivalent (as in the statement of Result 3) is the same as to assume that neither of them is self-contragredient. Also, it is worth mentioning that the definition of J_0 in Result 3 is unambiguous because the decomposition of T^J into a (complex orthogonal) direct sum of irreducible components is unique, T^J being multiplicity free; thus, P_1 and P_2 are completely identified by their being the projections that decompose T^J into irreducibles.

Notice that T^{J_0} (respectively, T^{-J_0}) cannot be decomposed into a direct sum of irreducibles not both unitarily equivalent to the component of T^J relative to P_1 (respectively, P_2), because the decomposition of T^{J_0} (or T^{-J_0}) is unique up to unitary isomorphisms, T^{J_0} (or T^{-J_0}) being a factor representation. Therefore, for $K \in C(M,S,B)$, the way T^K decomposes into irreducibles can be used as a criterion to distinguish between $K = \pm J_0$ and K being one of the operators $J_{(\alpha,\beta,\gamma)}$. Indeed, we have the following uniqueness condition for the complexification operators of (M,S,T), when for (M,S,T) the conditions of Result 3 hold: The operators $\pm J_0$ are the only complexification operators that turn T into a unitary representation whose irreducible components are copies of the same unitary representation.

Another uniqueness condition that determines the complexification operators to be just $\pm J_0$ is found if the "symmetry group"

$$\{T(g), g \in G\}$$

of a one particle picture that satisfies the conditions of Result 3 is suitably extended. Indeed, define the "gauge transformations"

$$\{\exp \lambda J, \lambda \in \mathbb{R}\}$$

[observe that $\exp \lambda J$ is a well defined orthogonal operator on M since J is a bounded linear operator on M and $\exp \lambda J$ is naturally a (complex) unitary operator on M^J] and the "charge conjugation"

$$C \equiv BP_1 + B^{-1}P_2$$

(where B is as in the proof of Result 3), and consider the set

$$\hat{T} \equiv \{T(g), g \in G; \exp \lambda J, \lambda \in \mathbb{R}; C\}$$

of orthogonal operators on M. Clearly $C(M,S,\hat{T})$ is a subset of $C(M,S,T)$, and since the latter is known it is just a matter of direct computation to show that $C(M,S,\hat{T})$ contains the complexification operators $\pm J_0$ and nothing else. This can be seen also on the basis of Result 1, since $\hat{T}$ defines in M^{J_0} a family of (complex) unitary operators that can be easily shown to be irreducible , not self-contragredient and self-adjoint (and therefore Result 1, which depends on Schur's theorem, still applies). Observe that the Stone theorem self-adjoint generator in M^{J_0} of the gauge transformations is the "charge" operator $P_1 - P_2$, since

$$\exp \lambda J = \exp \lambda J_0(P_1 - P_2).$$

Since $CP_1 = P_2 C$ and $C^2 = 1$, C interchanges the two eigenspaces of the charge, and this explains the terms used in inverted commas.

We end this section proving a result concerning the unitary inequivalence of the representations of the c.a.r. over (M,S) associated with different complexification operators. To this aim we first need another result. This appears here as a lemma, but its scope is wider, as it provides a simple criterion for the unitary equivalence of the Segal representations of the c.a.r. that are associated with different complexification operators of (M,S). Quite naturally, we call complexification operators of (M,S) the operators on M that satisfy (C1) and (C2) and denote the set of them by $C(M,S)$. Obviously $C(M,S,T)$ is a subset of $C(M,S)$ for any "symmetry group" T.

Lemma 3: If J and J' are two complexification operators of $C(M,S)$, the corresponding representations of the c.a.r. over (M,S) are unitarily equivalent if and only if $J - J' \in \mathscr{J}_2(M)$, the set of the Hilbert–Schmidt operators on the real Hilbert space M.

Proof: A first step consists in constructing a unitary operator from M^J onto $M^{J'}$. For this purpose, take two orthonormal bases $\{u_\alpha\}$ and $\{u'_\beta\}$ in M^J and $M^{J'}$, respectively; as $\{u_\alpha, Ju_\alpha\}$ and $\{u'_\beta, J'u'_\beta\}$ are two orthonormal bases for the same real Hilbert space M, we can take the same index set for both the bases. Define V as the unitary operator from M^J onto $M^{J'}$ such that

$$Vu_\alpha = u'_\alpha \quad \text{and}\ VJu_\alpha = J'u'_\alpha.$$

Notice that in particular $VJ = J'V$. If R and R' are the representations of the c.a.r. corresponding to J and J', respectively, a unitary operator $\Gamma(V)$ exists from the Fock space over M^{J} onto the Fock space over $M^{J'}$ such that $\forall m \in M$

$$\Gamma(V)\,R(m)\,\Gamma(V)^{-1} = (R'\circ V)(m);$$

therefore R and $R'\circ V$ are unitarily equivalent (recall that whenever V is an orthogonal operator on M, $R'\circ V$ is a representation of the c.a.r. if R' is). Thus R and R' are unitarily equivalent if and only if R' and $R'\circ V$ are. At this point use can be made of a criterion to be found in Ref. 7: Here, it amounts to saying that R' and $R'\circ V$ are unitarily equivalent if and only if $VJ' - J'V \in \mathscr{J}_2(M)$; as $V^{-1}J'V = J$, this is equivalent to $J' - J \in \mathscr{J}_2(M)$. This ends the proof of the lemma.

As a side remark, notice that it also follows from the proof of the lemma that if R' is a representation of the c.a.r. associated with a given complexification operator J', then for any orthogonal operator V on M there is a complexification operator J such that $R\circ V$ is (unitarily equivalent to) the representation corresponding to J, and vice-versa. Indeed, $J = V^{-1}J'V$. In other words, the Fock space representations (if any) of the c.a.r. over (M,S) can be classified by the orthogonal operators on M. This result, as well as the idea underlying the proof of Lemma 3, can also be found in Ref. 8.

We can now prove the result on the inequivalence of the representations of the c.a.r.

Result 4: In the situations described in Results 1, 2, and 3, under the obvious proviso that the space involved are all infinite dimensional, all the complexification operators contained in $C(M,S,T)$ give rise to unitarily inequivalent representations of the c.a.r. over (M,S).

Proof: In view of Lemma 3, we just have to look at $J - J'$. To begin with, notice that J and $-J$ are immediately seen to be inequivalent. Moreover

$$J_{(\alpha,\beta,\gamma)} - J_{(\alpha',\beta',\gamma')} = (\alpha-\alpha')\,J + (\beta-\beta')A + (\gamma-\gamma')\,JA$$

is, up to the factor $[(\alpha-\alpha')^2 + (\beta-\beta')^2 + (\gamma-\gamma')^2]^{1/2}$, an isometric operator on M in both the situations of Results 2 and 3. Finally, $J_{(\alpha,\beta,\gamma)} - J_0$, in the situation described in Result 3, is not even a compact operator. This completes the proof.

The last result shows that in the situations described in this paper, when the complexification operator is not unique, its nonuniqueness is an essential one, since it leads to unitarily inequivalent representations of the c.a.r.

3. EXAMPLES

In this section we use the results obtained in the previous one to discuss two very well known models. Also if these models are so simple that little really new can be said about them, some features of them that are usually overlooked are pinpointed here. However, the real goal of this section is to show how the theory discussed in this paper should be used in the analysis of linear systems (e.g., linear quantum fields in an external field).

The first application of the results proved in the previous section is to the one particle space defined by the free Schroedinger equation. Here, the one particle symmetries are described by a complex irreducible unitary representation of a central extension of the Galilei group. In the Bose case the Segal quantization is shown to be unique by our result of Ref. 5, since the relevant representation is real irreducible.[9] As for the Fermi case, either Result 1 or Result 2 of the present paper must apply, since the relevant representation is complex irreducible; as it is not self-contragredient,[10] Result 1 applies and uniqueness (which here always means to within a sign) of the Segal quantization obtains.

As a second example, we examine the one particle space whose quantization leads to the free relativistic Dirac field. On a suitable space of solutions of the Dirac equation an action of the proper orthochronous Poincaré group $P^{\uparrow}_{+}$ can be defined in a natural way; as is well known, this defines a representation of $P^{\uparrow}_{+}$ that splits into two irreducible components. To describe them, we refer to the structure one is likely to consider at first, defining on the space of solutions an L^2-type Hilbert space structure, where multiplication by the imaginary unit is plainly defined pointwise. If we denote this complexification operator by J, the space of solutions by M, the real part of the L^2-type inner product by S, and the "natural" action of $P^{\uparrow}_{+}$ on M by T, we find ourselves exactly in the conditions of Result 3. Indeed, T turns with respect to J into the direct sum of the two continuous irreducible representations of $P^{\uparrow}_{+}$ determined by the pairs of Casimir operator eigenvalues $(m,1/2)$ and $(-m,1/2)$. As this situation has been recognized unsatisfactory very soon, owing to the presence of negative mass states, various (equivalent) schemes have been contrived to have the space of the "antiparticle" states transform according to a "physical" representation of $P^{\uparrow}_{+}$. The way of getting round this difficulty we can make the most of to illustrate our results consists in redefining the structure of the one particle space (rather than redefining things after quantization); this is achieved by changing the sign of the multiplication by the imaginary unit on the support of the component $(-m,1/2)$, which means changing J into J_0 (see, e.g., the second of Bongaart's papers quoted in Ref. 1). In fact this turns T into the"physical"sum of $(m,1/2)$ with itself. Notice that this is not *a priori* forbidden since the relevant representation is not irreducible and therefore Results 1 and 2 do not apply. One might reasonably ask whether this is the only possibility, i.e., how many complexification operators there are that give rise to a "physical" representation of $P^{\uparrow}_{+}$. Result 3 answers this question and shows that J_0 is the only choice.

A few more remarks on the free Dirac field. As we have just seen, if one constructs the "naive" complexification operator J as a first step, then he is unambiguously led to construct the "physical" J_0 as a second step if he wants both particle and antiparticle states to transform according to physical representations of $P^{\uparrow}_{+}$. If J_0 instead was adopted right at the outset, the uniqueness of J_0 rather than its construction from J would be the main point of interest.

For this purpose, observe that J_0 is uniquely determined by the remarks that follow Result 3. Further, notice that another uniqueness condition for J_0 can easily be derived from Weinless's uniqueness condition quoted in the Introduction. Finally, it is worth mentioning that, as a consequence of Result 4, the representations of the c.a.r. over the free Dirac one particle space corresponding to the "naive" and the "physical" complexification operators are unitarily inequivalent. This makes it relevant to have criteria, such as those described above, to make a choice between them.

[1] I. E. Segal, "Caractérization mathématique des observables en théorie quantique des champs et ses consequences pour la structure des particules libres," in *Les Problèmes Mathématiques de la Théorie Quantique des Champs, Lille 3–8 Juin 1957* (CNRS, Paris, 1959); I. E. Segal, *Mathematical Problems of Relativistic Physics* (American Mathematical Society, Providence, R. I., 1963); D. Shale and W. F. Stinespring, "States of the Clifford Algebra," Ann. Math. **80**, 365–381 (1964); P. J. M. Bongaarts, "The Electron-Positron Field, Coupled to External Electromagnetic Potentials, as an Elementary C^* Algebra Theory," Ann. Phys. **56**, 108–139 (1970); P. J. M. Bongaarts, "Linear Fields According to I. E. Segal," in *Mathematics of Contemporary Physics*, edited by R. F. Streater (Academic, London, 1972); see also Ref. 2.

[2] M. Weinless, "Existence and Uniqueness of the Vacuum for Linear Quantized Fields," J. Funct. Anal. **4**, 350–379 (1969).

[3] J. M. Cook, "The Mathematics of Second Quantization," Am. Math. Soc. Trans. **74**, 222–245 (1953). Second quantization is called Segal quantization in, e.g., M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II.

[4] F. Gallone and A. Sparzani, "Segal quantization of dynamical systems," J. Math. Phys. **20**, 1375–1384 (1979).

[5] F. Gallone and A. Sparzani, "On the uniqueness of the Segal quantization of linear Bose systems," J. Phys. A: Math. Gen. **14**, 1341–1350 (1981); see also F. Gazllone and A. Sparzani, "A uniqueness result for the Segal quantization of a classical system with symmetries," in *Group Theoretical Methods in Physics, Ninth International Colloquium, Cocoyoc*, 1980 (Springer, Berlin, 1980).

[6] The complex linear combinations of operators that appear here and in related formulas must be understood to be defined with respect to the complex vector space structure of M^J.

[7] D. Shale and W. F. Stinespring, "Spinor representations of infinite orthogonal groups," J. Math. Mech. **14**, 315–322 (1965).

[8] E. Balslev, J. Manuceau, and A. Verbeure, "Representations of anticommutation relations and Bogolioubov transformations," Commun. Math. Phys. **8**, 315–326 (1968).

[9] This follows for instance from Remark 1.5 in Ref. 2, since in the representation there is a one-parameter subgroup with strictly positive generator (proportional to the one-particle space energy).

[10] See, e.g., V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, New York, 1970), Vol. II, p. 244.

Measurement systems and Jordan algebras

John R. Faulkner[a)]
Mathematics Department, University of Virginia, Charlottesville, Virginia 22903

(Received 5 January 1982; accepted for publication 9 April 1982)

An axiomatization of the measuring process leads to a Jordan algebra structure on the observables. The novel features in this development include a proof of the existence of the sum of observables, a proof of the quadratic nature of the square of an observable, the lack of a finite dimensionality assumption, and the exploitation of the change in the measuring process due to a change in the counting observable.

PACS numbers: 03.65.Bz, 03.65.Fd

I. INTRODUCTION

In this paper we give a full exposition of previously announced results linking a Jordan algebra structure to the measuring process.[1] The first axiomatization of the measuring process leading to Jordan algebras was done by Jordan[2] and Jordan, von Neumann, and Wigner.[3] They associated with each observable A another observable A^n which returns the nth power of the value returned by A on each measurement. Also, assumed was the existence of the *sum* $A+B$ of observables A and B satisfying

$$E_x(A+B)=E_x(A)+E_x(B)$$

for all states x, where $E_x(A)$ is the expected value of A on state x. With the assumption that A^n is given by an algebra structure, i.e., that

$$A\cdot B=\tfrac{1}{2}[(A+B)^2-A^2-B^2]$$

is bilinear, one is led to a power associative algebra which is formally real ($\Sigma A_i^2=0$ implies $A_i=0$). Finite dimensional formally real power associative algebras were shown to be Jordan; i.e.,

$$A^2\cdot(B\cdot A)=(A^2\cdot B)\cdot A.$$

The physical grounds for the existence of the sum of observables and the quadratic nature of A^2 are not clear, while the finite dimensionality assumption severely limits the applications. Indeed, in the Jordan model, if A and B are not "compatible," then $A+B$ can take values not among sums of those of A and B. Thus, it is difficult to assign a physical meaning to $A+B$, when A and B cannot be simultaneously measured. Also, the bilinearity of $A\cdot B$ is difficult to justify for noncompatible observables.

In our axiomatization, we retain the concept of taking a function of an observable, technically represented in axiom (MS2) (see Sec. II). However, we do not assume the existence of $A+B$, the quadratic nature of A^2, nor finite dimensionality. The basic new element which we introduce is a change in the measurement process due to a change in the counting observable. This is merely an exploitation of the obvious remark that the expected or average value of a collection of events depends on the total value and the count of the number of events. The changes from one counting observable to another are given by a group, reminiscent of the changes of observers in relativity theory. In our development, we show the existence of $A+B$, roughly by changing the counting observable to make A and B compatible. Also, the quadratic nature of A^2 is an involved consequence of our axioms.

In Sec. II, the foundational concepts and definitions are given, including the axioms of a measurement system. The algebraic properties of such a system are developed in Sec. III while the analytic properties are given in Sec. IV. In Sec. V, we give the main theorem stating that the observables form a formally real Jordan algebra and a normed linear space.

II. MEASUREMENT SYSTEMS

One can view the "classical" measuring process as a set Ω of *states* and the set $\mathfrak{a}$ of (bounded) functions on Ω called *observables*. The value of the observable f in state p is simply $f(p)$. In a "statistical" measuring process the *states* are probability measures on a set Ω which are absolutely continuous with respect to a fixed measure μ with $\mu(\Omega)<\infty$. The *observables* are the functions in $\mathfrak{a}_\Omega=L^\infty(\Omega,\mu)$ and the *expectation* of the observable f in state ν is $E_\Omega(f,\nu)=\int_\Omega f\,d\nu$.

The following trivial observation will play a central role in the sequel. The intuitive meaning of the expectation is that it is the average value of f over a large ensemble of independent occurrences of the state ν. Thus, if one thinks of f as additive on such occurrences, the expectation is approximated by the total value of f on an ensemble divided by the number of occurrences in the ensemble. Clearly, if the total value of f remains the same but the count of the number of occurrences changes, then the expectation of f will change.

To represent these changes in the statistical measuring process, we view $a\nu$ as an *ensemble* of a occurrences of the state ν, where $a>0$. Thus, the set $\mathscr{E}_\Omega$ of ensembles is just the set of non-negative, nonzero, finite measures η on Ω which are absolutely continuous with respect to μ. Note $\eta=a\nu$, with $a=\eta(\Omega)$. Identifying the state ν with the ray $\mathbb{R}^+\nu=\mathbb{R}^+\eta\equiv\bar{\eta}$ and defining the value of $f\in\mathfrak{a}_\Omega$ on η to be

$$\langle f,\eta\rangle_\Omega\equiv\int_\Omega f\,d\eta, \tag{1}$$

we obtain

$$E_\Omega(f,\eta)=\langle f,\eta\rangle_\Omega/\langle 1,\eta\rangle_\Omega \tag{2}$$

independent of the ensemble η in state $\bar{\eta}$. Here we have used the constant observable 1 to get the count of $\langle 1,\eta\rangle_\Omega=\eta(\Omega)$ occurrences of state $\bar{\eta}$. If we use instead the observable h to count occurrences of $\bar{\eta}$, we get

[a)] Supported in part by the National Science Foundation.

$$E'_\Omega(f,\bar\eta)=\langle f,\eta\rangle_\Omega/\langle h,\eta\rangle_\Omega. \tag{3}$$

If $h>0$, we can define $h\cdot\eta$ by $d(h\cdot\eta)=h\,d\eta$ and $h\cdot\bar\eta=\overline{(h\cdot\eta)}$. We now have

$$E'_\Omega(f,\bar\eta)=E_\Omega(h^{-1}f,h\cdot\bar\eta). \tag{4}$$

Thus, $f\to h^{-1}f$, $\eta\to h\cdot\eta$ represents the change due to changing the counting observable from 1 to h.

We shall need the following definitions. If $\mathfrak{a}$ and $\mathscr{E}$ are sets we say $\langle,\rangle$: $\mathfrak{a}\times\mathscr{E}\to\mathbb{R}$ is a *nondegenerate pairing* provided

$$\text{and}\quad\begin{aligned}&\langle A,x\rangle=\langle B,x\rangle \quad \text{for all } x\in\mathscr{E} \text{ implies } A=B\\ &\langle A,x\rangle=\langle A,y\rangle \quad \text{for all } A\in\mathfrak{a} \text{ implies } x=y.\end{aligned}$$

If $(\mathfrak{a}',\mathscr{E}',\langle,\rangle')$ is another such triple and if for θ: $\mathfrak{a}\to\mathfrak{a}'$ there exists ϕ: $\mathscr{E}'\to\mathscr{E}$ satisfying

$$\langle\theta(A),x\rangle'=\langle A,\phi(x)\rangle \quad \text{for } A\in\mathfrak{a},\ x\in\mathscr{E}',$$

then we say θ is a *homomorphism*. Since ϕ is uniquely determined by θ, we write $\phi=\theta^*$. As an example, we note that for $h, h^{-1}\in\mathfrak{a}_\Omega$, $h>0$, ω_h given by $\omega_h(f)=hf$, $\omega_h^*(\eta)=h\cdot\eta$ is an automorphism of $(\mathfrak{a}_\Omega,\mathscr{E}_\Omega,\langle,\rangle_\Omega)$ corresponding to changing the counting observable from 1 to h^{-1}. A homomorphism θ: $\mathfrak{a}_\Omega\to\mathfrak{a}$ with θ injective and θ^* surjective is a *representation* of (Ω,μ) in $(\mathfrak{a},\mathscr{E},\langle,\rangle)$. Representations θ_i, $i=1,2$, of (Ω_i,μ_i) in $(\mathfrak{a},\mathscr{E},\langle,\rangle)$ are *equivalent* if there is a measure space isomorphism α: $\Omega_1\to\Omega_2$ with $\theta_2(f)=\theta_1(f\circ\alpha)$ for all $f\in\mathfrak{a}_{\Omega_2}$. A family $\mathscr{T}$ of representations θ: $\mathfrak{a}_{\Omega_\theta}\to\mathfrak{a}$ *covers* $\mathfrak{a}$ provided

$$\mathfrak{a}=\bigcup_{\theta\in\mathscr{T}}\theta(\mathfrak{a}_\theta). \tag{5}$$

We write $\mathfrak{a}_\theta=\mathfrak{a}_{\Omega_\theta}$, $\mathscr{E}_\theta=\mathscr{E}_{\Omega_\theta}$, and $\langle,\rangle_\theta=\langle,\rangle_{\Omega_\theta}$.

A *measurement system* $\mathscr{M}=(\mathfrak{a},\mathscr{E},\langle,\rangle,\mathscr{W},I,x_I,\mathscr{T})$ consists of a set $\mathfrak{a}$ of *observables*, a set $\mathscr{E}$ of *ensembles*, a nondegenerate pairing $\langle,\rangle$: $\mathfrak{a}\times\mathscr{E}\to\mathbb{R}$, a group $\mathscr{W}$ of automorphisms of $(\mathfrak{a},\mathscr{E},\langle,\rangle)$ called *changes in counting*, a fixed observable $I\in\mathfrak{a}$ called the *counting observable*, a fixed ensemble $x_I\in\mathscr{E}$ called the *fundamental ensemble*, and a family $\mathscr{T}$ of representations θ: $\mathfrak{a}_\theta\to\mathfrak{a}$ covering $\mathfrak{a}$, satisfying

(MS1) $\theta(1)=I$, $\quad\theta^*(x_I)=\mu_\theta$,

(MS2) If $\theta_1,\theta_2\in\mathscr{T}$, if $\theta_i(f_i)=A$ for $f_i\in\mathfrak{a}_{\theta_i}$, and if u: $\mathbb{R}\to\mathbb{R}$ has $u\circ f_1\in\mathfrak{a}_{\theta_1}$, then $u\circ f_2\in\mathfrak{a}_{\theta_2}$ and $\theta_1(u\circ f_1)=\theta_2(u\circ f_2)$. The common value is denoted $u_I(A)$.

(MS3) Given $\theta\in\mathscr{T}$ and $h\in\mathfrak{a}_\theta$, then $\theta(h)=W(I)$ for some $W\in\mathscr{W}$ if and only if $h^{-1}\in\mathfrak{a}_\theta$, $h>0$. In this case, there is $\tilde\theta\in\mathscr{T}$ with $W\tilde\theta$ equivalent to $\theta\omega_h$.

(MS4) If $W\in\mathscr{W}$ and j: $\mathbb{R}\setminus\{0\}\to\mathbb{R}$ is $j(t)=t^{-1}$, then there is $\hat W\in\mathscr{W}$ with $j_IW=\hat Wj_I$, where j_I is defined.

III. ALGEBRAIC PROPERTIES

Throughout the remainder of this paper let $\mathscr{M}=(\mathfrak{a},\mathscr{E},\langle,\rangle,\mathscr{W},I,x_I,\mathscr{T})$ be a measurement system. Let $\mathscr{C}$ be the orbit $\mathscr{W}(I)$ of I under $\mathscr{W}$, and call $A\in\mathscr{C}$ a *counting observable*. By (MS3), we see $\mathscr{C}=\{\theta(h)|\theta\in\mathscr{T};\ h,h^{-1}\in\mathfrak{a}_\theta;\ h>0\}=\{\theta(h)|\theta\in\mathscr{T};\ h\geqslant\epsilon>0 \text{ for some } \epsilon\}=\{\theta(h)|\theta\in\mathscr{T};\ \langle h,v\rangle_\theta\geqslant\epsilon\langle 1,v\rangle_\theta \text{ for some } \epsilon>0 \text{ for all } v\in\mathscr{E}_\theta\}=\{A\in\mathfrak{a}|\langle A,x\rangle\geqslant\epsilon\langle I,x\rangle \text{ for some } \epsilon>0, \text{ all } x\in\mathscr{E}\}$.

Lemma 1: $\mathfrak{a}$ has the structure of a vector space over $\mathbb{R}$ so that

(a) for $x\in\mathscr{E}$, the map $A\to\langle A,x\rangle$ is a linear function on $\mathfrak{a}$,

(b) $\mathscr{W}$ acts linearly on $\mathfrak{a}$,

(c) $\mathscr{C}$ spans $\mathfrak{a}$, indeed $\mathfrak{a}=\mathbb{R}I+\mathscr{C}$.

Proof: If $A_i\in\mathfrak{a}$ and $a\in\mathbb{R}$, write $B=aA_1$ and $C=A_1+A_2$ for B, $C\in\mathfrak{a}$ provided $\langle B,x\rangle=a\langle A_1,x\rangle$ and $\langle C,x\rangle=\langle A_1,x\rangle+\langle A_2,x\rangle$ for all $x\in\mathscr{E}$. By the nondegeneracy of $\langle,\rangle$, B and C are unique, if they exist. It is also clear by uniqueness, that if B and C exist for all choices of A_i, a, then $\mathfrak{a}$ is a vector space satisfying (a). Moreover, $\langle W(A),x\rangle=\langle A,W^*(x)\rangle$ shows (b). To show B exists, we use (5) to write $A_1=\theta(f)$ with $f\in\mathfrak{a}_\theta$ and note that for $B=\theta(af)$, we have $\langle B,x\rangle=\langle af,\theta^*(x)\rangle_\theta=a\langle f,\theta^*(x)\rangle_\theta=a\langle A_1,x\rangle$. To show C exists first assume $A_1\in\mathbb{R}\mathscr{C}$, say $A_1=aW(I)$, $a\in\mathbb{R}$, $W\in\mathscr{W}$. Write $A_2=\theta(f)$ and set $C=W(\theta(a1+f))$. We see $\langle C,x\rangle=\langle a1+f,\theta^*W^*(x)\rangle_\theta=a\langle 1,\theta^*W^*(x)\rangle_\theta+\langle f,\theta^*W^*(x)\rangle_\theta=a\langle W(I),x\rangle+\langle A_2,x\rangle$. In general, write $A_1=\theta(f)$ and choose $m>0$ so that $h=m1+f\geqslant 1$. Since h, $h^{-1}\in\mathfrak{a}_\theta$ with $h>0$ we see $mI+A_1=\theta(h)\in\mathscr{C}$, showing (c). Moreover, $C=(-mI)+((mI+A_1)+A_2)$ exists and $C=A_1+A_2$.

Lemma 2: If $A\in\mathscr{C}$ with $A=W(I)$, then $U_A=W\hat W^{-1}\in\mathscr{W}$ and $x_A=\hat W^{*-1}x_I\in\mathscr{E}$ depend only on A, and $U_I=id$, the identity of $\mathscr{W}$. Moreover, $x_A=x_B$, A, $B\in\mathscr{C}$ implies $A=B$.

Proof: First assume $A=I$. If $B\in\mathscr{C}$, write $B=\theta(h)$. Since $W^{-1}(I)=I=\theta(1)$, we see by (MS3) there is $\tilde\theta\in\mathscr{T}$ with $W^{-1}\tilde\theta$ equivalent to $\theta\omega_1=\theta$. Thus, there is a measure space isomorphism α: $\Omega_{\tilde\theta}\to\Omega_\theta$ with $W^{-1}\tilde\theta(f\circ\alpha)=\theta(f)$ or $\tilde\theta(f\circ\alpha)=W\theta(f)$ for $f\in\mathfrak{a}_\theta$. Since $j_IWj_I(B)=j_IWj_I\theta(h)=j_IW\theta(h^{-1})=j_I\tilde\theta(h^{-1}\circ\alpha)=\tilde\theta((h^{-1}\circ\alpha)^{-1})=\tilde\theta(h\circ\alpha)=W\theta(h)=W(B)$, we see $\hat W=W$ on $\mathscr{C}$ and hence on $\mathfrak{a}$ by Lemma 1. Thus, $U_I=id$. Also, if $B=\theta(f)\in\mathfrak{a}$, then $W(I)=\theta(1)$ gives $\tilde\theta\in\mathscr{T}$ and α: $\Omega_{\tilde\theta}\to\Omega_\theta$ with $W\tilde\theta(f\circ\alpha)=\theta\omega_1(f)=\theta(f)$. Hence, $\langle B,W^{*-1}(x_I)\rangle=\langle W^{-1}(B),x_I\rangle=\langle W^{-1}\theta(f),x_I\rangle=\langle\tilde\theta(f\circ\alpha),x_I\rangle=\langle f\circ\alpha,\tilde\theta(x_I)\rangle_{\tilde\theta}=\langle f\circ\alpha,\mu_{\tilde\theta}\rangle_{\tilde\theta}=\langle f,\mu_\theta\rangle_\theta=\langle B,x_I\rangle$. Thus $W^{*-1}(x_I)=x_I$. Now let $A\in\mathfrak{a}$ with $W_1(I)=W_2(I)=A$. If $W=W_1^{-1}W_2$ so $W_2=W_1W$ with $W(I)=I$, and $\hat W_2=j_IW_2j_I=\hat W_1\hat W$, then $W_2\hat W_2^{-1}=W_1W\hat W^{-1}\hat W_1^{-1}=W_1\hat W_1^{-1}$ and $\hat W_2^{*-1}(x_I)=(\hat W^*\hat W_1^*)^{-1}(x_I)=\hat W_1^{*-1}\hat W^{*-1}(x_I)=\hat W_1^{*-1}(x_I)$.

If $x_A=x_B$ with $A=W_1(I)$, $B=W_2(I)$, $W_i\in\mathscr{W}$, then $\hat W_2^*\hat W_1^{*-1}x_I=x_I$. Let $W=W_1^{-1}W_2$, so that $\hat W^{*-1}x_I=x_I$. Now $A=B$ will follow from $W(I)=I$. Let $\hat W(I)=\theta(h)$ and let $\tilde\theta\in\mathscr{T}$ with $\hat W\tilde\theta$ equivalent to $\theta\omega_h$ via α: $\Omega_{\tilde\theta}\to\Omega_\theta$. If i_α: $f\to f\circ\alpha$, then $\hat W\tilde\theta i_\alpha=\theta\omega_h$ and $\omega_h^{*-1}i_\alpha^*\tilde\theta^*=\theta^*\hat W^{*-1}$. Thus, $\mu_\theta=\theta^*\hat W^{*-1}x_I=\omega_h^{*-1}i_\alpha^*\tilde\theta^*(x_I)=\omega_h^{*-1}i_\alpha^*(\mu_{\tilde\theta})=\omega_h^{*-1}\mu_\theta$ or $d\mu_\theta=h^{-1}d\mu_\theta$ and $h=1$. Thus, $\hat W(I)=I$, so $\hat W=(\hat W)\hat{}=W$ and $W(I)=I$.

Lemma 3: If $A\in\mathscr{C}$ and $A=W(I)$, then $\mathscr{M}_W=(\mathfrak{a},\mathscr{E},\langle,\rangle_A,\mathscr{W},A,x_A,\mathscr{T}_W)$ is a measurement system where $\langle B,x\rangle_A=\langle U_A^{-1}(B),x\rangle$ and $\mathscr{T}_W=\{\theta_W=W\theta\,|\theta\in\mathscr{T}\}$. $\mathscr{M}_W$ is uniquely determined by A up to equivalence of the representations $\mathscr{T}_W$. Also, $(\mathscr{M}_W)_{W'}=\mathscr{M}_{W'W}$.

Proof: Clearly, $\langle,\rangle_A$ is a nondegenerate pairing. If ρ: $\mathfrak{a}'\to\mathfrak{a}$ is a homomorphism relative to $\langle,\rangle$, the $\langle\rho(B'),x\rangle_A=\langle U_A^{-1}\rho(B'),x\rangle=\langle\rho(B'),U_A^{*-1}(x)\rangle=\langle B',\rho^*U_A^{*-1}(x)\rangle'$. Thus, ρ is also a homomorphism relative to

$\langle,\rangle_A$ with $\phi=\rho^* U_A^{*-1}: \mathscr{C}\to\mathscr{C}'$. Similarly, a homomorphism $\rho: \mathfrak{a}\to\mathfrak{a}'$ for $(\mathfrak{a}, \mathscr{C}, \langle,\rangle)$ is also a homomorphism for $(\mathfrak{a}, \mathscr{C}, \langle,\rangle_A)$ with $\phi' = U_A^* \rho^*: \mathscr{C}'\to\mathscr{C}$. In particular, $\theta_W: \mathfrak{a}_\theta\to\mathfrak{a}$ is a representation of $(\Omega_\theta, \mu_\theta)$ in $(\mathfrak{a}, \mathscr{C}, \langle,\rangle_A)$ with $\theta_W^{*A} = (\theta_W)^* U_A^{*-1} = \theta^* W^*(W\hat{W}^{-1})^{*-1} = \theta^* \hat{W}^*$. Also, if $V\in\mathscr{W}$ then V is an isomorphism of $(\mathfrak{a}, \mathscr{C}, \langle,\rangle)$ with $(\mathfrak{a}, \mathscr{C}, \langle,\rangle_A)$ and hence an automorphism of $(\mathfrak{a}, \mathscr{C}, \langle,\rangle_A)$ with $V^{*A} = U_A^* V^* U_A^{*-1}$.

Clearly $\mathscr{T}_W$ covers $\mathfrak{a}$. If $W(I)=A=I$, then as in the proof of Lemma 2, for each $\theta\in\mathscr{T}$ there is $\tilde{\theta}\in\mathscr{T}$ with $\theta_W = W\theta$ equivalent to $\tilde{\theta}$. In general, we see that if $W(I) = W'(I)$ then $\mathscr{T}_W$ is equivalent to $\mathscr{T}_{W'}$.

Since $\theta_W(1) = W\theta(1) = A$ and $\theta_W^{*A}(x_A) = \theta^* \hat{W}^* \hat{W}^{*-1}(x_I) = \mu_\theta$, (MS1) holds for $\mathscr{M}_W$. If $W\theta_i\in\mathscr{T}_W$ with $W\theta_i(f_i) = B$ and if $u\circ f_1\in\mathfrak{a}_{\theta_1}$, then $\theta_i(f_i) = W^{-1}(B)$ so $u\circ f_2\in\mathfrak{a}_{\theta_2}$ and $W\theta_1(u\circ f_1) = W\theta_2(u\circ f_2)$ showing (MS2). Let $u_W(B) = Wu_I W^{-1}(B)$ be the common value. If $W'(I) = A$, then $W\theta_1$ is equivalent to some $W'\tilde{\theta}$ so there is $\alpha: \Omega_{\tilde{\theta}}\to\Omega_\theta$ with $B = W\theta_1(f_1) = W'\tilde{\theta}(f_1\circ\alpha)$ and $u_{W'}(B) = W'\tilde{\theta}(u\circ f_1\circ\alpha) = W\theta_1(u\circ f_1) = u_W(B)$. Write $u_A(B) = u_W(B)$ and note $u_{W(I)} = Wu_I W^{-1}$ so

$$Wu_A W^{-1} = u_{W(A)} \quad \text{for } A\in\mathscr{C},\ W\in\mathscr{W}. \tag{6}$$

If $\theta_W\in\mathscr{T}_W$ and $h\in\mathfrak{a}_\theta$, then $\theta_W(h) = V(A)$ for some $V\in\mathscr{W}$ if and only if $\theta(h) = W^{-1}VW(I)$ if and only if $h^{-1}\in\mathfrak{a}_\theta, h>0$. In this case $\theta_W\omega_h = W\theta\omega_h$ is equivalent to $W(W^{-1}VW)\tilde{\theta} = V\tilde{\theta}_W$ for some $\tilde{\theta}_W\in\mathscr{T}_W$, showing (MS3). To show (MS4), we let $V\in\mathscr{W}$, and compute $j_A V = Wj_I W^{-1}V = Wj_I W^{-1}VWW^{-1} = W\hat{W}^{-1}\hat{V}\hat{W}W^{-1}Wj_I W^{-1} = (U_A \hat{V} U_A^{-1}) j_A$.

Finally, if $W, W'\in\mathscr{W}$ then $B = W'(A) = (W'W)(I)$, $(U_A \hat{W}' U_A^{-1})^{*A^{-1}} x_A = \hat{W}'^{*-1}\hat{W}^{*-1}x_I = (W'W)\hat{}^{*-1}(x_I) = x_B$, $W'\theta_W = (W'W)\theta$, and $U_B^A = W'(U_A\hat{W}'U_A^{-1})^{-1} = W'W\hat{W}^{-1}\hat{W}'^{-1}U_A^{-1} = U_B U_A^{-1}$ so $\langle (U_B^A)^{-1}C, x\rangle_A = \langle U_A^{-1}U_A U_B^{-1}C, x\rangle = \langle C, x\rangle_B$. Thus $(\mathscr{M}_W)_{W'} = \mathscr{M}_{W'W}$.

We define $s: \mathbb{R}\to\mathbb{R}$ by $s(t) = t^2$. Since $f\in\mathfrak{a}_\theta$ implies $f^2\in\mathfrak{a}_\theta$, $s_C(A)$ is defined for $C\in\mathscr{C}$, $A\in\mathfrak{a}$. If $B\in\mathfrak{a}$, set $s_C(A,B) = s_C(A+B) - s_C(A) - s_C(B)$.

Lemma 4: If $A, B, C\in\mathscr{C}$, then

(a) $j_A^2 = id$,

(b) $j_A(A) = A$,

(c) $U_B^A = j_B j_A$,

(d) $j_A j_B j_A = j_{j_A(B)}$,

(e) $u_B(A) = \tilde{u}_A(B)$, for $u: \mathbb{R}\setminus\{0\}\to\mathbb{R}$, $\tilde{u}(t) = tu(t^{-1})$,

(f) $s_B(A) = j_A(B)$,

(g) $\langle A, x_B\rangle_C = \langle B, x_A\rangle_C$,

(h) $\langle A, x_A\rangle_C = \langle s_C(A), x_C\rangle_C$,

(i) $2\langle A, x_B\rangle_C = \langle s_C(A,B), x_C\rangle_C$,

(j) $\langle U_A^C B, x_B\rangle_C = \langle U_B^C A, x_A\rangle_C$.

Proof: Since by Lemma 2, replacing $\mathscr{M}$ by $\mathscr{M}_V$, $V\in\mathscr{W}$ does not change any of the expressions, we may assume that any one of A, B, C is I. For (a)–(e), let $A = I$ and $B = W(I) = \theta(h)$. Now (a) and (b) are trivial. For (c), $U_B = W\hat{W}^{-1} = Wj_I W^{-1} j_I = j_B j_I$ by (6). Also, $\hat{W}(I) = j_I W j_I(I) = j_I(B)$ implies $j_{j_I(B)} j_I = U_{j_I(B)} = \hat{W}W^{-1} = U_B^{-1} = j_I j_B$, yielding (d). For (e), let $W\tilde{\theta}$ be equivalent to $\theta\omega_h$ via $\alpha: \Omega_{\tilde{\theta}}\to\Omega_\theta$. Now $u_I(B) = \theta(u\circ h) = \theta\omega_h(h^{-1}(u\circ h)) = \theta\omega_h((\tilde{u}\circ j)\circ h) = \tilde{\theta}_W((\tilde{u}\circ j)\circ h\circ\alpha) = (\tilde{u}\circ j)_B\tilde{\theta}_W(h\circ\alpha)$. Letting $u(t) = 1$ so $\tilde{u}(t) = t$, we see $I = u_I(B) = j_B(\tilde{\theta}_W(h\circ\alpha))$ or $\tilde{\theta}_W(h\circ\alpha) = j_B(I)$. In general, we have $u_I(B) = (\tilde{u}\circ j)_B j_B(I) = \tilde{u}_B(I)$. Since $\tilde{s} = j$, (f) follows from (e). For (g)–(j), let $C = I$ and $A = W(I)$. We note $\hat{U}_A = (W\hat{W}^{-1})\hat{} = U_A^{-1}$ and $j_I(A) = j_I j_A A = U_A^{-1}W(I)$. Thus, $\langle B, x_{j_I(A)}\rangle = \langle B, (U_A^{-1}W)\hat{}^{*-1}x_I\rangle = \langle B, U_A^{*-1}\hat{W}^{*-1}x_I\rangle = \langle U_A^{-1}B, x_A\rangle = \langle B, x_A\rangle_A$ and $\langle j_I(A), x_B\rangle = \langle U_A^{-1}A, x_B\rangle = \langle A, x_B\rangle_A$. Replacing A by $j_I(A)$ shows $\langle B, x_A\rangle = \langle A, x_B\rangle$ follows from $\langle B, x_A\rangle_A = \langle A, x_B\rangle_A$. Thus, for (g), we can assume $A = C = I$. Let $r(t) = t^{1/2}$ for $t>0$, and note $D = r_I(B)$ exists. Also $U_D(I) = j_D(I) = s_I(D) = B$. Thus, $\langle I, x_B\rangle = \langle I, U_D^{*-1}x_I\rangle = \langle U_D(I), x_I\rangle = \langle B, x_I\rangle$ showing (g). If $D = r_I(A)$, then $U_D(I) = j_D(I) = A$ and $U_D A = j_D j_I j_D(I) = j_A(I) = s_I(A)$. Thus $\langle A, x_A\rangle = \langle A, U_D^{*-1}x_I\rangle = \langle U_D(A), x_I\rangle = \langle s_I(A), x_I\rangle$, yielding (h). For (i), we see $\langle s_I(A,B), x_I\rangle = \langle s_I(A+B), x_I\rangle - \langle s_I(A), x_I\rangle - \langle s_I(B), x_I\rangle = \langle A+B, x_{A+B}\rangle - \langle A, x_A\rangle - \langle B, x_B\rangle = \langle A, x_{A+B}\rangle + \langle B, x_{A+B}\rangle - \langle A, x_A\rangle - \langle B, x_B\rangle = \langle A+B, x_A\rangle + \langle A+B, x_B\rangle - \langle A, x_A\rangle - \langle B, x_B\rangle = 2\langle A, x_B\rangle$. For (j), $\langle U_A B, x_B\rangle = \langle j_A(j_I j_B)B, x_B\rangle = \langle U_B j_A U_B^{-1}B, x_B\rangle_B = \langle j_{U_B(A)}B, x_B\rangle_B = \langle s_B(U_B A), x_B\rangle_B = \langle U_B A, x_{U_B A}\rangle_B = \langle A, x_{U_B A}\rangle = \langle U_B A, x_A\rangle$.

Lemma 5: If $C\in\mathscr{C}$, then $s_C(A,B)$ is linear in A and in B.

Proof: Let $C = I$. If $A, B\in\mathscr{C}$, then $A+B\in\mathscr{C}$ since $\langle A+B, x\rangle > \langle A, x\rangle \geqslant \epsilon\langle I,x\rangle$ for some $\epsilon>0$. Set $U_{A,B} = U_{A+B} - U_A - U_B$ as endomorphisms of the vector space $\mathfrak{a}$. Since $\langle U_A B, x\rangle = \langle j_A j_I B, x\rangle = \langle j_I j_{j_I(A)}B, x\rangle = \langle B, x\rangle_{j_I(A)}$, we see for $D\in\mathscr{C}$, $\langle U_A B, x_D\rangle = \langle U_A D, x_B\rangle$. Also, $\langle A, x_{B+D}\rangle = \langle B, x_A\rangle + \langle D, x_A\rangle = \langle A, x_B\rangle + \langle A, x_D\rangle$. Replacing B by $B+D$ in $\langle U_A B, x_B\rangle = \langle U_B A, x_A\rangle$ yields $2\langle U_A B, x_D\rangle = \langle U_{B,D}A, x_A\rangle$. Replacing A by $A+E$, with $E\in\mathscr{C}$, gives

$$\langle U_{A,E}B, x_D\rangle = \langle U_{B,D}A, x_E\rangle.$$

Letting $A = I$ and noting $U_{B,D}I = j_{B+D}I - j_B I - j_D I = s_I(B, D)$, we see

$$\langle U_{I,E}B, x_D\rangle = \langle s_I(B,D), x_E\rangle \quad \text{for } B,D,E\in\mathscr{C}. \tag{7}$$

If $A, B\in\mathscr{C}$ with $\langle A, x_E\rangle = \langle B, x_E\rangle$ for all $E\in\mathscr{C}$ then $\langle E, x_A\rangle = \langle E, x_B\rangle$. Since $\mathscr{C}$ spans $\mathfrak{a}$, we see $x_A = x_B$ and $A = B$ by Lemma 2. Again since $\mathscr{C}$ spans $\mathfrak{a}$, for $Q\in\mathfrak{a}$, $\langle Q, x_E\rangle = 0$ for all $E\in\mathscr{C}$ implies $Q = 0$. Since the left side of (7) is linear in B, we see

$$s_I(B_1 + B_2, D) = s_I(B_1, D) + s_I(B_2, D) \quad \text{for } B_i, D\in\mathscr{C}. \tag{8}$$

For arbitrary $Q, P, R\in\mathfrak{a}$, we have

$$s_I(Q+P, R) = s_I(Q+R, P) + s_I(Q, R) - s_I(Q, P), \tag{9}$$

as is easily checked from the definition. If $Q = \theta(f)$, then $s_I(Q, mI) = \theta((f+m1)^2) - \theta(f^2) - \theta((m1)^2) = \theta(2mf) = 2mQ$. If $P = mI$ in (9), we see $s_I(Q+mI, R) = 2m(Q+R) + s_I(Q, R) - 2mQ = s_I(Q, R) + 2mR$. Letting $B = Q + mI$, $D = R + nI$, we see

$$s_I(Q, R) = s_I(B, D) - 2mR - 2nB. \qquad (10)$$

For Q_i, $R \in \mathfrak{a}$, choose m_i, n with $B_i = Q_i + m_I I$, $D = R + nI \in \mathscr{C}$ and let $Q = Q_1 + Q_2$, $m = m_1 + m_2$. From (8) and (10), we get $s_I(Q_1 + Q_2, R) + s_I(Q_2, R)$.

Since $a > 0$, $B \in \mathscr{C}$ implies $aB \in \mathscr{C}$, we have from (7) that

$$s_I(aB, D) = as_I(B, D) \quad \text{for } a > 0,\ B,\ D \in \mathscr{C}. \qquad (11)$$

From (10), we then have $s_I(aQ, R) = as_I(Q, R)$ for $a > 0$, Q, $R \in \mathfrak{a}$. If $a = 0$, it is trivial. For $a < 0$, it follows by additivity that $s_I(aQ, R) = -s_I(-aQ, R) = as_I(Q, R)$.

IV. ANALYTIC PROPERTIES

If $A \in \mathfrak{a}$, $C \in \mathscr{C}$, then by Lemmas 1 and 3, there is $m > 0$ with $mC \pm A \in \mathscr{C}$. Thus, $m\langle C, x\rangle_C \pm \langle A, x\rangle_C \geqslant 0$ and

$$\frac{|\langle A, x\rangle_C|}{\langle C, x\rangle_C} \leqslant m \quad \text{for all } x \in \mathscr{E}.$$

Thus,

$$\|A\|_C = \sup_{x \in \mathscr{E}} \frac{|\langle A, x\rangle_C|}{\langle C, x\rangle_C}$$

is well defined

Lemma 6: The following hold for $A \in \mathfrak{a}$, B, $C \in \mathscr{C}$:

(a) $\|\ \|_C$ is a norm on $\mathfrak{a}$,

(b) $\|A\|_C = \|f\|_\theta$, if $\theta(f) = A$, $\theta \in \mathscr{T}_W$, $C = W(I)$,

(c) $\|A\|_C \leqslant \|j_B(C)\|_B \|A\|_B$,

(d) $\|\ \|_C$ and $\|\ \|_B$ are equivalent norms,

(e) $\|U_B A\|_C \leqslant \|A\|_C \|B\|_C^2$,

(f) U_B is continuous.

Proof: Clearly, $\|aA\|_C = |a|\ \|A\|_C$. Also $\|A\|_C = 0$ implies $\langle A, x\rangle = 0$ for all $x \in \mathscr{E}$ and $A = 0$. Since $|\langle A_1 + A_2, x\rangle_C| \leqslant |\langle A_1, x\rangle_C| + |\langle A_2, x\rangle_C|$, we have $\|A_1 + A_2\|_C \leqslant \|A_1\|_C + \|A_2\|_C$ showing (a). Since $\theta^*(\mathscr{E}) = \mathscr{E}_\theta$, we see

$$\|A\|_C = \sup_{v \in \mathscr{E}_\theta} \frac{|\int_{\Omega_\theta} f\, dv|}{v(\Omega_\theta)} = \|f\|_\theta,$$

the (essential) supremum of $|f|$ over Ω_θ. For (c), let $B = I$ and $C = \theta(h)$, $\theta \in \mathscr{T}$. Set $M = \|j_I(C)\|_I = \|h^{-1}\|_\theta$, so $h > M^{-1}$. Thus, for all $x \in \mathscr{E}$, $\langle C, x\rangle \geqslant M^{-1}\langle I, x\rangle$ or

$$0 < \frac{\langle I, x\rangle}{\langle C, x\rangle} \leqslant M.$$

Now

$$\begin{aligned}\frac{|\langle A, x\rangle_C|}{\langle C, x\rangle_C} &= \frac{|\langle U_C^{-1}A, x\rangle|}{\langle U_C^{-1}C, x\rangle} = \frac{|A, U_C^{*-1}x\rangle|}{\langle C, U_C^{*-1}x\rangle}\\ &= \frac{|\langle A, U_C^{*-1}x\rangle|}{\langle I, U_C^{*-1}x\rangle}\frac{\langle I, U_C^{*-1}x\rangle}{\langle C, U_C^{*-1}x\rangle}\\ &\leqslant M\|A\|_I = \|j_I(C)\|_I \|A\|_I.\end{aligned}$$

Clearly, (d) follows from (c). For (e), we let $C = I$, and compute

$$\begin{aligned}&\|U_B^{-1}A\|_I\\ &\quad = \sup_{x \in \mathscr{E}} \frac{|\langle U_B^{-1}A, x\rangle|}{\langle U_B^{-1}B, x\rangle}\frac{\langle U_B^{-1}B, x\rangle}{\langle I, x\rangle} \leqslant \|A\|_B \|U_B^{-1}B\|_I\\ &\quad \leqslant \|A\|_I \|j_I(B)\|_I^2\end{aligned}$$

since $U_B^{-1}B = j_I j_B(B) = j_I(B)$. Replacing B by $j_I(B)$ and noting $U_{j_I(B)}^{-1} = (j_I j_B j_I j_I)^{-1} = U_B$ yields (e). Finally, (f) follows from (e).

Recall that if $\mathscr{X}$ is an associative algebra with 1 and a normed linear space with $\|xy\| \leqslant \|x\|\ \|y\|$, then $\|x^{-1}\| \leqslant \|x^{-1}\|^2\|x\|$ so $\|x\|^{-1} \leqslant \|x^{-1}\|$. Thus, $\|x\|^{-1} = \|x^{-1}\|$ and $\|1\| = 1$. Now $1 = \|1 + x - x\| \leqslant \|1 + x\| + \|x\|$ so $\|(1 + x)^{-1}\| \leqslant 2$ if $\|x\| \leqslant \frac{1}{2}$. If $z = (1 + x)^{-1} - (1 - x + x^2) = -(1 + x)^{-1}x^3$, we see $\lim_{x \to 0}(\|z\|/\|x\|^2) = 0$. Hence,

$$(1 + x)^{-1} = 1 - x + x^2 + z, \quad \text{with } \lim_{x \to 0} \frac{\|z\|}{\|x\|^2} = 0. \qquad (12)$$

Lemma 7: If $A \in \mathfrak{a}$, B, $C, \in \mathscr{C}$ then

(a) $j_I(C + A) = j_I(C) - U_C^{-1}A + U_C^{-1}s_C(A) + Z$,

where

$$\lim_{A \to 0} \frac{\|Z\|_I}{\|A\|_I^2} = 0,$$

(b) $U_{B,I}C = U_{C,I}B = s_I(B, C)$.

Proof: First let $C = I$ and $A = \theta(f)$. By (12) for $\mathfrak{a}_\theta$, we have $(1 + f)^{-1} = 1 - f + f^2 + g$ with $\lim_{f \to 0}(\|g\|_\theta/\|f\|_\theta^2) = 0$. Applying θ, we get

$$j_I(I + A) = I - A + s_I(A) + Z, \quad \text{where } \lim_{A \to 0} \frac{\|Z\|_I}{\|A\|_I^2} = 0. \qquad (13)$$

Replacing I by C in (13) and applying $U_C^{-1} = j_I j_C = U_{J_I(C)}$ gives (a), since $\lim_{A \to 0}(\|Z\|_C/\|A\|_C^2) = 0$ implies $\lim_{A \to 0}(\|U_{j_I(C)}Z\|_I/\|A\|_I^2) = 0$ by Lemma 6(d) and 6(e).

For (b), let $t > 0$ and let $C = I$, $A = t(B + C)$ in (a) to obtain

$$j_I(I + tb + tC) = I - t(B + C) - t^2 s_I(B + C) + Z_t, \qquad (14)$$

where $\lim_{t \to 0^+} t^{-2}Z_t = 0$ since $\lim_{t \to 0^+}\|t^{-2}Z_t\| = \lim_{t \to 0^+}(\|Z_t\|_I/\|tB + tC\|_I^2)\|B + C\|_I^2 = 0$. Let $C = I + tB$, $A = tC$ in (a) to obtain

$$\begin{aligned}&j_I(I + tB + tC)\\ &\quad = j_I(I + tB) - U_{I + tB}^{-1}(tC) + U_{I + tB}^{-1}s_{I + tB}(tC) + W_t,\end{aligned} \qquad (15)$$

where $\lim_{t \to 0^+} t^{-2}W_t = 0$. Also, $C = I$, $A = tB$ in (a) gives

$$j_I(I + tB) = I - tB + t^2 s_I(B) + Y_t, \qquad (16)$$

where $\lim_{t \to 0^+} t^{-2}Y_t = 0$. Note $U_B C = j_B j_I(C) = s_{j_I(C)}(B)$ and $U_{B,D}C = s_{j_I(C)}(B, D)$, so

$$U_{tB} = t^2 U_B,\ U_{tB,D} = tU_{B,D} \quad \text{for } t > 0, \quad B, D \in \mathscr{C} \qquad (17)$$

by (11), with $j_I(C)$ replacing I.

Since the continuous maps from $\mathfrak{a}$ to itself form an algebra with norm

$$\|T\| = \sup_{\|A\|_I < 1} \|T(A)\|_I$$

satisfying $\|TS\| \leqslant \|T\|\ \|S\|$, (17) and (12) yield

$$U_{I + tB}^{-1} = id - t(U_{I,B} + tU_B) + V_t, \qquad (18)$$

where $\lim_{t \to 0^+} t^{-1}V_t = 0$, since $\lim_{t \to 0^+} t^{-1}\|V_t\| = \lim_{t \to 0^+}[\|V_t\|/\|t(U_{I,B} + tU_B)\|]\|U_{I,B} + tU_B\| = 0$. Also, $s_{I + tB}(C) = j_C(I + tB) = j_C j_I j_I(I + tB) = j_C(I)$

$+ R_t$, where $\lim_{t\to 0^+} t^{-2}R_t = 0$ by the proof of (13). Hence

$$s_{I+tB}(tC) = t^2 s_I(C) + Q_t, \tag{19}$$

where $\lim_{t\to 0^+} t^{-2}Q_t = 0$. Combining (15), (16), (18), and (19) yields $j_I(I + tB + tC) = I - tB + t^2 s_I(B) + Y_t$ $= (id - tU_{I,B} - t^2U_B + V_t)tC + (id - tU_{I,B} - t^2U_B + V_t)$ $(t^2 s_I(C) + Q_t) + W_t$ or

$$j_I(I + tB + tC)$$
$$= I - t(B + C) + t^2(s_I(B) + s_I(C) + U_{I,B}C) + R_t, \tag{20}$$

where $\lim_{t\to 0^+} t^{-2}R_t = 0$. Using (14) and (20 to compute $\lim_{t\to 0^+} t^{-2}[\, j_I(I + tB + tC) - (I - tB - tC)]$ yields (b).

V. RELATION TO JORDAN ALGEBRAS

We are now in a position to show that the space $\mathfrak{a}$ of observables form a Jordan algebra. It will be convenient to use a special case of the McCrimmon[4] formulation of quadratic Jordan algebras which is equivalent, for the fields considered, to the classical definition. Thus, a vector space $\mathscr{X}$ over a field of characteristic zero, a quadratic map U: $\mathscr{X} \to \mathrm{End}(\mathscr{X})$, and *unit* $c \in \mathscr{X}$ form a *quadratic Jordan algebra* provided for $x, y, z \in \mathscr{X}$,

(QJ1) $U_c = id,$

(QJ2) $U_x U_y U_x = U_{U_x y},$

(QJ3) $U_x U_{y,z} x = U_{U_x y, x} z,$

where $U_{y,z} = U_{y+z} - U_y - U_z$.

Theorem: If $\mathscr{M}$ is a measurement system, then the set $\mathfrak{a}$ of observables has the structure of a quadratic Jordan algebra and a normed linear space over $\mathbb{R}$ so that

(a) I is the unit of $\mathfrak{a}$,

(b) $U_A = j_A j_I$ for $A \in \mathscr{C}$,

(c) $U_A B$ is continuous in $A, B \in \mathfrak{a}$,

(d) $\|A^2\| = \|A\|^2, A \in \mathfrak{a}$,

(e) $\Sigma A_i^2 = 0$ implies $A_i = 0$.

Proof: We say a vector valued function l on $\mathscr{C}$ is $\mathscr{C}$-*linear* if

$$l(A + B) = l(A) + l(B), \quad A, B \in \mathscr{C}, \tag{21}$$

$$l(aA) = al(A), \quad a > 0, A \in \mathscr{C}. \tag{22}$$

In this case, l has a unique extension to a linear function $\tilde{l}$ on $\mathfrak{a}$. Indeed, $\tilde{l}$ is unique if it exists, since $\mathscr{C}$ spans $\mathfrak{a}$ by Lemma 1. Also, $\tilde{l}$ will exist if $\Sigma a_i A_i = 0$, $a_i \in \mathbb{R}$, $A_i \in \mathscr{C}$ implies $\Sigma a_i l(A_i) = 0$. The condition $\Sigma a_i A_i = 0$ can be rewritten as $\Sigma b_j B_j = \Sigma c_k C_k$ with $b_j, c_k > 0$, $B_j, C_k \in \mathscr{C}$. But $\Sigma b_j l(B_j) = \Sigma c_k l(C_k)$ follows from (21) and (22). If b: $\mathscr{C} \times \mathscr{C} \to \mathscr{V}$ is $\mathscr{C}$-bilinear, then $b_A(C) = b(A, C), A, C \in \mathscr{C}$ extends to a linear map $\tilde{b}_A : \mathfrak{a} \to \mathscr{C}$. Since for $A, B \in \mathscr{C}$, $a > 0$, $\tilde{b}_A + \tilde{b}_B = \tilde{b}_{A+B}$ and $a\tilde{b}_A = \tilde{b}_{aA}$ by uniqueness, we see f: $A \to \tilde{b}_A$ extends to a linear map $\tilde{f}$: $\mathfrak{a} \to \mathrm{End}(\mathfrak{a}, \mathscr{V})$. Now $\tilde{b}(A, B) \equiv \tilde{f}(A)(B) = b(A, B)$ for $A, B \in \mathscr{C}$, so $\tilde{b}$ is a bilinear map extending b. By induction, we see any $\mathscr{C}$-multilinear map has a unique extension to a multilinear map on $\mathfrak{a}$. In particular, since $U_{A,B}C = s_{j_I(C)}(A, B)$ for $A, B, C \in \mathscr{C}$ we see, by Lemma 1(b) and Lemma 5, that $A \times B \times C \to U_{A,B}C$ extends to a unique trilinear map $A \times B \times C \to \widetilde{U}_{A,B}C$ on $\mathfrak{a}$. Define $\widetilde{U}_A = \frac{1}{2}\widetilde{U}_{A,A}$ and note for $A \in \mathscr{C}$, $\widetilde{U}_A = \frac{1}{2}(U_{2A} - 2U_A) = U_A$. We now simply write U_A and $U_{A,B}$ for $\widetilde{U}_A$ *and* $\widetilde{U}_{A,B}$, $A, B \in \mathfrak{a}$. Clearly, $A \to U_A$ is a quadratic map U: $\mathfrak{a} \to \mathrm{End}(\mathfrak{a})$. Also, $U_I = id$ by Lemma 2, so (QJ1) holds. If $A, B \in \mathscr{C}$, then $U_A U_B U_A = j_A j_I j_B j_I j_A j_I = j_{j_A j_I(B)} j_I = U_{U_A(B)}$ by Lemma 4(d), so (QJ2) holds for $A, B \in \mathscr{C}$. Replacing A by $\Sigma t_i A_i$, $t_i > 0$, $A_i \in \mathscr{C}$ and B by $\Sigma s_j B_j$, $s_j > 0$, $B_j \in \mathscr{C}$ and picking out the coefficient of $t_1 t_2 t_3 t_4 s_1 s_2$, we see the full linearization of (QJ2) holds for $\mathscr{C}$. By linearity, the linearization and (QJ2) itself hold for $\mathfrak{a}$. Similarly, it suffices to show (QJ3) for $\mathscr{C}$. If $A, B, C, D \in \mathscr{C}$, then $U_D U_{B,C} U_D I = U_{U_D B, U_D C} I = s_I(U_D B, U_D C) = U_{U_D B, I} U_D C$ by (QJ2) and Lemma 7(b). If $E = j_I(D)$, then $U_A^E = j_A j_E = j_A j_I j_D j_I = U_A U_D$. Hence, we have

$$U_I^E U_{B,C}^E I = U^E_{U_I^E B, I} C. \tag{23}$$

Replacing I by A and E by I in (23) gives (QJ3) for $\mathscr{C}$. Thus, $\mathfrak{a}$ is a Jordan algebra.

Since $A^2 = U_A I = s_I(A)$, we see for $A = \theta(f)$, $\|A^2\| = \|s_I(A)\|_I = \|f^2\|_\theta = \|f\|_\theta^2 = \|A\|^2$, showing (d). For $A \cdot B = \frac{1}{2}s_I(A, B)$, we see $\|A \cdot B\| \leqslant \frac{1}{2}(\|A + B\|^2 + \|A\|^2 + \|B\|^2) \leqslant \|A\|^2 + \|B\|^2 + \|A\|\,\|B\| = \|A\|\,\|B\|(1 + \|A\|/\|B\| + \|B\|/\|A\|)$. Replacing A by tA, where $t = \|B\|/\|A\|$, we see $\|A \cdot B\| = t^{-1}\|(tA) \cdot B\| \leqslant 3\|A\|\,\|B\|$. Hence $A \cdot B$ is continuous in A and B. Since for any Jordan algebra, $U_A B = 2A \cdot (A \cdot B) - A^2 \cdot B$, (c) holds. Finally, $\Sigma A_i^2 = 0$ and $\langle A_i^2, x\rangle \geqslant 0$ implies $\langle A_i^2, x\rangle = 0$ so $A_i^2 = 0$ and $A_i = 0$.

[1] J. R. Faulkner, "An apology for Jordan algebras in quantum theory," Contemporary Math. (to appear).

[2] P. Jordan, Z. Phys. **80**, 285 (1933).

[3] P. Jordan, J. von Neumann, and E. Wigner, Ann Math. **35**, 29 (1934).

[4] K. McCrimmon, Proc. Natl. Acad. Sci. U.S.A. **56**, 1072 (1966).

Upper and lower bounds to zeroth order Coulombic hyperangular interaction integrals

Metin Demiralp and N. Abdülbaki Baykara

Applied Mathematics Division, Marmara Scientific and Industrial Research Institute, P.O. Box 141, Kadıköy-Istanbul, Turkey

(Received 3 February 1981; accepted for publication 19 June 1981)

In recent years it has been shown that the use of hyperspherical coordinate representation of the Schrödinger equation for electrically charged particles necessitates the evaluation of certain kinds of hyperangular interaction integrals. The analytic evaluation of rather simple cases has also been accomplished. On the other hand certain numerical devices have been utilized and the complications that have arisen have been discussed for their computation. We have attempted in this work to find upper and lower bounds for all types of zeroth order hyperangular interaction integrals having Coulombic potentials. A nesting procedure has been developed for obtaining close bounds which can possibly be used to evaluate the desired value of integrals under consideration. In this context a theorem has been established for the evaluation of a similar type of integral and possible ways towards the generalization of the theorem have also been discussed. For three particle systems some applications have also been presented.

PACS numbers: 03.65.Ge

I. INTRODUCTION

To determine the energy values of a system of electrically charged particles, one can write the Schrödinger equation in hyperspherical coordinates. We then see that almost all of the mathematical complications are encountered in the evaluation of certain hyperangular interaction integrals.[1] Among these integrals the simplest one having only two potential interactions was analytically evaluated.[2] For the general case a computational scheme has also been proposed.[3] There we have also discussed the convergence of the offered scheme and pointed out that in certain cases the speed of convergence is quite slow. Although various techniques can be adopted to accelerate the convergence, determination of upper and lower bounds for these entities gains importance at least in checking numerically obtained values. Besides, the establishment of a nesting procedure gives the possibility of evaluating those integrals within any desired precision. A slow convergence might show up. This, however, is beyond the scope of the present work. In this context we shall only be interested in searching for upper and lower bounds and in making the difference between them narrower with the aid of nesting procedure.

II. PRESENTATION OF THE SCHEME TO FIND THE UPPER AND LOWER BOUNDS

Consider the following hyperangular interaction integral:

$$\chi_p\left(\begin{array}{c|c} m_1,\ldots & u_1,\ldots \\ \ldots m_p & \ldots u_{p+1} \end{array}\right) \equiv \int_{S_\xi} \theta_0^*(\xi^T A_1 \xi)^{-1/2} \mathscr{L}^{m_1}(\xi^T A_2 \xi)^{-1/2} \times \mathscr{L}^{m_2}\ldots(\xi^T A_p \xi)^{-1/2}\mathscr{L}^{m_p}(\xi^T A_{p+1}\xi)^{-1/2}\theta_0 dS_\xi, \quad (2.1)$$

where $m_1,\ldots,m_p$ represent some positive integers and ξ is a $3N$-dimensional unit vector. The integration is performed over the domain of hyperangles. The effect of the operator $\mathscr{L}$ on any arbitrary function φ which is in the space spanned by the hyperspherical harmonics is given as below,

$$\mathscr{L}_\varphi = -\sum_{k=1}^{\infty}\sum_{l=1}^{d_k}\frac{1}{k(k+2\alpha)}\theta_k^{(l)}\int_S \theta_k^{(l)*}\varphi dS, \quad (2.2)$$

where $\theta_k^{(l)}$ denotes a k th order hyperspherical harmonic with an upper index l showing its position among the d_k, k th order hyperspherical harmonics. In other words d_k represents the degeneracy. The star appearing as a superscript of $\theta_k^{(l)*}$ implies the complex conjugate of $\theta_k^{(l)}$. For a system having $N+1$ particles the parameter α is given by the relationship

$$\alpha = (3N-2)/2.$$

The A_j's are certain idempotent matrices which can be derived from some unit $3N$-dimensional vectors u_j's and three dimensional unit matrix I_3 with the aid of direct product operation as follows:

$$A_j = u_j u_j^T \otimes I_3, \quad j = 1,\ldots,p+1.$$

After certain intermediate steps presented in a previous work,[2] the effect of the m consecutive applications of $\mathscr{L}$ on the function φ can be expressed as

$$\mathscr{L}_\varphi^m = -\frac{\Gamma(\alpha+1)}{(m-1)!2\pi^{\alpha+1}}\sum_{J=0}^{m-1}\frac{\Gamma(2m-J-1)(-1)^J}{J!(m-J-1)!(2\alpha)^{2m-J-1}}\int_{S_\eta}\int_0^1\frac{1-(-1)^J x^{2\alpha}}{x}\left[\frac{1-x^2}{(1-2(\xi^T\eta)x+x^2)^{\alpha+1}}-1\right] \times(-\ln x)^J\varphi(\eta)dS_\eta. \quad (2.3)$$

0022-2488/82/091622-06$02.50

The utilization of this formula in Eq. (2.1) yields

$$\chi_p\left(\begin{array}{c|c} m_1,\ldots & u_1,\ldots \\ \ldots m_p & \ldots,u_{p+1}\end{array}\right) = \left(\frac{\Gamma(\alpha+1)}{2\pi^{\alpha+1}}\right)^{p+1}\sum_{J_1=0}^{m_1-1}\cdots\sum_{J_p=0}^{m_p-1}\prod_{l=1}^{p}\frac{\Gamma(2m_l-J_l-1)(-1)^{J_l+1}}{J_l!(m_l-J_l-1)!(2\alpha)^{2m_l-J_l-1}}\Lambda_p^\alpha\left(\begin{array}{c|c} J_1,\ldots & u_1,\ldots \\ \ldots J_p & \ldots u_{p+1}\end{array}\right), \tag{2.4}$$

where Λ_p^α is defined as follows:

$$\Lambda_p^\alpha\left(\begin{array}{c|c} J_1,\ldots & u_1,\ldots \\ \ldots J_p & \ldots u_{p+1}\end{array}\right) = \int_0^1\cdots\int_0^1\int_{S_{p+1}}\cdots\int_{S_1}\prod_{q=1}^{p+1}(\xi_q^TA_q\xi_q)^{-1/2}$$
$$\times\prod_{q=1}^{p}\frac{1-(-1)^{J_p}x_q^2}{x_q}\left[\frac{1-x_q^2}{(1-2(\xi_q^T\xi_{q+1})x_q+x_q^2)^{\alpha+1}}-1\right]dx_q\prod_{q=1}^{p+1}dS_q. \tag{2.5}$$

The expansion of the integrand above into the powers of terms like $(\xi_q^T\xi_{q+1})$ results in

$$\Lambda_p^\alpha\left(\begin{array}{c|c} J_1,\ldots & u_1,\ldots \\ \ldots J_p & \ldots u_{p+1}\end{array}\right) = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_p=0}^{\infty}\tau_{k_1\ldots k_p}\prod_{q=1}^{p}\left[\left(\frac{\alpha+1}{2}\right)_{k_q}\left(\frac{\alpha+2}{2}\right)_{k_q}\Big/\left(\tfrac{1}{2}\right)_{k_q}k_q!\right]\beta_{k_q}^{J_q}(\alpha), \tag{2.6}$$

where $\beta_{k_q}^J(\alpha)$'s denote the integrals over x_q's. In a previous work[2] an analytic formula was given for these integrals.

$$\beta_k^J(\alpha) = \frac{1}{2^J}\left\{\frac{d^J}{dt^J}\left[\frac{2^{2k}\Gamma(k+\alpha+t)\Gamma(k-t)}{\Gamma(2k+a+1)}+\frac{\delta_{k0}}{t(t+\alpha)}\right]\left(t+\frac{\alpha}{2}\right)\right\}_t=0. \tag{2.7}$$

The integrals $\tau_{k_1\ldots k_2}$ (the dependence on α and u's of which are not shown explicitly for the sake of simplicity) on the other hand carry mathematical complications, since they characterize all potential interactions on unit hypersphere. Their explicit representation is

$$\tau_{k_1\ldots k_p} = \int_{S_{p+1}}\cdots\int_{S_1}\prod_{q=1}^{p+1}(\xi_q^TA_q\xi_q)^{-1/2}\prod_{q=1}^{p}(\xi_q\xi_{q+1})^{2k_q}\prod_{q=1}^{p+1}dS_q. \tag{2.8}$$

It can easily be shown that for all non-negative integer J values and physically meaningful α values ($\alpha = 0.5, 2, 3.5, 5, \ldots$) β_0^J is negative. For all other k, however, $\beta_k^J(\alpha)$ is positive. In constructing inequalities, this type of sign distinction creates certain difficulties. To get rid of this complication one can employ the following entity instead of $\Lambda_p^\alpha\left(\begin{smallmatrix} J_1,\ldots \\ \ldots J_p\end{smallmatrix}\middle|\begin{smallmatrix} u_1,\ldots \\ \ldots u_{p+1}\end{smallmatrix}\right)$.

$$\Omega_p^\alpha\left(\begin{array}{c|c} J_1,\ldots & \ldots u_{p+1} \\ \ldots J_p & u_1,\ldots\end{array}\right) = \sum_{k_1=1}^{\infty}\cdots\sum_{k_p=1}^{\infty}\tau_{k_1\ldots k_p}\prod_{q=1}^{p}\left[\left(\frac{\alpha+1}{2}\right)_{k_q}\left(\frac{\alpha+2}{2}\right)_{k_q}\Big/\left(\frac{1}{2}\right)_{k_q}k_q!\beta_{k_q}^{J_q}(\alpha)\right]. \tag{2.9}$$

The negativity of β_0^J can be dealt with after constructing inequalities for this entity Ω_p^α.

To find bounds for $\tau_{k_1\ldots k_p}$ one can take into consideration the following identity given in a previous publication.[2]

$$\int_{S_1}\frac{(\xi_1^T\xi_2)^{2k_1}}{\sqrt{(\xi_1^TA_1\xi_1)}}dS_1 = \frac{4\pi^\alpha\Gamma(k_1+\frac{1}{2})}{\Gamma(k_1+\alpha+\frac{1}{2})}\int_0^1(1-\sigma^2t^2)^{k_1}dt, \tag{2.10}$$

$$\sigma^2 = \xi_2^TA_1\xi_2. \tag{2.11}$$

A careful look at the integrand above reveals the following inequalities:

$$(1-t^2)^{k_1} < (1-\sigma^2t^2)^{k_1} < 1 \tag{2.12}$$

and these in turn imply that

$$\frac{4\pi^\alpha\Gamma(k_1+\frac{1}{2})\Gamma(k_1+1)}{\Gamma(k_1+\alpha+\frac{1}{2})(\frac{3}{2})_{k_1}} < \int_{S_1}\frac{(\xi_1^T\epsilon_2)^{2k_1}}{\sqrt{(\xi_1^TA_1\xi_1)}}dS_1$$
$$< \frac{4\pi^\alpha\Gamma(k_1+\frac{1}{2})}{\Gamma(k_1+\alpha+\frac{1}{2})}. \tag{2.13}$$

The consecutive use of these inequalities in Eq. (2.8) leads to

$$\left(\frac{4\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac{1}{2})}\right)^{p+1}\prod_{q=1}^{p}\frac{(\frac{1}{2})_{k_q}(1)_{k_q}}{(\alpha+\frac{1}{2})_{k_q}(\frac{3}{2})_{k_q}} < \tau_{k_1\ldots k_q}$$
$$< \left(\frac{4\pi^{\alpha+1/2}}{(\alpha+\frac{1}{2})}\right)^{p+1}\prod_{q=1}^{p}\frac{(\frac{1}{2})_{k_q}}{(\alpha+\frac{1}{2})_{k_q}}. \tag{2.14}$$

Now let us define the following entities:

$$U_J(\alpha) = \sum_{k=1}^{\infty}\left[\left(\frac{\alpha+1}{2}\right)_k\left(\frac{\alpha+2}{2}\right)_k\Big/(\alpha+\tfrac{1}{2})_k k!\right]\beta_k^J(\alpha), \tag{2.15}$$

$$L_J(\alpha) = \sum_{k=1}^{\infty}\left[\left(\frac{\alpha+1}{2}\right)_k\left(\frac{\alpha+2}{2}\right)_k\Big/(\alpha+\tfrac{1}{2})_k(\tfrac{3}{2})_k\right]\beta_k^J(\alpha). \tag{2.16}$$

One can express $U_J(\alpha)$ in terms of a Gaussian hypergeometric function having unit argument and write the following closed form.

$$U_J(\alpha) = \frac{1}{2^J}\left\{\frac{d^J}{dt^J}\frac{(t+\alpha/2)}{\Gamma(\alpha+1)}\Gamma(\alpha+t)\Gamma(-t)\right.$$
$$\left.\times\left[\frac{\Gamma(\alpha+\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2}-t)\Gamma(\alpha+\frac{1}{2}+t)}-1\right]\right\}_{t=0}. \tag{2.17}$$

The analytic evaluation of $L_j(\alpha)$, however, needs a little bit more effort. In this case a generalized hypergeometric function ${}_3F_2$ with unit argument shows up. Using certain transformations for this kind of function with unit argument[5] one can obtain the following result.

$$L_J(\alpha)=\frac{1}{2^J}\left\{\frac{d^J}{dt^J}\frac{(t+\alpha/2)}{\Gamma(\alpha+1)}\frac{\Gamma(\alpha+\frac12)\Gamma(\frac32)}{(t+\frac12)(\alpha+t-\frac12)}\right.$$
$$\left.\times\left[\frac{\Gamma(1-t)\Gamma(\alpha+t+1)}{\Gamma(\frac32)\Gamma(\alpha+\frac12)}-1\right]\right\}_{t=0}. \quad (2.18)$$

Therefore, one concludes

$$\left(\frac{4\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac12)}\right)^{p+1}\prod_{q=1}^{p}L_{J_q}(\alpha)<\Omega_p^\alpha\left(\begin{array}{c|c}J_1,\ldots & u_1,\ldots\\ \ldots J_p & \ldots u_{p+1}\end{array}\right)$$
$$<\left(\frac{4\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac12)}\right)^{p+1}\prod_{q=1}^{p}U_{J_q}(\alpha). \quad (2.19)$$

III. A NESTING PROCEDURE TO EVALUATE TIGHTER UPPER AND LOWER BOUNDS

The gap between upper and lower bounds given by Eq. (2.19) may be narrowed, with the aid of a nesting procedure. To this end let us rewrite Eq. (2.9) in the following manner where all $_{k_q}$'s are positive integers,

$$\Omega_p^\alpha=\sum_{k=0}^{n}\sum_{k}^{*}\tau_{k_1,\ldots,k_p}\prod_{q=1}^{p}\left[\left(\frac{\alpha+1}{2}\right)_{k_q}\left(\frac{\alpha+2}{2}\right)_{k_q}\Big/(\tfrac12)_{k_q}k_q!\right]\beta_{k_q}^{J_q}(\alpha)$$
$$+\sum_{k=n+1}^{\infty}\sum_{k}^{*}\tau_{k_1,\ldots,k_p}\prod_{q=1}^{p}\left[\left(\frac{\alpha+1}{2}\right)_{k_q}\left(\frac{\alpha+2}{2}\right)_{k_q}\Big/(\tfrac12)_{k_q}k_q!\right]\beta_{k_q}^{J_q}(\alpha). \quad (3.1)$$

Starred summation is to be performed in such a way that all indexed quantities are summed for all values of the indices $k_1,\ldots,k_p$ which fulfil the condition $k_1+\ldots+k_p=k+p$. Inserting the inequality (2.14) into Eq. (3.1) one can after some algebra conclude

$$B_L^{(n)}<\Omega_p^\alpha\left(\begin{array}{c|c}\ldots J_p & \ldots u_{p+1}\\ J_1,\ldots & u_1,\ldots\end{array}\right)<B_U^{(n)}, \quad (3.2)$$

where dependence of $B_L^{(n)}$ and $B_U^{(n)}$ on $\alpha, J_1,\ldots,J_p, u_1,\ldots,u_{p+1}$ is not shown explicitly for the sake of simplicity. The bounds $B_U^{(n)}$ and $B_L^{(n)}$ may be defined by the following recursive relations:

$$B_U^{(n)}=B_U^{(n-1)}+\Sigma_n^*\left[\tau_{k_1\ldots k_p}-(4\pi^{\alpha+1/2}/\Gamma(\alpha+\tfrac12))^{p+1}\prod_{q=1}^{p}\frac{(\frac12)_{k_q}}{(\alpha+\frac12)_{k_q}}\right]\prod_{q=1}^{p}\left[\left(\frac{\alpha+1}{2}\right)_{k_q}\left(\frac{\alpha+2}{2}\right)_{k_q}\Big/(\tfrac12)_{k_q}k_q!\right]\beta_{k_q}^{J_q}(\alpha), \quad (3.3)$$

$$B_U^{(0)}=\left(\frac{4\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac12)}\right)^{p+1}\prod_{q=1}^{p}U_{J_q}(\alpha) \quad (3.4)$$

$$B_L^{(n)}=B_L^{(n-1)}+\Sigma_n^*\left[\tau_{k_1\ldots k_p}-\left(\frac{4\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac12)}\right)^{p+1}\prod_{q=1}^{p}\frac{(\frac12)_{k_q}(1)_{k_q}}{(\alpha+\frac12)_{k_q}(\frac32)_{k_q}}\right]\prod_{q=1}^{p}\left[\left(\frac{\alpha+1}{2}\right)_{k_q}\left(\frac{\alpha+2}{2}\right)_{k_q}\Big/(\tfrac12)_{k_q}k_q!\right]\beta_{k_q}^{J_q}(\alpha), \quad (3.5)$$

$$B_L^{(0)}=\left(\frac{4\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac12)}\right)^{p+1}\prod_{q=1}^{p}L_{J_q}(\alpha). \quad (3.6)$$

A careful look at these relations immediately shows that the following inequalities are correct:

$$B_L^{(0)}<B_L^{(1)}<\ldots<B_L^{(n)}<\Omega_p^\alpha\left(\begin{array}{c|c}\ldots J_p & \ldots u_{p+1}\\ J_1,\ldots & u_1,\ldots\end{array}\right)<B_U^{(n)}$$
$$<\ldots B_U^{(1)}<B_U^{(0)} \quad (3.7)$$

and the sequences of $B_L^{(n)}$'s and $B_U^{(n)}$'s converge to Ω_p^α,

$$_{n\to\infty}B_L^{(n)}={}_{n\to\infty}B_U^{(n)}=\Omega_p^\alpha\left(\begin{array}{c|c}\ldots J_p & \ldots u_{p+1}\\ J_1,\ldots & u_1,\ldots\end{array}\right). \quad (3.8)$$

Therefore by using this nesting procedure (nth bounds are nested by $n-1$th bounds) one can evaluate Ω_p's in any desired precision. Of course the speed of convergence for the sequences of $B_L^{(n)}$'s and $B_U^{(n)}$'s will determine the amount of effort needed to obtain a given accuracy. This subject is beyond the scope of this work.

IV. EVALUATION OF THE BOUNDS

To evaluate $B_L^{(n)}$ and $B_U^{(n)}$ one needs the values of $\beta_k^J(\alpha)$ and $\tau_{k_1\ldots k_p}$. One can recall that $\beta_k^J(\alpha)$ was analytically evaluated. The structure of $\tau_{k_1\ldots k_p}$ is, however, more complicated. An efficient technique is to diagonalize the quadratic form $\xi_q^TA_q\xi_q$ into $\xi_q^T\hat{I}\xi_q$ (all elements of $\hat{I}$ are zero except for the first three diagonal elements which equal unity). The integral $\tau_{k_1\ldots k_p}$ can be finitely expressed in terms of u_j vectors, since the quadratic form $\xi_q^T\hat{I}\xi_q$ can be reduced to the product of sine squares of hyperangles corresponding to ξ_q. This structure essentially reduces the integrals over ξ_q's to certain polynomials of the elements of orthonormal matrices which diagonalize the quadratic forms involved in the integrand of $\tau_{k_1\ldots k_p}$. In the case of smaller p and n values this evaluation can be handled fairly easily. Larger values of p, however, enlarge the dimensions of the matter. Even for small values of $k_1\ldots,k_p$ evaluation of $\tau_{k_1\ldots k_p}$ may end up being inconvenient for hand calculations. We find it appropriate to start with the simplest case where all k's are unity and gradually extend to other cases with some extra effort. As a start we need the following theorem.

Theorem 4.1: If we define the following operator,

$$\mathcal{N}_k(\alpha,A,\xi)f(\xi)=\int_{S_\eta}\frac{(\xi^T\eta)^{2k}}{\sqrt{(\eta^TA\eta)}}f(\eta)dS_\eta,$$
$$k=1,2,3,\ldots \quad (4.1)$$

where f is an integrable function of hyperangles, (a) $\mathcal{N}_k(\alpha,A,\xi)$ transforms any homogeneous polynomial of ξ into a kth order homogeneous multinomial; (b) the following relation holds between the kernel matrices of the quadratic forms $\xi^T B\xi$ and $\xi^T C\xi$, the latter of which is obtained from the former one by the operator $\mathscr{N}_1(\alpha,A,\xi)$,

$$\xi^T C\xi = \mathscr{N}_1(\alpha,A,\xi)\xi^T B\xi, \tag{4.2}$$

$$C = [\pi^{\alpha+1/2}/(\alpha+\tfrac{5}{2})]\, Q\,\{(\mathrm{Tr}B)I + 2B - \tfrac{1}{3}(\mathrm{Tr}\hat{I}B\hat{I})I - \tfrac{1}{3}(\mathrm{Tr}B)\hat{I} - \tfrac{2}{3}(\hat{I}B + B\hat{I}) + \tfrac{1}{5}(\mathrm{Tr}\hat{I}B\hat{I})\hat{I} + \tfrac{2}{5}\hat{I}B\hat{I}\}Q^T, \tag{4.3}$$

where Tr and Q stand for trace and the orthonormal matrix which diagonalizes the matrix A.

Proof: Since part (a) can be proved directly expanding the integrand in terms of ξ_j's we can start with Eqs. (4.2) and (4.1) to prove part (b). After a diagonalization procedure one can explicitly write

$$\mathscr{N}_1(\alpha,A,\xi)\xi^T B\xi = \sum_{j,k,l,m=1}^{3N} (Q^T BQ)_{jk}(Q^T\xi\xi^T Q)_{lm}\int_{S_\eta} \frac{\eta_j\eta_k\eta_l\eta_m}{\sqrt{(\eta_1^2+\eta_2^2+\eta_3^2)}}\, dS_\eta, \tag{4.4}$$

where an indexed parenthesis shows the corresponding element of the matrix products shown inside the parenthesis. The use of hyperangular coordinates ($\eta_{3N} = \cos\theta_{3N-1}$, $\eta_{3N-1} = \sin\theta_{3N-1}\cos\theta_{3N-2}, \ldots$) verifies

$$\int_S \frac{\eta_j\eta_k\eta_l\eta_m}{\sqrt{(\eta_1^2+\eta_2^2+\eta_3^2)}}\, dS_\eta = \Big\{\delta_{jk}\delta_{lm}\int_{S_\eta}\frac{\eta_j^2\eta_l^2}{\sqrt{(\eta_1^2+\eta_2^2+\eta_3^2)}}\,dS_\eta + (\delta_{jl}\delta_{km}+\delta_{jm}\delta_{kl}) \times\int_S \frac{\eta_j^2\eta_k^2\,dS}{(\eta_1^2+\eta_2^2+\eta_3^2)^{1/2}} - 2\delta_{jk}\delta_{kl}\delta_{lm}\int\frac{\eta_j^4}{(\eta_1^2+\eta_2^2+\eta_3^2)^{1/2}}\,dS_\eta\Big\}$$

$$(\delta_{jk} = \text{Kronecker's symbols}). \tag{4.5}$$

Due to the fact that any orthonormal transformation which affects only $\eta_4, \eta_5, \ldots, \eta_{3N}$ or η_1, η_2, η_3 does not alter the values of integrals above but changes the indices appearing in the integrands, after some intermediate steps one can summarize

$$\int_S \frac{\eta_j^4}{\sqrt{(\eta_1^2+\eta_2^2+\eta_3^2)}}\,dS = \begin{cases} \dfrac{8\pi^{\alpha+1/2}}{5\Gamma(\alpha+\frac{5}{2})}, & j=1,2,3 \\[2ex] \dfrac{3\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac{5}{2})}, & j=4,5,\ldots,3N \end{cases} \tag{4.6}$$

$$\int_S \frac{\eta_j^2\eta_k^2}{\sqrt{(\eta_1^2+\eta_2^2+\eta_3^2)}}\,dS_\eta = \begin{cases} \dfrac{8\pi^{\alpha+1/2}}{15\Gamma(\alpha+\frac{5}{2})}, & j,k=1,2,3, \qquad j\neq k \\[2ex] \dfrac{2\pi^{\alpha+1/2}}{3\Gamma(\alpha+\frac{5}{2})}, & j=1,2,3 \qquad k=4,5,\ldots 3N \text{ (or vice versa)}. \\[2ex] \dfrac{\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac{5}{2})}, & j,k=4,5,\ldots,3N, \qquad j\neq k \end{cases} \tag{4.7}$$

These equations can be condensed into the following single one by virtue of Kronecker's symbols,

$$\int_S \frac{\eta_j^2\eta_k^2}{\sqrt{(\eta_1^2+\eta_2^2+\eta_3^2)}}\,dS_\eta = \frac{\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac{5}{2})}\{1-\tfrac{1}{3}(\delta_{j1}+\delta_{j2}+\delta_{j3}+\delta_{k1}+\delta_{k2}+\delta_{k3}) + (2-\tfrac{14}{15}[\delta_{j1}+\delta_{j2}+\delta_{j3}])\delta_{jk} + \tfrac{1}{5}(\delta_{j1}+\delta_{j2}+\delta_{j3})(\delta_{k1}+\delta_{k2}+\delta_{k3})\}. \tag{4.8}$$

The utilization of this equality together with Eq. (4.5) in Eq. (4.4) completes the proof.

Corollary 1: The evaluation of $\tau_{11\ldots11}$ can be accomplished by means of the following recursive relation:

$$B_q = [\pi^{\alpha+1/2}/\Gamma(\alpha+\tfrac{5}{2})]\, Q_q\{(\mathrm{Tr}B_{q-1})I + 2B_{q-1} - \tfrac{1}{3}(\mathrm{Tr}\hat{I}B_{q-1}\hat{I})I - \tfrac{1}{3}(\mathrm{Tr}B_{q-1})\hat{I} + \tfrac{2}{3}(\hat{I}B_{q-1} + B_{q-1}\hat{I}) + \tfrac{1}{5}(\mathrm{Tr}\hat{I}B_{q-1}\hat{I})\hat{I} + \tfrac{2}{5}\hat{I}B_{q-1}\hat{I}\}Q_q^T, \quad q = 2, 3, 4, \ldots, p. \tag{4.9}$$

The initial value B_1 of this recursion can be found from

$$\xi^T B_1\xi = \int_{S_\eta} \frac{(\xi^T\eta)^2}{\sqrt{(\eta^T A_1\eta)}}\,dS_\eta \tag{4.10}$$

with the aid of the following equation which can be derived in a similar but easier manner from the evaluation of the integral given by Eq. (4.5),

$$\int_{S_\eta} \eta_j\eta_k(\eta_1^2+\eta_2^2+\eta_3^2)^{-1/2}dS_\eta = \frac{2\pi^{\alpha+1/2}}{(\alpha+\frac{3}{2})}[1-\tfrac{1}{3}(\delta_{j1}+\delta_{j2}+\delta_{j3})]\,\delta_{jk}. \tag{4.11}$$

One can write

$$B_1 = \frac{2\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac{3}{2})}Q_1(I-\tfrac{1}{3}\hat{I})Q_1^T \tag{4.12}$$

and after some intermediate manipulations

$$\tau_{11\ldots11} = \frac{2\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac{3}{2})}\,\mathrm{Tr}\{B_p - \tfrac{1}{3}\hat{I}Q_{p+1}^T B_p Q_{p+1}\hat{I}\}. \tag{4.13}$$

In the preceeding formulas Q_q stands for the orthonormal matrix which diagonalizes A_q.

Corollary 2: The form of A_j given by Eq. (2.2) implies

$$Q_j = \bar{Q}_j \otimes I_3, \qquad (4.14)$$

where I_3 and $\bar{Q}_j$ are three-dimensional unit and N-dimensional orthonormal matrices, respectively, the latter which diagonalize the idempotent matrix $u_j u_j^T$. This allows us to write

$$B_q = \frac{4\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac{1}{2})}\Big)^q \bar{B}_q \otimes I_3 \qquad (4.15)$$

and

$$\hat{I} = \bar{I} \otimes I_3, \quad \bar{I} = e_1 e_1^T, \qquad (4.16)$$

where e_1 is an N-dimensional unit Cartesian vector, namely

$$e_1^T = [1, 0, \ldots 0]. \qquad (4.17)$$

These and the identity

$$\bar{I}\bar{B}_q\bar{I} = \{\mathrm{Tr}\bar{I}\bar{B}_q\bar{I}\}\bar{I} \qquad (4.18)$$

lead to the following equations after some algebra:

$$\bar{B}_1 = [1/(2\alpha+1)]\bar{Q}_1(I_N - \tfrac{1}{3}\bar{I})\bar{Q}_1^T; \quad I_N\text{:}N\text{-dimensional unit matrix} \qquad (4.19)$$

$$\begin{aligned}\bar{B}_q = {} & [1/(2\alpha+1)(2\alpha+3)]\bar{Q}_q\{3(\mathrm{Tr}\bar{B}_{q-1})I_N + 2\bar{B}_{q-1} \\ & - (\mathrm{Tr}\bar{I}\bar{B}_{q-1}\bar{I})I_N - (\mathrm{Tr}\bar{B}_{q-1})\bar{I} + (\mathrm{Tr}\bar{I}\bar{B}_{q-1}\bar{I})\bar{I} \\ & - \tfrac{2}{3}(\bar{I}\bar{B}_{q-1} + \bar{B}_{q-1}\bar{I})\}\bar{Q}_q^T, \end{aligned} \qquad (4.20)$$

$$\begin{aligned}\tau_{11\ldots11} = {} & \left(\frac{4\pi^{\alpha+1/2}}{\Gamma(\alpha+\frac{1}{2})}\right)^{p+1}[1/(2\alpha+1)] \\ & \times[3\mathrm{Tr}\bar{B}_q - u_{p+1}^T\bar{B}_p u_{p+1}], \end{aligned} \qquad (4.21)$$

the last of which has a more compact form than Eq. (4.13).

Corollary 3: In the case of adjacent unit indices, i.e., $\tau_{k_1\ldots k_j 11\ldots 11 k_{j+1}\ldots k_p}$, where $k_j \neq 1$, $k_{j+1} \neq 1$, part (a) guarantees that the integrals corresponding to the indices $k_1,\ldots,k_j$, $k_{j+1},\ldots,k_p$ can be expressed in quadratic forms; then the consecutive use of Eq. (4.3) finalizes the problem as to the evaluation of an integral over hyperangles, the intergrand of which is a product of two quadratic forms and the reciprocal of the square root of another quadratic form. Although the corresponding τ looks hard to evaluate, this last integral does not have a complicated structure for evaluation and can be tackled by making use of the aforementioned approaches.

V. CERTAIN APPLICATIONS AND CONCLUSION

It is because of the fact that three-particle systems are the most realistic ones among those characterized by symmetric wavefunctions that we make our applications in this section for $\alpha = 2$. For this case $(2^J/J!)U_J(2)$ and $(2^J/J!)L_J(2)$ values required for the determination of bounds to Ω_p^2 are calculated. The related sets of values for roughest upper and lower bounds are (1.333333, 2.245179, 2.029793, 2.240185) and (0.547935, 1.086775, 1.231266, 1.417371) for $J = 0, 1, 2, 3$, respectively. A quick glance at Eq. (2.4) will show that recombination of Ω_p^2 with $(2^J/J!)$ appearing in the coefficient of A_p^α will lead us to a healthier analysis, especially in a numerical sense. Although the structure of $U_J(2)$ and $L_J(2)$ can be given analytically, the algebraic effort to take the Jth derivatives will become extremely tedious for higher values of J. One does need to use more explicit expressions for $U_J(2)$ and $L_J(2)$. To this end the serial expansion of the products $\Gamma(1+t)\Gamma(1-t)$ and $\Gamma(\frac{1}{2}+t)\Gamma(\frac{1}{2}-t)$ into the powers of t can be made by virtue of Bernoulli numbers[4] or better still Riemann's Zeta[4] function with non-negative integer-valued arguments (excluding the singular point 1). After some intermediate algebra we can conclude

$$U_J(2) = (J!/2^J)(\mathscr{D}_{J-1} + 2\mathscr{D}_J + \mathscr{D}_{J+1}), \qquad (5.1)$$

where

$$\begin{aligned}\mathscr{D}_J = {} & \frac{3}{2}\sum_{k=0}^{J}(-1)^k 2^k(1-3^{-k-1})\frac{1+(-1)^{J-k}}{2}\zeta(J-k) \\ & + \frac{1+(-1)^J}{2}(1-2^{-J+1})\zeta(J), \quad \mathscr{D}_{-1} \equiv 0, \end{aligned} \qquad (5.2)$$

$$\begin{aligned}L_J(2) = {} & (J!/2^J)(\mathscr{E}_{J-3} + 4\mathscr{E}_{J-2} + 5\mathscr{E}_{J-1} + 2\mathscr{E}_J \\ & + \mathscr{F}_{J-1} + \mathscr{F}_J), \end{aligned} \qquad (5.3)$$

where

$$\begin{aligned}\mathscr{E}_J = {} & 2\sum_{k=0}^{J}(-1)^k 2^k(1-3^{-k-1})(1-2^{-J+k+1}) \\ & \times\frac{1+(-1)^{J-k}}{2}\zeta(J-k), \ \mathscr{E}_J \equiv 0, J<0, \end{aligned} \qquad (5.4)$$

$$\mathscr{F}_J = -\tfrac{3}{8}\pi(-1)^J 2^J(1-3^{-J-1}), \quad \mathscr{F}_{-1} \equiv 0. \qquad (5.5)$$

These types of expressions are not restricted to the case where $\alpha = 2$. Similar but more complicated formulas can be derived for integer α values. These are beyond the scope of this work. On the other hand half-integer α values seem to create some trouble due to the existence of products like $\Gamma(\frac{1}{2}-t)\Gamma(t)$. In these cases one can, however, expand these types of products in the powers of t, but the expansion coefficients seem not to be expressible in terms of easily calculable quantities.

For higher values of J one can derive an asymptotic form for β_k^J's from its integral representation,[2]

$$\begin{aligned}\beta_k^J = {} & \int_0^1 [1-(-1)^J x^{2\alpha}](1-x^2)(2x)^{2k-1} \\ & \times(1+x^2)^{-\alpha-1-2k}dx. \end{aligned} \qquad (5.6)$$

Indeed, serial expansion of the last factor in the integrand above into the powers of x, and term by term integra-

TABLE I. Nested bounds to the simplest $\Omega_1^2(J/u_1,u_2)$. (a) $\gamma = u_1^T u_2$, (2) values are to be multiplied by $(16\pi^2/3)^2$.

J	0			3		
γ^2	0.0	0.5	1.0	0.0	0.5	1.0
B_U^1	1.2533	1.2667	1.2800	1.4034	1.4495	1.4958
B_U^3	1.1598	1.1849	1.2132	1.3278	1.3837	1.4420
B_U^5	1.1039	1.1344	1.1710	1.3108	1.3683	1.4291
Ω_1^2	0.7225	0.7474	0.8161	1.2915	1.3500	1.4132
B_L^5	0.6606	0.6908	0.7274	1.2878	1.3451	1.4059
B_L^3	0.6431	0.6680	0.6963	1.2824	1.3381	1.3963
B_L^1	0.6013	0.6146	0.6279	1.2475	1.2936	1.3398

tion yields the following equation after certain elementary steps:

$$((2^J/J!)\beta_k^J)_{J\to\infty} = 2\delta_{k1}. \tag{5.7}$$

The use of this fact in the definition of $U_J(2)$ and $L_J(2)$ gives the limit values 2.4 and 1.6, respectively. As a matter of fact given sets of four values tend to these limits.

To obtain tighter bounds we use this last discussion and the findings of the previous sections to build Table I. There, the comparison of the nested bounds with the actual values of Ω_p^2 has been given for the first few n values and $J = 0$, $J = 3$ in the simplest case. A careful look at Table I reveals that the rate of the convergence of the nesting procedure increases as J increases. This implies that the nesting procedure can be effectively used for the evaluation of Ω_p^2 for higher J values.

To evaluate first nested bounds to the general case of Ω_p^α we need the value of $\tau_{11\ldots11}$. As can be recalled from the aforementioned theorem and its corollaries a recursion formula can be made use of for this purpose. For the sake of computational simplicity we can reorganize the elements of $\overline{B}_q$ in a vectorial form, making use of its symmetry. This, after some elementary matrix algebra, yields for the most realistic case of $\alpha = 2$,

$$\tau_{11\ldots11} = \left(\frac{16\pi^2}{3}\right)^{p+1} 3\omega_{p+1}^T \mathscr{M}_p \mathscr{M}_{p-1} \ldots \mathscr{M}_2 \omega_1, \tag{5.8}$$

where

$$\omega_q^T = \tfrac{1}{5}[\,1 - \tfrac{1}{3}\cos^2\theta_q, 1 - \tfrac{1}{3}\sin^2\theta_q, \quad -\tfrac{1}{3}\sin\theta_q\cos\theta_q\,], \qquad q = 1, p+1 \tag{5.9}$$

and

$$\mathscr{M}_q = \tfrac{2}{35}\begin{bmatrix} 1 + \tfrac{1}{3}\cos^2\theta_q & 1 + \tfrac{3}{2}\sin^2\theta_q & -\tfrac{1}{3}\sin\theta_q\cos\theta_q \\ 1 + \tfrac{1}{3}\sin^2\theta_q & 1 + \tfrac{3}{2}\cos^2\theta_q & \tfrac{1}{3}\sin\theta_q\cos\theta_q \\ \tfrac{1}{3}\sin\theta_q\cos\theta_q & -\tfrac{3}{2}\sin\theta_q\cos\theta_q & \tfrac{2}{3}(\cos^2\theta_q - \sin^2\theta_q) \end{bmatrix}, \qquad q = 2,3,\ldots,p, \tag{5.10}$$

where θ_q is originated from the following representation of the vector u_q appearing in the definition of the potential matrix A_q

$$u_q^T = [\cos\theta_q, \sin\theta_q,] \quad q = 1,\ldots,p+1. \tag{5.11}$$

For other positive integer values of α the sample approach can be used but higher-dimensional matrices and vectors appear.

For the equipotential case further simplifications can be made making use of the fact that all θ_q's are equivalent and can be chosen to be zero. This produces the following compact form:

$$\tau_{11\ldots11} = \left(\frac{16\pi^2}{3}\right)^{p+1} 3\overline{\omega}^T \overline{\mathscr{M}}^{p-1}\omega, \tag{5.12}$$

where

$$\overline{\omega}^T = \tfrac{1}{5}[\,\tfrac{2}{3}, 1] \tag{5.13}$$

and

$$\overline{\mathscr{M}} = \frac{2}{35}\begin{bmatrix} \tfrac{4}{3} & 1 \\ 1 & \tfrac{5}{2} \end{bmatrix}. \tag{5.14}$$

By employing the theory of matrices one can arrive at more explicit results which it is unnecessary to give here.

Therefore in the light of these applications we can say that upper and lower bounds to Ω_p^α's may be effective in their actual evaluation for certain cases. However, to increase the precision required other types of τ integrals must be evaluated. But this necessitates dealing with tensors. We have left this subject out of this work. This may be the subject of a future work.

[1] M. Demiralp, J. Chem. Phys. **72**, 3828–3826 (1980).

[2] M. Demiralp and N. A. Baykara, "Analytic Evaluation of Certain Zeroth Order Coulombic Hyperangular Interaction Integrals," J. Math. Phys. **22**, 2427 (1981).

[3] N. A. Baykara and M. Demiralp, "Numerical Evaluation of the Zeroth Order Coulombic Hyperangular Interaction Integrals" (submitted for publication to J. Chem. Phys.)

[4] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, Berlin, 1966).

[5] Y. L. Luke, *The Special Functions and Their Approximations*, Vol. 1 (Academic, New York, 1969).

Mass dependence of Schrödinger wavefunctions for an exponential potential

R. K. Roychoudhury
Electronics Unit, Indian Statistical Institute, Calcutta-700035, India

(Received 17 February 1981; accepted for publication 8 May 1981)

It is shown that the condition $G(r) \geqslant 0$ for $r \geqslant 0$, where $G(r) = (\partial/\partial m)\int_0^r (u(r))^2\, dr$, does not hold in the case of exponential type potentials.

PACS numbers: 03.65.Ge, 03.65.Nk, 14.40. – n

I. INTRODUCTION

Recently Leung and Rosner[1] studied mass dependence of wave funtions in the nonrelativistic case for some types of potentials. This study is useful in the investigation of the spectrum of bound states like the ψ/J or Υ particles.[2] They showed that for a power law potential of the type $V(r) = r^\epsilon (\epsilon > 0)$ and also for logarithmic potentials the condition

$$G(r) \geqslant 0 \quad \text{for } r \geqslant 0,$$

where

$$G(r) = \frac{\partial}{\partial m} \int_0^r \{u(r)\}^2\, dr$$

is satisfied. Here $u(r)$ denotes normalized Schrödinger wavefunctions. It is not very difficult to obtain the mass dependence of the Schrödinger wavefunctions and energies[3] for a power law potential if one uses the simple scaling arguments. However, for a potential which is a transcendental function of r, scaling arguments cannot be used to obtain meaningful results. Hence, mass dependence of wavefunctions for some well known monotonic potentials like $e^{-r/a}$ (exponential type), $(1/r)e^{-kr}$ (Yukawan), or $A/\cosh^2 \alpha r$ (modified Pöschteller) cannot be determined in a trivial manner. In this note we show that the condition mentioned above fails to hold in the case of exponential type potentials. Our investigation was limited to the case of s wave Schrödinger wavefunctions.

II. SOLUTION OF SCHRÖDINGER EQUATION

The exact s-wave solutions of the Schrödinger equation for an exponential potential are very well known. However, for the sake of completeness, we present here the essential steps. As usual we write the radial part of the s-wave Schrödinger wavefunction $\psi(r)$ as

$$\psi(r) = u(r)/r, \tag{1}$$

where the radial wavefunction $u(r)$ satisfies the differential equation (taking $\hbar = 1$)

$$u'' + 2\mu(E - V(r))u = 0. \tag{2}$$

If we consider the equation for a bound state, then μ is the reduced mass and $\mu = m/2$, where m is the mass of any of the constituent particles assumed to be of equal masses. If we take

$$V(r) = Ae^{-kr} \quad (A < 0) \tag{3}$$

then (2) can be written as

$$v''(z) + (4m/k^2)(E - Ae^{2z})v(z) = 0, \tag{4}$$

where $u(r) = v(z)$ and $z = -kr/2$. The solution of (4) is given by

$$u(r) = a_0 J_\nu(\lambda e^{-kr/2}), \tag{5}$$

where

$$\nu^2 = -4mE/k^2, \quad \lambda^2 = -4Am/k^2 \tag{6}$$

and a_0 is a normalization constant to be determined from the relation

$$\int_0^\infty (u(r))^2\, dr = 1. \tag{7}$$

Eigenvalues are obtained from the boundary condition

$$J_\nu(\lambda) = 0. \tag{8}$$

Now

$$\int_0^\infty J_\nu(\lambda e^{-kr/2}) J_\mu(\lambda e^{-kr/2})\, dr = \frac{2}{k} \int_0^1 J_\mu(\lambda t) J_\nu(\lambda t) t^{-1}\, dt. \tag{9}$$

The right hand side of (9) can be evaluated by using the explicit expressions for integrals like[4]

$$\int^z z^{-1} J_\mu(\lambda z) J_\nu(\lambda z)\, dz.$$

After some straightforward calculation, we see that the right-hand side of (9) is equal to zero if $\mu \neq \nu$ and for $\mu = \nu$

$$\int_0^1 (J_\nu(\lambda t))^2 t^{-1}\, dt = \frac{\lambda}{2\nu}\left(J_{\nu+1} \frac{\partial J_\nu(\lambda)}{\partial \nu}\right), \tag{10}$$

$$\left(\frac{\partial J_\nu(\lambda)}{\partial \nu} \equiv \frac{\partial J_\nu(z)}{\partial \nu}\right)\bigg|_{z=\lambda},$$

where we have also used the result (8).

Hence, orthonormal eigenfunctions for an exponential potential are given by

$$u_n = a_n(\nu) J_{\nu_n}(\lambda e^{-kr/2}), \tag{11}$$

where ν_n is the value of ν corresponding to the nth zero of $J_\nu(\lambda)$ for fixed $\lambda; a_n$ is given by

$$a_n(\nu) = \left(\frac{k\nu}{\lambda}\right)^{1/2} \left(J_{\nu+1}(\lambda) \frac{\partial J_\nu(\lambda)}{\partial \nu}\right)^{-1/2} \tag{12}$$

III. CALCULATION OF THE MASS DEPENDENCE OF WAVEFUNCTIONS

Let us now calculate $P(r)$ defined by

$$P(r) = \int_0^r (u(r))^2\, dr. \tag{13}$$

0022-2488/82/091628-02$02.50

Using (5) we have

$$P(r)=\frac{2}{k}\int_{e^{-kr/2}}^{1}a_0^2(J_\nu(\lambda t)^2)t^{-1}\,dt$$

$$=\frac{2}{k}\int_0^1 a_0^2(J_\nu(\lambda t)^2)t^{-1}\,dt \tag{14}$$

$$-\frac{2}{k}a_0^2\int_0^{e^{-kr/2}}(J_\nu(\lambda t))^2t^{-1}\,dt. \tag{15}$$

Using (10) and (12)

$$P(r)=1-f(m,r), \tag{16}$$

where

$$f(m,r)=\Big[e^{-kr/2}\Big\{J_{\nu+1}(\lambda e^{-kr/2})\frac{\partial J_\nu}{\partial\nu}(\lambda e^{-kr/2}) - J_\nu(\lambda e^{-kr/2})\frac{\partial J_{\nu+1}}{\partial\nu}(\lambda e^{-kr/2})\Big\} + \frac{(J_\nu(\lambda e^{-kr/2})^2}{\lambda})\Big]\div\Big(J_{\nu+1}(\lambda)\frac{\partial J_\nu(\lambda)}{\partial\nu}\Big). \tag{17}$$

If $\partial P(r)/\partial m \geqslant 0$ then $P(r)$ is monotonically increasing with respect to m and hence $f(m,r)$ must be a monotonically decreasing function of m. We show by counterexample that this cannot be true. In calculating ν for given m, one should note that

$$\frac{\partial\nu^2}{\partial m}>0,$$

i.e., ν^2 is a strictly monotonically increasing function of m[5]. This can be shown in the following way. From (6)

$$\nu^2=-4mE/k^2. \tag{18}$$

Hence

$$\frac{\partial\nu^2}{\partial m}=-\frac{4}{k^2}\Big[E+m\frac{\partial E}{\partial m}\Big]. \tag{19}$$

But

$$\frac{\partial E}{\partial m}=-\frac{\langle T\rangle}{m} \tag{20}$$

(from the Feynman–Hellmann theorem)

$$\therefore\frac{\partial\nu^2}{\partial m}=-\frac{4}{k^2}\langle V\rangle>0. \tag{21}$$

Also, ν is determined from the equation

$$J_\nu(\lambda)=0.$$

Now, for our counterexample, we take

$$e^{-kr_0/2}=\frac{j_{1,1}}{j_{1,2}},\quad m_1=-j_{1,1}^2k^2/4A<m_2$$

$$=-j_{2,2}^2k^2/4A, \tag{22}$$

where $j_{\nu,n}$ is the nth zero of $J_\nu(z)$

Hence $\nu_1=1,\nu_2=2$. Using these values, the right-hand side of (17) can be calculated numerically.[6] $\partial J_\nu(z)/\partial\nu$ are calculated by using the formula[7]

$$\Big[\frac{\partial}{\partial\nu}J_\nu(z)\Big]_{\nu=n}=\frac{\pi}{2}Y_n(z) + \frac{n!(z/2)^{-n}}{2}\sum_{k=0}^{n-1}\frac{(z/2)^kJ_k(z)}{(n-k)k!}. \tag{23}$$

It is found that

$$f(m_1,r)=0.773<f(m_2,r)=0.880.$$

Hence, $f(m_1,r)<f(m_2,r)$ for $m_1<m_2$ and the condition that $f(m,r)$ is monotonically decreasing fails to hold.

IV. CONCLUSIONS

As pointed out by Leung and Rosner,[1] the condition $G(r)\geqslant 0, 0\leqslant r<\infty$ is a quantitative statement of the condition that a bound particle falls deeper into the well as μ increases. We have shown by a counterexample that this condition is not satisfied for the case of exponential type potentials. In fact, numerous counterexamples can be found to show that $f(m,r)$ is not a monotonic function.

[1] C. N. Leung and J. Rosner, J. Math. Phys. **20**, 1435 (1979).

[2] A brief but excellent review is given by C. Quigg, in Lectures on Charmed Particles, Fermi Lab. Conf. 78/137THY (1978) (unpublished).

[3] G. Cocconi, Commun. Nucl. Particle Phys. VII **6**, 117 (1978).

[4] L. Y. Luke, *Integrals of Bessel Functions* (McGraw-Hill, New York, 1962), pp. 254–5.

[5] I am grateful to Mr. C. N. Leung of the University of Minnesota for pointing this out and also for some helpful comments.

[6] Numerical values of Bessel functions of the first kind for arguments up to three places of decimals are taken from *The Annals of the Computation Laboratory of Harvard University* (Harvard, Boston, MA, 1947), Vols. III and IV (unpublished). Values of Bessel functions of the second kind are taken from E. A. Chistova, *Tables of Bessel Functions* (Pergamon, New York, 1959).

[7] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970), p. 362.

On the effect of space-time isometries on the neutrino field

Charalampos A. Kolassis

Institut Henri Poincaré, Laboratoire de Physique Théorique, Equipe de Recherche Associée au C.N.R.S. No. 533, 11, rue Pierre et Marie Curie, 75231 Paris Cedex 05, France

(Received 14 November 1981; accepted for publication 16 June 1981)

We consider a neutrino field in interaction with a space-time admitting an isometry group and we attempt to derive the symmetries imposed on the neutrino flux-vector and on the neutrino field for solutions of the Einstein-Weyl field equations. It is proved that if one of the following two constraints is ìmposed, (i) the neutrino field is of class E_1, (ii) the neutrino flux-vector is collinear with one of the principal null directions of the Weyl tensor, so that $\Psi_0 = 0$, then

$$\mathscr{L}_n \xi^A = -(1/2) s \xi^A \quad \text{and} \quad \mathscr{L}_n l^\mu = 0,$$

where ξ^A is the neutrino field, l^μ is the neutrino flux-vector, n^μ is a Killing vector field, and s is a real constant. However, in the cases of a pure-radiation field with diverging rays and a pure-radiation field with nondiverging rays and $\Psi_3 = 0$ the above formulas become

$$\mathscr{L}_n \xi^A = \tfrac{1}{2}(p - is)\xi^A \quad \text{and} \quad \mathscr{L}_n l_\mu = p l^\mu,$$

where now p and s are in general real functions of the coordinates.

PACS numbers: 04.20.Cv

I. INTRODUCTION

To solve the Einstein field equations with a nonvanishing source, very often we have to find approximation methods or at best make restrictive hypotheses on the form of the space-time metric. These hypotheses generally consist in admitting that space-time possesses a certain group of isometries. Aside from these restrictions, in practice it is often convenient to assume that the source of the gravitational field (i.e., a scalar field, an electromagnetic field, etc.) inherits the symmetries of the space-time metric. Naturally a question will arise as to whether this assumption really constitutes a restriction or is just a consequence, through the Einstein field equations, of the hypotheses of symmetry made on the metric. Concerning the scalar field, the electromagnetic field, and the perfect fluid, recently several authors have examined this question.[1] Here, we have the same problem in view but for a spinor field, that of the neutrino.

To proceed, we adopt the 2-spinor formalism which in general seems to be more adequate than the 4-spinor formalism for the treatment of the neutrino field in a curved space-time. The important advantage of this formalism is the possibility of introducing directly from the neutrino field in each point of the space-time a null tetrad in terms of which the Weyl equation and the neutrino energy-momentum tensor take a very convenient form. For the reader's convenience and in order to fix our notation we give in Sec. II a number of preliminary results; however, a familiarity with the spin coefficient formalism of Newman and Penrose[2] is assumed.

II. THE 2-COMPONENT NEUTRINO FIELD IN A CURVED SPACE-TIME

The interaction of a neutrino field and a gravitational field is described in general relativity by the Einstein–Weyl coupled equations

$$R_{\mu\nu} = -T_{\mu\nu}, \tag{II.1}$$

$$\sigma^\mu_{A\dot{X}} \xi^A_{\ ;\mu} = 0, \tag{II.2}$$

where

$$T_{\mu\nu} = (i/4)(\bar{\xi}_{\dot{X}} \sigma^{A\dot{X}}_\mu \xi_{A;\nu} + \bar{\xi}_{\dot{X}} \sigma^{A\dot{X}}_\nu \xi_{A;\mu} - \text{c.c.}) \tag{II.3}$$

is the neutrino energy-momentum tensor. The spinor [Throughout this article we adopt the notations of Pirani.[3] Thus, small Greek indices are tensor indices (values 0,1,2,3) while capital Latin indices are spinor indices (values 1,2). The signature of the space-time is taken to be $-,-,-,+$.] ξ^A is the neutrino field and the quantities $\sigma^\mu_{A\dot{X}}$ are the generalized Pauli matrices satisfying the following anticommutation relations:

$$\sigma^{A\dot{X}}_\mu \sigma_{\nu A\dot{Y}} + \sigma^{A\dot{X}}_\nu \sigma_{\mu A\dot{Y}} = g_{\mu\nu} \delta_{\dot{Y}}^{\ \dot{X}}. \tag{II.4}$$

Following Griffiths and Newing[4] we introduce the 2-spinor χ^A in a way that (ξ^A, χ^A) form a dyad or spinor frame; that is,

$$\xi_A \chi_B - \xi_B \chi_A = \epsilon_{AB}. \tag{II.5}$$

We then construct the null tetrad or null frame:

$$l^\mu = \sigma^\mu{}_{A\dot{X}} \xi^A \bar{\xi}^{\dot{X}}, \tag{II.6a}$$

$$\kappa^\mu = \sigma^\mu{}_{A\dot{X}} \chi^A \bar{\chi}^{\dot{X}}, \tag{II.6b}$$

$$m^\mu = \sigma^\mu{}_{A\dot{X}} \xi^A \bar{\chi}^{\dot{X}}, \tag{II.6c}$$

$$\bar{m}^\mu = \sigma^\mu{}_{A\dot{X}} \chi^A \bar{\xi}^{\dot{X}}. \tag{II.6d}$$

The vector l^μ is interpreted as the neutrino flux-vector. The completeness relation which arises from (II.5) is

$$l_\mu \kappa_\nu + l_\nu \kappa_\mu - m_\mu \bar{m}_\nu - m_\nu \bar{m}_\mu = g_{\mu\nu}. \tag{II.7}$$

For a given ξ^A the most general transformation on χ^A which preserves (II.5) may be written in the form

$$\xi^A = \xi'^A, \tag{II.8a}$$

$$\chi^A = \chi'^A + \Psi \xi'^A, \tag{II.8b}$$

where Ψ is an arbitrary complex function of the coordinates. This transformation generates the following transformation for the corresponding null tetrad, called a "null rotation

0022-2488/82/091630-09$02.50

about l^μ":

$$l^\mu = l'^\mu, \qquad \text{(II.9a)}$$
$$\kappa^\mu = \kappa'^\mu + \Psi m'^\mu + \bar{\Psi}\bar{m}'^\mu + \Psi\bar{\Psi} l'^\mu, \qquad \text{(II.9b)}$$
$$m^\mu = m'^\mu + \bar{\Psi} l'^\mu. \qquad \text{(II.9c)}$$

In terms of the spin coefficients $\alpha, \beta, \gamma, \epsilon, \rho, \kappa, \sigma, \tau$ associated with the null tetrad (see Appendix A), the Weyl equation and the neutrino energy-momentum tensor assume the equivalent tetrad form[5]

$$\rho = \epsilon, \qquad \text{(II.10a)}$$
$$\beta = \tau, \qquad \text{(II.10b)}$$

$$\begin{aligned} T_{\mu\nu} &\\ = -\tfrac{1}{4}\{ &\Lambda l_\mu l_\nu + 2\omega[g_{\mu\nu} - 4l_{(\mu}\kappa_{\nu)}] + 2i[(\alpha - 2\bar{\tau})l_{(\mu}m_{\nu)} \\ &- (\bar{\alpha} - 2\tau)l_{(\mu}\bar{m}_{\nu)} + \kappa\kappa_{(\mu}\bar{m}_{\nu)} - \bar{\kappa}\kappa_{(\mu}\, m_{\nu)} \\ &+ \bar{\sigma} m_\mu m_\nu - \sigma\bar{m}_\mu\bar{m}_\nu]\}\,, \end{aligned} \qquad \text{(II.11)}$$

where

$$\Lambda = 2i(\bar{\gamma} - \gamma), \qquad \text{(II.12)}$$
$$\omega = (i/2)(\rho - \bar{\rho}). \qquad \text{(II.13)}$$

The round brackets on the indices in Eq. (II.11) mean symmetrization. The energy density of a field as measured by an observer with future-pointing velocity u^μ is

$$E(u) = T_{\mu\nu}u^\mu u^\nu.$$

The expression (II.11) for the neutrino energy-momentum tensor can be simplified essentially by the adoption for the neutrino field of the following physically relevant condition:

$$E(u) \neq 0$$

for all the observers at every point in which $T_{\mu\nu} \neq 0$. In the terminology of Wainwright[6] we say that these neutrino fields satisfy the weak energy condition E_1 or equivalently that they are neutrino fields of class E_1. If this condition is fulfilled we can prove[6] the existence of a null tetrad with respect to which the energy-momentum tensor (II.11) reduces to the form

$$\begin{aligned} T_{\mu\nu} = -\tfrac{1}{4}[&\Lambda l_\mu l_\nu + 2\omega g_{\mu\nu} - 4\omega(l_\mu\kappa_\nu + l_\nu\kappa_\mu) \\ &+ 2i\bar{\sigma} m_\mu m_\nu - 2i\sigma\bar{m}_\mu\bar{m}_\nu], \end{aligned} \qquad \text{(II.14)}$$

with the following restrictions on the spin coefficients:

$$\kappa = 0, \qquad \text{(II.15)}$$
$$\alpha - 2\bar{\tau} = 0. \qquad \text{(II.16)}$$

Here, ω and σ are proportional, respectively, to the twist and the shear of the neutrino principal null congruence and they fulfill the relation

$$\sigma\bar{\sigma} - 4\omega^2 \leqslant 0. \qquad \text{(II.17)}$$

If in particular $\sigma\bar{\sigma} - 4\omega^2 < 0$ then the null tetrad is determined uniquely with respect to the null rotations about l^μ and if $\sigma\bar{\sigma} - 4\omega^2 = 0$ then there exists a freedom in the choice of the null tetrad with the restriction that

$$2\omega\Psi - i\bar{\sigma}\bar{\Psi} = 0. \qquad \text{(II.18)}$$

So, in this second case we can perform the null rotation (II.9a)–(II.9c) and give any value to the real part of Ψ.

III. LIE DERIVATIVES OF THE NULL TETRAD OVER A KILLING VECTOR FIELD

Let us consider a null tetrad $(l^\mu, \kappa^\mu, m^\mu, \bar{m}^\mu)$ and a real Killing vector field n^μ. Because the vectors l^μ, κ^μ, m^μ are null, it is obvious that their Lie derivatives with respect to n^μ are given in general by

$$\mathscr{L}_n l^\mu = p l^\mu - q m^\mu - \bar{q}\bar{m}^\mu, \qquad \text{(III.1a)}$$
$$\mathscr{L}_n \kappa^\mu = f\kappa^\mu - r m^\mu - \bar{r}\bar{m}^\mu, \qquad \text{(III.1b)}$$
$$\mathscr{L}_n m^\mu = e l^\mu + g\kappa^\mu + h m^\mu, \qquad \text{(III.1c)}$$

where p, f are real and q, r, e, g, h complex. By Lie differentiation with respect to n^μ of the completeness relation (II.7) and the help of (III.1a)–(III.1c) we obtain

$$f + p = g + \bar{q} = e + \bar{r} = h + \bar{h} = 0. \qquad \text{(III.2)}$$

Conversely, if (III.2) are valid it is clear that n^μ is a Killing vector field. So, we have the following theorem.

Theorem A

A vector field n^μ is a Killing vector field if and only if the Lie derivatives with respect to n^μ of the null tetrad vectors l^μ, κ^μ, m^μ are expressed by the formulas

$$\mathscr{L}_n l^\mu = p l^\mu - q m^\mu - \bar{q}\bar{m}^\mu, \qquad \text{(III.3a)}$$
$$\mathscr{L}_n \kappa^\mu = -p\kappa^\mu - r m^\mu - \bar{r}\bar{m}^\mu, \qquad \text{(III.3b)}$$
$$\mathscr{L}_n m^\mu = -\bar{r} l^\mu - \bar{q}\kappa^\mu - is m^\mu, \qquad \text{(III.3c)}$$

where p, s are real and q, r complex.

Let us now define the Lie derivative of a spinor ξ^A with respect to a vector field n^μ so as to satisfy the following properties[7]:

(i) If ξ^A is a 2-spinor, then $\mathscr{L}_n\xi^A$ is also a 2-spinor.

(ii) $\mathscr{L}_n\sigma^\mu_{A\dot{X}} = 0$ if and only if n^μ is a Killing vector field.

Thus using (II.6a)–(III.6d) the spinor equivalents of (III.3a)–(III.3c) are given by

$$\mathscr{L}_n\xi^A = \tfrac{1}{2}(p - is)\xi^A - \bar{q}\chi^A, \qquad \text{(III.4a)}$$
$$\mathscr{L}_n\chi^A = -r\xi^A - \tfrac{1}{2}(p - is)\chi^A. \qquad \text{(III.4b)}$$

If the 2-spinor ξ^A is the neutrino field, then from (III.3a) and (III.4a) we observe that the symmetry properties of the neutrino flux-vector and of the neutrino field depend critically on whether or not the quantity q vanishes. However, under a null rotation about l^μ the quantities q, p, s, r are transformed as follows:

$$q' = q, \qquad \text{(III.5a)}$$
$$p' = p - \bar{q}\Psi - q\bar{\Psi}, \qquad \text{(III.5b)}$$
$$is' = is + \bar{q}\Psi - q\bar{\Psi}, \qquad \text{(III.5c)}$$
$$r' = r + (p - is)\Psi - \bar{q}\Psi^2 + \mathscr{L}_n\Psi. \qquad \text{(III.5d)}$$

From (III.5a) q cannot be put to zero in general. If it happens to vanish, then p and s become invariant with respect to (II.9a)–(II.9c).

It is well known[8] that the Lie derivatives of the Weyl tensor $C_{\mu\nu\rho\sigma}$ and of the Ricci tensor $R_{\mu\nu}$ with respect to a Killing vector field are zero:

$$\mathscr{L}_n C_{\mu\nu\rho\sigma} = 0,$$
$$\mathscr{L}_n R_{\mu\nu} = 0,$$

or equivalently,

$$\mathscr{L}_n \Psi_{ABCD} = 0, \qquad \text{(III.6)}$$
$$\mathscr{L}_n \Phi_{AB\dot{X}\dot{Y}} = 0, \qquad \text{(III.7)}$$

where Ψ_{ABCD} and $\Phi_{AB\dot{X}\dot{Y}}$ are the spinor equivalents of the Weyl tensor and of the Ricci tensor, respectively.

Considering the dyad components of the Weyl spinor [see Appendix A, Eqs. (A5)] after a straightforward calculation and with the help of (III.4a) and (III.4b) and (III.6) we obtain

$$\mathscr{L}_n\Psi_0 = 2(p - is)\Psi_0 - 4\bar{q}\Psi_1, \tag{III.8a}$$
$$\mathscr{L}_n\Psi_1 = -r\Psi_0 + (p - is)\Psi_1 - 3\bar{q}\Psi_2, \tag{III.8b}$$
$$\mathscr{L}_n\Psi_2 = -2r\Psi_1 - 2\bar{q}\Psi_3, \tag{III.8c}$$
$$\mathscr{L}_n\Psi_3 = -3r\Psi_2 - (p - is)\Psi_3 - \bar{q}\Psi_4, \tag{III.8d}$$
$$\mathscr{L}_n\Psi_4 = -4r\Psi_3 - 2(p - is)\Psi_4. \tag{III.8e}$$

Let us now consider the dyad components Φ_{ij} of the Ricci spinor [see Appendix A, Eqs. (A6)]. By a straightforward calculation and with the help of (III.4a)–(III.4b) and (III.7) we obtain

$$\mathscr{L}_n\Phi_{00} = 2(p\Phi_{00} - q\Phi_{01} - \bar{q}\bar{\Phi}_{01}), \tag{III.9a}$$
$$\mathscr{L}_n\Phi_{11} = -q\Phi_{12} - \bar{q}\bar{\Phi}_{12} - r\Phi_{01} - \bar{r}\bar{\Phi}_{01}, \tag{III.9b}$$
$$\mathscr{L}_n\Phi_{01} = (p - is)\Phi_{01} - 2q\Phi_{11} - q\Phi_{02} - r\Phi_{00}, \tag{III.9c}$$
$$\mathscr{L}_n\Phi_{12} = -(p + is)\Phi_{12} - r\Phi_{02} - \bar{q}\Phi_{22} - 2\bar{r}\Phi_{11}, \tag{III.9d}$$
$$\mathscr{L}_n\Phi_{02} = -2(\bar{r}\Phi_{01} + \bar{q}\Phi_{12} + is\Phi_{02}), \tag{III.9e}$$
$$\mathscr{L}_n\Phi_{22} = -2(p\Phi_{22} + r\Phi_{12} + \bar{r}\bar{\Phi}_{12}). \tag{III.9f}$$

From the Einstein field.equations (II.1) and the expression (II.11) for the neutrino energy-momentum tensor we derive the following equations which can be considered as the dyad components of the Einstein field equations.

$$\Phi_{00} = 0, \tag{III.10a}$$
$$\Phi_{01} = (i/8)\kappa, \tag{III.10b}$$
$$\Phi_{02} = (i/4)\sigma, \tag{III.10c}$$
$$\Phi_{11} = \tfrac{1}{4}\omega, \tag{III.10d}$$
$$\Phi_{12} = -(i/8)(\bar{\alpha} - 2\tau), \tag{III.10e}$$
$$\Phi_{22} = -\tfrac{1}{8}\Lambda, \tag{III.10f}$$

By substitution of (III.10a)–(III.10f) into (III.9a)–(III.9f) we obtain

$$q\kappa - \bar{q}\,\bar{\kappa} = 0, \tag{III.11a}$$
$$\mathscr{L}_n\omega = (i/2)[q(\bar{\alpha} - 2\tau) - \bar{q}(\alpha - 2\bar{\tau}) + \overline{r\kappa} - r\kappa], \tag{III.11b}$$
$$\mathscr{L}_n\kappa = (p - is)\kappa + 2i(2\omega\bar{q} + i\sigma q), \tag{III.11c}$$
$$\mathscr{L}_n(\bar{\alpha} - 2\tau) = -(p + is)(\bar{\alpha} - 2\tau) - 2i(2\omega\bar{r} + i\sigma r - \tfrac{1}{2}\Lambda\bar{q}), \tag{III.11d}$$
$$\mathscr{L}_n\sigma = \bar{q}(\bar{\alpha} - 2\tau) - \bar{r}\kappa - 2is\sigma, \tag{III.11e}$$
$$\mathscr{L}_n\Lambda = -2p\Lambda + 2i\bar{r}(\alpha - 2\bar{\tau}) - 2ir(\bar{\alpha} - 2\tau). \tag{III.11f}$$

It must be noted that equations (III.11a)–(III.11f) can also be considered as the intergrability conditions $\mathscr{L}_nR_{\mu\nu} = -\mathscr{L}_nT_{\mu\nu}$ of the Einstein field equations.

For the neutrino fields of class E_1, we must introduce in the equations (III.11a)–(III.11f) the restrictions (II.15) and (II.16). We obtain thus,

$$\mathscr{L}_n\omega = 0, \tag{III.12a}$$
$$2\omega\bar{q} + i\sigma q = 0, \tag{III.12b}$$
$$2\omega\bar{r} + i\sigma r - \tfrac{1}{2}\Lambda\bar{q} = 0, \tag{III.12c}$$
$$\mathscr{L}_n\sigma = -2is\sigma, \tag{III.12d}$$
$$\mathscr{L}_n\Lambda = -2p\Lambda. \tag{III.12e}$$

In the next sections, in order to exploit the above equations we will use extensively and often without explicit reference the various restrictions satisfied by the spin coefficients [e.g. Eqs. (II.10a) and (II.10b), (II.15), (II.16), etc.] and the Einstein equations (III.10a)–(III.10f). Also, we will refer to some of the Ricci and Bianchi identities. For these, we will adopt a special reference notation; e.g., by (R.2) [Respectively, (B.2)] we will mean the second Ricci identity (respectively, second Bianchi identity) in the listing given by Pirani[9] or by Flaherty.[10] However, for the reader's convenience all the Ricci and Bianchi identities which are used in this paper together with the commutation relations of the D, Δ, δ, and $\bar{\delta}$ differential operators are displayed in Appendix A.

IV. THE EFFECT OF SPACE-TIME ISOMETRIES ON THE NEUTRINO FIELDS OF CLASS E_1

Because of Eqs. (III.12b), (III.12c), and the restriction (II.17) satisfied by the neutrino fields of class E_1, we must separately consider the two cases $4\omega^2 - \sigma\bar{\sigma} > 0$ and $4\omega^2 - \sigma\bar{\sigma} = 0$. The pure radiation field ($\omega = 0$ and $\sigma = 0$) will be considered as the third case. Finally, the results of the three cases may be resumed in Theorem B given below.

Case 1

We consider the case where

$$4\omega^2 - \sigma\bar{\sigma} > 0. \tag{IV.1.1}$$

So, from (III.12b) and (III.12c) follows that

$$q = 0, \tag{IV.1.2}$$
$$r = 0. \tag{IV.1.3}$$

In order to exploit Eqs. (III.12a), (III.12d), and (III.12e) we must apply a theorem which states that the Lie derivative with respect to a Killing vector field commutes with the covariant derivative.[8] Here, and in the remainder of the paper this theorem in conjunction with Eqs. (III.3a)–(III.3c) will be used extensively without explicit reference.

First, let us consider Eq. (III.12a). This can be written in the form

$$\mathscr{L}_n[(l_{\mu;\nu} - l_{\nu;\mu})m^\mu\bar{m}^\nu] = 0.$$

Inserting (III.3a) and (III.3c) together with (IV.1.2) and (IV.1.3) into the above expression yields

$$p\omega = 0.$$

Since $\omega = 0$ is excluded from (IV.1.1), we obtain

$$p = 0. \tag{IV.1.4}$$

Now Eq. (III.12d) becomes an identity and we must consider Eq. (III.12e). This, with a similar procedure as previously and with the help of (IV.1.2), (IV.1.3), and (IV.1.4), yields

$$\Delta s = 0. \tag{IV.1.5}$$

Further restrictions on s can be obtained from the Lie derivatives with respect to n^μ of the Weyl equation and the help of (IV.1.2), (IV.1.3), and (IV.1.4). Thus, from $\mathscr{L}_n\epsilon = \mathscr{L}_n\rho$ we obtain

$$Ds = 0, \tag{IV.1.6}$$

and from $\mathscr{L}_n\beta = \mathscr{L}_n\tau$ we obtain

$$\delta s = 0. \tag{IV.1.7}$$

[For the definition of the Δ, D, and δ operators, see Appendix A, Eqs. (A2)]. Equations (IV.1.5), (IV.1.6), and (IV.1.7) imply that s is constant. Now, by virtue of (IV.1.2), (IV.1.3), (IV.1.4), and (IV.1.7) the equations $\mathscr{L}_n \kappa = 0$ and $\mathscr{L}_n(\alpha - 2\bar{\tau}) = 0$ become an identity and therefore we cannot derive further restrictions on s.

Case 2

We consider the case where

$$4\omega^2 - \sigma\bar{\sigma} = 0, \qquad \text{(IV.2.1)}$$

with

$$\omega \neq 0 \text{ and } \sigma \neq 0. \qquad \text{(IV.2.2)}$$

To prove that q vanishes we will proceed by contradiction. So, let us assume that

$$q \neq 0. \qquad \text{(IV.2.3)}$$

By virtue of (IV.2.1) and (IV.2.3) Eq. (III.12c) splits into the equations

$$2\omega\bar{r} + i\sigma r = 0, \qquad \text{(IV.2.4)}$$
$$\gamma = \bar{\gamma}. \qquad \text{(IV.2.5)}$$

From (IV.2.1) it follows that we can perform the null rotation (II.9a)–(II.9c) subject to the restriction (II.18). Under such a transformation and by virtue of (III.12b), (III.12d), and (IV.2.4) we may observe that the quantities ω, σ, $\gamma - \bar{\gamma}$, q, s, and arg$r$ are invariant [see Appendix A, Eqs. (A4a)–(A4e)]. However, by virtue of (IV.2.3) we can always choose the real part of Ψ so that

$$p = 0. \qquad \text{(IV.2.6)}$$

Now, let us consider Eqs. (III.12a), (III.12d) and the integrability conditions

$\mathscr{L}_n\epsilon = \mathscr{L}_n\rho$, $\mathscr{L}_n\beta = \mathscr{L}_n\tau$, $\mathscr{L}_n\kappa = \mathscr{L}_n(\alpha - 2\bar{\tau}) = 0$.

From these, if we take into account (III.12b), (IV.2.4), (IV.2.5), and (IV.2.6) we can derive the following equations:

$$\delta q = 3q\tau - \tfrac{1}{2}[q(\bar{\pi} + \tau) + \bar{q}(\pi + \bar{\tau})], \qquad \text{(IV.2.7)}$$
$$\bar{\delta} q = 3q\bar{\tau}, \qquad \text{(IV.2.8)}$$
$$3\Delta q = 4q\gamma - q\bar{\mu} - \bar{q}\lambda, \qquad \text{(IV.2.9)}$$
$$Dq = q(2\bar{\rho} + \rho). \qquad \text{(IV.2.10)}$$

From (R.1) follows

$$D\omega = 2(\rho + \bar{\rho})\omega. \qquad \text{(IV.2.11)}$$

By virture of (IV.2.10), (R.2), (IV.2.3), and (IV.2.11) the D differentiation of (III.12b) gives

$$\Psi_0 = 2i\omega\sigma. \qquad \text{(IV.2.12)}$$

By substituting (IV.2.12) into (III.8a) and (III.8b) and using (III.12a), (III.12d), (IV.2.3), (IV.2.6), and (III.12b) we obtain

$$\Psi_1 = 0, \qquad \text{(IV.2.13)}$$
$$\Psi_2 = \frac{4\bar{r}}{3\bar{q}}\omega^2. \qquad \text{(IV.2.14)}$$

We can observe that from (III.12b), (IV.2.4), and (IV.2.14) follows

$$\Psi_2 = \bar{\Psi}_2. \qquad \text{(IV.2.15)}$$

From (R.3), (R.4), (R.5), and (R.11) we derive

$$\bar{\delta}\omega = -2\omega(\pi + \bar{\tau}) + i\bar{\sigma}(\bar{\pi} + \tau) + 4\omega\bar{\tau}, \qquad \text{(IV.2.16)}$$
$$\bar{\delta}\sigma = 4\sigma\bar{\tau}. \qquad \text{(IV.2.17)}$$

With the help of (IV.2.12), (IV.2.16), and (IV.2.17) Eq. (B1) yields

$$2\omega(\bar{\pi} + \tau) + i\sigma(\pi + \bar{\tau}) - \omega\bar{\pi} = 0. \qquad \text{(IV.2.18)}$$

On the other hand by $\bar{\delta}$ differentiation of (III.12b) and the help of (IV.2.7), (IV.2.8), (IV.2.16), and (IV.2.17) we obtain

$$10\omega(\bar{\pi} + \tau) + 3i\sigma(\pi + \bar{\tau}) = 0. \qquad \text{(IV.2.19)}$$

By virtue of (IV.2.1) Eqs. (IV.2.18) and (IV.2.19) yield

$$\tau = 0, \qquad \text{(IV.2.20)}$$
$$\pi = 0. \qquad \text{(IV.2.21)}$$

By Lie differentiation with respect to n^μ of (IV.2.20) and with the help of (IV.2.4) we obtain

$$\Delta\bar{q} = 2\bar{q}\gamma + \overline{r\rho}. \qquad \text{(IV.2.22)}$$

By Lie differentiation with respect to n^μ of (IV.2.21) and with the help of (IV.2.5), (IV.2.9), and (IV.2.22) we obtain

$$Dr = q(\bar{\mu} - \mu + 2\gamma) + r\rho. \qquad \text{(IV.2.23)}$$

With the help of (III.12b), (R.2), (IV.2.5), (IV.2.11), (IV.2.12), and (IV.2.23) the D differentiation of (IV.2.4) yields

$$r(\rho - \bar{\rho}) = q(\mu - \bar{\mu}). \qquad \text{(IV.2.24)}$$

Now, by virtue of (IV.2.24), Eq. (IV.2.14) can be written in the form

$$\Psi_2 = -\tfrac{1}{3}(\rho - \bar{\rho})(\mu - \bar{\mu}). \qquad \text{(IV.2.25)}$$

Acting on q with the commutator of the δ- and $\bar{\delta}$-differential operators and taking into account Eqs. (IV.2.7), (IV.2.8), (IV.2.10), (IV.2.20), (IV.2.21), (IV.2.22), and (IV.2.24) we obtain

$$\gamma(\rho - \bar{\rho}) = (\rho + \bar{\rho})(\bar{\mu} - \mu). \qquad \text{(IV.2.26)}$$

With the help of (IV.2.5), (IV.2.15), and (IV.2.20) the Ricci identities (R.6), (R.12), and (R.17) yield

$$\gamma(\rho - \bar{\rho}) = \bar{\rho}\bar{\mu} - \rho\mu,$$
$$\lambda\sigma - \bar{\lambda}\bar{\sigma} = \bar{\rho}\mu - \rho\bar{\mu}.$$

These equations together with (IV.2.26) yield

$$\rho\bar{\mu} = \bar{\rho}\mu, \qquad \text{(IV.2.27)}$$
$$\lambda\sigma = \bar{\lambda}\bar{\sigma}. \qquad \text{(IV.2.28)}$$

With the use of (IV.2.15), (IV.2.21), (IV.2.27), and (IV.2.28) we can derive from (R.8)

$$D(\mu - \bar{\mu}) = -(\rho + \bar{\rho})(\mu - \bar{\mu}). \qquad \text{(IV.2.29)}$$

Now with the help of (IV.2.11) and (IV.2.29) and by D differentiation of (IV.2.25) we obtain

$$D\Psi_2 = (\rho + \bar{\rho})\Psi_2. \qquad \text{(IV.2.30)}$$

By virtue of (IV.2.11), (IV.2.13), and (IV.2.15) the real part of (B.3) becomes

$$2D\Psi_2 = 3(\rho + \bar{\rho})\Psi_2 + 2(\rho + \bar{\rho})\Phi_{11}.$$

Comparing this with (IV.2.30), we obtain

$$(\rho + \bar{\rho})(\Psi_2 + 2\Phi_{11}) = 0.$$

Now, if $\rho + \bar{\rho} = 0$, then by virtue of (R.1) and (IV.2.1) follows $\rho = \omega = 0$, which contradicts (IV.2.2). On the other hand, if $\Psi_2 + 2\Phi_{11} = 0$, then by D differentiation of this equation and the help of (IV.2.11) and (IV.2.30) we arrive again at the contradiction $\Phi_{11} = \omega = 0$. So we must conclude that

$$q = 0.$$

Now by similar calculations as in Case 1 we may derive from (III.12a) that

$$p = 0,$$

and from (III.12e) and the Lie derivative with respect to n^μ of the Weyl equation (II.10a) and (II.10b) that

$$s = \text{constant}.$$

Case 3

We consider the case where

$$\omega = 0, \qquad \text{(IV.3.1)}$$
$$\sigma = 0, \qquad \text{(IV.3.2)}$$

i.e., the neutrino field is a pure radiation field. Because of this, Eqs. (III.12a), (III.12b), and (III.12d) become identities and from (III.12c) follows

$$q = 0. \qquad \text{(IV.3.3)}$$

Otherwise the neutrino field would be a ghost field. These fields, because they do not interact with space-time, are excluded from our discussion.

From Eq. (III.12e) and the integrability conditions $\mathscr{L}_n\epsilon = \mathscr{L}_n\rho$, $\mathscr{L}_n\beta = \mathscr{L}_n\tau$, $\mathscr{L}_n(\alpha - 2\bar{\tau}) = 0$, we obtain

$$Ds = 0, \qquad \text{(IV.3.4)}$$
$$\delta s = 0, \qquad \text{(IV.3.5)}$$
$$Dp = 0, \qquad \text{(IV.3.6)}$$
$$\delta p = 0, \qquad \text{(IV.3.7)}$$
$$\Delta s = \tfrac{1}{2}p\Lambda. \qquad \text{(IV.3.8)}$$

The condition $\mathscr{L}_n\kappa = 0$ is irrelevant and therefore Eqs. (IV.3.4)–(IV.3.8) constitute all the restrictions which we can formulate on s and p. The investigation of these equations in conjunction with the commutation relations of the D, Δ, δ, and $\bar{\delta}$ differential operators shows that they are compatible both with the cases of a pure radiation neutrino field with diverging rays (i.e., $\rho \neq 0$) and with a pure radiation neutrino field with nondiverging rays (i.e., $\rho \neq 0$) and $\Psi_3 = 0$ without any further restriciton on p and s. Thus, in these cases we have

$$\mathscr{L}_n l^\mu = pl^\mu,$$
$$\mathscr{L}_n \xi^A = \tfrac{1}{2}(p - is)\xi^A,$$

where p and s are in general real functions of the coordinates. In fact, as is shown in Appendix B, this agrees with the results of Collinson and Morris.[11]

Let us now consider the case which is characterized by the equations

$$\rho = 0, \qquad \text{(IV.3.9)}$$
$$\Psi_3 \neq 0. \qquad \text{(IV.3.10)}$$

It is well known that pure radiation neutrino fields can exist only in space-times with algebraically specialized Weyl tensors,[12] i.e.,

$$\Psi_0 = \Psi_1 = 0. \qquad \text{(IV.3.11)}$$

With the help of (IV.3.2), (IV.3.9), (IV.3.11), and the other restrictions satisfied by the spin coefficients, Eqs. (R.12), (R.16), (R.17), and (B.4) yield

$$\Psi_2 = 0, \qquad \text{(IV.3.12)}$$
$$\tau = 0. \qquad \text{(IV.3.13)}$$

Here it must be noted that because of (IV.3.11) and (IV.3.12) the condition (IV.3.10) is preserved under the null rotation (II.9a)–(II.9c) [see Appendix A, Eqs. (A7)]. By virtue of(IV.3.2), (IV.3.9), and (IV.3.13) Eqs. (R.15) and (R.18) take the form

$$\delta\gamma = 0, \qquad \text{(IV.3.14)}$$
$$\delta\bar{\gamma} = \bar{\Psi}_3. \qquad \text{(IV.3.15)}$$

Now, acting on s with the commutator of the δ and Δ differential operators and using (IV.3.4), (IV.3.5), (IV.3.7), (IV.3.8), (IV.3.13), (IV.3.14) and (IV.3.15) we obtain

$$p\bar{\Psi}_3 = 0,$$

and thus, because of (IV.3.10),

$$p = 0.$$

From this last equation and Eqs. (IV.3.4), (IV.3.5), and (IV.3.8) follows that

$$s = \text{constant}.$$

The results of the three cases can be recapitulated in the following theorem.

Theorem B

If a neutrino field of class E_1 interacts according to the Einstein–Weyl coupled equations with a gravitational field which admits a Killing vector field n^μ, then

$$\mathscr{L}_n \xi^A = -(i/2)s\xi^A,$$
$$\mathscr{L}_n l^\mu = 0,$$

where ξ^A is the neutrino field, l^μ is the neutrino flux-vector, and s is a real constant. However, in the cases of a pure-radiation neutrino field with diverging rays ($\rho \neq 0$) and a pure-radiation neutrino field with nondiverging rays ($\rho = 0$) and $\Psi_3 = 0$ and only in these cases, we have

$$\mathscr{L}_n \xi^A = \tfrac{1}{2}(p - is)\xi^A,$$
$$\mathscr{L}_n l^\mu = pl^\mu,$$

where now p and s are in general real functions of the coordinates.

V. THE EFFECT OF SPACE-TIME ISOMETRIES ON THE NEUTRINO FIELDS WITH $\Psi_0 = 0$

In this section the general neutrino field is considered in interaction with a space-time admitting a Killing vector field n^μ and subject to the restriction that the neutrino flux-vector be collinear with one principal null direction of the Weyl tensor; that is,

$$\Psi_0 = 0. \qquad \text{(V.1)}$$

Of course, this restriction is preserved under the null rotation (II.9a)–(II.9c) [see Appendix A, Eqs. (A7)]. From (V.1) and (III.8a)–(III.8c) it is easily seen that unless the space-time is flat,

$$q = 0. \qquad \text{(V.2)}$$

Considering the integrability conditions

$\mathscr{L}_n\epsilon = \mathscr{L}_n\rho$, $\mathscr{L}_n\beta = \mathscr{L}_n\tau$ of the Weyl equation and taking into account Eq. (V.2) we derived by a straightforward calculation

$$Dp = 0, \qquad \text{(V.3)}$$

$$Ds = 0, \qquad \text{(V.4)}$$

$$\delta p - i\delta s = 0. \qquad \text{(V.5)}$$

Also, with a direct calculation and with the help of (V.2), we obtain

$$\mathscr{L}_n(\bar{\alpha} - 2\tau) = \tfrac{1}{2}\delta(p + is) - is(\bar{\alpha} - 2\tau) - 2i(2\omega\bar{r} + i\sigma r),$$

$$\mathscr{L}_n\Lambda = -2\Delta s - p\Lambda + 2i\bar{r}(\alpha - 2\bar{\tau}) - 2ir(\bar{\alpha} - 2\tau),$$

$$\mathscr{L}_n\kappa = (2p - is)\kappa,$$

$$\mathscr{L}_n\omega = p\omega + (i/2)(\overline{r\kappa} - r\kappa),$$

$$\mathscr{L}_n\sigma = p\sigma - \bar{r}\kappa - 2is\sigma.$$

The comparison of the above equations with Eqs. (III.11a)–(III.11f) and the use of (V.2) yield

$$\delta p + i\delta s = -2p(\bar{\alpha} - 2\tau), \qquad \text{(V.6)}$$

$$\Delta s = \tfrac{1}{2}p\Lambda, \qquad \text{(V.7)}$$

$$p\kappa = 0, \qquad \text{(V.8)}$$

$$p\omega = 0, \qquad \text{(V.9)}$$

$$p\sigma = 0. \qquad \text{(V.10)}$$

If $p = 0$ then from (V.4), (V.5), and (V.7) follows that

$$s = \text{const}$$

and therefore a theorem similar to Theorem B is established.

There remains to investigate the possibility

$$p \neq 0. \qquad \text{(V.11)}$$

We will show below that this assumption leads to a contradiction.

At first, from (V.8)–(V.11) follows

$$\kappa = 0, \qquad \text{(V.12)}$$

$$\omega = 0, \qquad \text{(V.13)}$$

$$\sigma = 0. \qquad \text{(V.14)}$$

These equations imply that the quantity $\alpha - 2\bar{\tau}$ remains invariant under the null rotation (II.9a)–(II.9c). Furthermore, if $\alpha - 2\bar{\tau} = 0$, then the neutrino field would be a pure radiation field which is considered in the preceding section. So, we must assume that

$$\alpha - 2\bar{\tau} \neq 0. \qquad \text{(V.15)}$$

Equations (V.12) and (V.14) together with (R.3), (R.5), and (R.11) imply that

$$\Psi_1 = 0. \qquad \text{(V.16)}$$

Taking into account Eqs. (V.2) and (V.16) Eq. (III.8c) can be written in the form

$$\mathscr{L}_n\Psi_2 = 0. \qquad \text{(V.17)}$$

Acting on s with the commutator of the D and Δ differential operators and using Eqs. (V.3), (V.4), (V.5), (V.6), (V.7), (V.11), (V.12), and (V.13) we obtain

$$\Psi_2 = \bar{\Psi}_2. \qquad \text{(V.18)}$$

As by virtue of (V.12) the spin coefficient ρ is invariant under the null rotation (II.9a)–(II.9c) [see Appendix A, Eq. (A4f)] it seems useful for our proof to distinguish the two cases $\rho \neq 0$ and $\rho = 0$ (i.e., neutrino fields with diverging rays and neutrino fields with nondiverging rays, respectively). Equations (V.1), (V.13), (V.14), (V.16), and (V.18) will be used in what follows without explicit reference.

Case 1

Let us assume that

$$\rho \neq 0. \qquad \text{(V.1.19)}$$

Then, using the null rotation (II.9a)–(II.9c), we make

$$\tau = 0, \qquad \text{(V.1.20)}$$

and consequently (V.15) is reduced to

$$\alpha \neq 0. \qquad \text{(V.1.21)}$$

By virtue of (V.2), (V.14), and (V.1.20) the Lie derivative with respect to n^μ of (V.1.20) yields

$$r = 0. \qquad \text{(V.1.22)}$$

From (R.17), (V.1.19), and (V.1.20) follows

$$\mu = \bar{\mu}. \qquad \text{(V.1.23)}$$

With the help of (V.5), (V.6), (V.11), (V.1.20), (V.1.23), and (R.12) the action on s of the commutator of the δ and $\bar{\delta}$ differential operators yields

$$\Psi_2 = \rho\mu - \alpha\bar{\alpha}. \qquad \text{(V.1.24)}$$

With a direct calculation and the help of (V.2) and (V.1.22) we derive

$$\mathscr{L}_n\rho = p\rho,$$

$$\mathscr{L}_n\mu = -p\mu.$$

Also, from (III.11d) and (V.1.20) follows

$$\mathscr{L}_n\alpha = (is - p)\alpha.$$

Thus the Lie derivative with respect to n^μ of (V.1.24) can be written

$$\mathscr{L}_n\Psi_2 = 2p\alpha\bar{\alpha}, \qquad \text{(V.1.25)}$$

which by virtue of (V.17) is reduced to

$$p\alpha\bar{\alpha} = 0.$$

But this last equation contradicts (V.11) and (V.1.21).

Case 2

Let us assume that

$$\rho = 0. \qquad \text{(V.2.19)}$$

Acting on s with the commutator of the δ and $\bar{\delta}$ differential operators and using (V.4), (V.5), (V.6), (V.11), (V.2.19), (R.12), and (R.17) we obtain

$$2\Psi_2 = -(\alpha - 2\bar{\tau})(\bar{\alpha} - 2\tau) - \tau\bar{\tau}. \qquad \text{(V.2.20)}$$

By a direct calculation using (V.2) and (V.2.19) we can derive

$$\mathscr{L}_n\tau = -is\tau.$$

Also, by virtue of (V.2) Eq. (III.11d) reduces to

$$\mathscr{L}_n(\bar{\alpha} - 2\tau) = -(p + is)(\bar{\alpha} - 2\tau).$$

Thus the Lie derivative with respect to n^μ of (V.2.20) can be written

$$\mathscr{L}_n\Psi_2 = p(\alpha - 2\bar{\tau})(\bar{\alpha} - 2\tau), \qquad \text{(V.2.21)}$$

which by virtue of (V.17) reduces to

$$p(\alpha - 2\bar{\tau})(\bar{\alpha} - 2\tau) = 0.$$

But this last equation contradicts (V.11) and (V.15).

So we have proven the following theorem.

Theorem C

If a neutrino field interacts according to the Einstein–Weyl coupled equations with a gravitational field which admits a Killing vector field n^μ and if the neutrino flux-vector is collinear with one principal null direction of the Weyl tensor, so that $\Psi_0 = 0$, then

$$\mathscr{L}_n \xi^A = -(i/2) s \xi^A$$

and

$$\mathscr{L}_n l^\mu = 0,$$

where ξ^A is the neutrino field, l^μ is the neutrino flux-vector, and s is a real constant. However, in the cases of a pure-radiation neutrino field with diverging rays ($\rho \neq 0$) and a pure-radiation neutrino field with nondiverging rays ($\rho = 0$), and $\Psi_3 = 0$ and only in these cases, we have

$$\mathscr{L}_n \xi^A = \tfrac{1}{2}(p - is)\xi^A,$$
$$\mathscr{L}_n l^\mu = p l^\mu,$$

where now p and s are in general real functions of the coordinates.

VI. CONCLUSION

With a method closely related to the spin coefficient formalism we have considered the effect of space-time isometries on the neutrino field. For a space-time admitting a Killing vector field n^μ and in a coordinate system adapted to this vector field (i.e., such that $n^\mu = \delta^\mu_{(\nu)}$) the main results of this paper are

(i) the neutrino flux-vector is independent of the coordinate x^ν,

(ii) the neutrino field ξ^A has the plane-wave form

$$\xi^A = e^{-(i/2)sx^\nu} \varphi^A,$$

where φ^A is a 2-spinor independent of the coordinate x^ν and where the "frequency" s is a real constant.

It is worthwhile remarking that these results fail to hold for the pure-radiation neutrino fields. This must be related to the fact that pure-radiation fields are not uniquely determined from the space-time metric. Furthermore, we do not know if the above results are valid for neutrino fields not belonging to the energy class E_1 and such that $\Psi_0 \neq 0$. This case will be investigated in a future work.

Recent experiments[13] seem to indicate that neutrinos are not massless. If that is the case, Weyl's equation will be an approximation for the description of their behavior and therefore we should consider the Dirac equation. However, it must be noted that the method used here for the study of the effect of space-time isometries on the zero rest-mass neutrino field can be applied also in the case of a nonzero rest-mass neutrino field for which only the restriction on the helicity is retained (i.e., the neutrino field is described by a 2-spinor). For this it is necessary to formulate the Dirac equation and the energy-momentum tensor of the Dirac field in the 2-spinorial formalism and in particular in terms of spin coefficients. This and the energy conditions for the Dirac field are examined by Griffiths[14] and Radford and Klotz.[15] We intend to investigate this question in a future paper. Finally we would like to remark that, as was pointed out by Henneaux,[16] the above results (i) and (ii) must be valid at least under some restrictions (unknown for the moment) and for the general Dirac field.

APPENDIX A

For convenience, we list the definitions of the spin coefficients and of the D, Δ, δ, $\bar\delta$ differential operators with respect to the null tetrad (l^μ, κ^μ, m^μ, $\tilde{m}^\mu$) as defined by Newman and Penrose.[17]

$$\begin{aligned}
&\kappa = l_{\mu;\nu} m^\mu l^\nu, \quad \pi = \bar m_{\mu;\nu} \kappa^\mu l^\nu, \quad \epsilon = \tfrac{1}{2}(l_{\mu;\nu}\kappa^\mu + \bar m_{\mu;\nu} m^\mu) l^\nu,\\
&\rho = l_{\mu;\nu} m^\mu \bar m^\nu, \quad \lambda = \bar m_{\mu;\nu}\kappa^\mu \bar m^\nu, \quad \alpha = \tfrac{1}{2}(l_{\mu;\nu}\kappa^\mu \cdot + \bar m_{\mu;\nu} m^\mu)\bar m^\nu,
\end{aligned} \tag{A1}$$

$$\begin{aligned}
&\sigma = l_{\mu;\nu} m^\mu m^\nu, \quad \mu = \bar m_{\mu;\nu}\kappa^\mu m^\nu, \quad \beta = \tfrac{1}{2}(l_{\mu;\nu}\kappa^\mu + \bar m_{\mu;\nu} m^\mu) m^\nu,\\
&\tau = l_{\mu;\nu} m^\mu \kappa^\nu, \quad \nu = \bar m_{\mu;\nu}\kappa^\mu \kappa^\nu, \quad \gamma = \tfrac{1}{2}(l_{\mu;\nu}\kappa^\mu + \bar m_{\mu;\nu} m^\mu)\kappa^\nu.\\
&D = l^\mu \nabla_\mu, \quad \Delta = \kappa^\mu \nabla_\mu, \quad \delta = m^\mu \nabla_\mu, \quad \bar\delta = \bar m^\mu \nabla_\mu.
\end{aligned} \tag{A2}$$

where ∇_μ means covariant differentiation.

With the help of the Weyl equation in its tetrad form (II.10a)–(II.10b) the commutation equations of the D, Δ, δ, and $\bar\delta$ operators acting on scalars are written in the form[18]

$$\begin{aligned}
\Delta D - D\Delta &= (\gamma + \bar\gamma)D + (\rho + \bar\rho)\Delta - (\tau + \bar\pi)\bar\delta - (\bar\tau + \pi)\delta,\\
\delta D - D\delta &= (\bar\alpha + \tau - \bar\pi)D + \kappa\Delta - \sigma\bar\delta - \rho\delta,\\
\delta\Delta - \Delta\delta &= -\bar\nu D - \bar\alpha\Delta + \overline{\lambda\delta} + (\mu - \gamma + \bar\gamma)\delta,\\
\bar\delta\delta - \delta\bar\delta &= (\bar\mu - \mu)D + (\bar\rho - \rho)\Delta - (\bar\alpha - \tau)\bar\delta - (\bar\tau - \alpha)\delta.
\end{aligned} \tag{A3}$$

Using the tetrad form (II.10a)–(II.10b) of the Weyl equation [which clearly is preserved under the null rotation (II.9a)–(II.9c)] we find that the quantities κ, ω, σ, $\alpha - 2\bar\tau$, $\gamma - \bar\gamma$, ρ, and τ are transformed under (II.9a) – (II.9c) as follows:

$$\kappa = \kappa', \tag{A4a}$$

$$\omega = \omega' + (i/2)(\kappa'\Psi - \bar\kappa'\bar\Psi), \tag{A4b}$$

$$\sigma = \sigma' + \kappa'\bar\Psi, \tag{A4c}$$

$$\begin{aligned}\alpha - 2\bar\tau = \alpha' - 2\bar\tau' + \kappa'\Psi^2 - 2\bar\kappa'\Psi\bar\Psi\\ - 2i(2\omega'\Psi - i\bar\sigma'\bar\Psi),\end{aligned} \tag{A4d}$$

$$\begin{aligned}\gamma - \bar\gamma = \gamma' - \bar\gamma' + \bar\Psi(\alpha' - 2\bar\tau') - \Psi(\bar\alpha' - 2\tau')\\ + \Psi\bar\Psi(\Psi\kappa' - \bar\Psi\bar\kappa') - i\Psi(2\bar\Psi\omega' + i\sigma'\Psi)\\ - i\bar\Psi(2\omega'\Psi - i\bar\sigma'\bar\Psi),\end{aligned} \tag{A4e}$$

$$\rho = \rho' + \Psi\kappa', \tag{A4f}$$

$$\tau = \tau' + \rho'\bar\Psi + \sigma'\Psi + \kappa'\Psi\bar\Psi. \tag{A4g}$$

The components of the Weyl spinor Ψ_{ABCD} and of the Ricci spinor $\Phi_{AB\dot X\dot Y}$ with respect to the dyad (ξ^A, χ^A) are defined by[19]

$$\begin{aligned}
\Psi_0 &= \Psi_{ABCD}\xi^A\xi^B\xi^C\xi^D,\\
\Psi_1 &= \Psi_{ABCD}\xi^A\xi^B\xi^C\chi^D,\\
\Psi_2 &= \Psi_{ABCD}\xi^A\xi^B\chi^C\chi^D,\\
\Psi_3 &= \Psi_{ABCD}\xi^A\chi^B\chi^C\chi^D,\\
\Psi_4 &= \Psi_{ABCD}\chi^A\chi^B\chi^C\chi^D.
\end{aligned} \tag{A5}$$

$$\Phi_{00}=\Phi_{AB\dot{X}\dot{Y}}\xi^A\xi^B\bar{\chi}^{\dot{X}}\bar{\xi}^{\dot{Y}},\quad \Phi_{01}=\bar{\Phi}_{10}=\Phi_{AB\dot{X}\dot{Y}}\xi^A\xi^B\bar{\xi}^{\dot{X}}\bar{\chi}^{\dot{Y}},$$
$$\Phi_{11}=\Phi_{AB\dot{X}\dot{Y}}\xi^A\chi^B\bar{\xi}^{\dot{X}}\bar{\chi}^{\dot{Y}},\quad \Phi_{12}=\bar{\Phi}_{21}=\Phi_{AB\dot{X}\dot{Y}}\xi^A\chi^B\bar{\xi}^{\dot{X}}\bar{\xi}^{\dot{Y}}, \tag{A.6}$$
$$\Phi_{22}=\Phi_{AB\dot{X}\dot{Y}}\chi^A\chi^B\bar{\chi}^{\dot{X}}\bar{\chi}^{\dot{Y}},\quad \Phi_{02}=\bar{\Phi}_{20}=\Phi_{AB\dot{X}\dot{Y}}\xi^A\xi^B\bar{\chi}^{\dot{X}}\bar{\chi}^{\dot{Y}}.$$

Under the null rotation (II.9a)–(II.9c) the dyad components of the Weyl spinor transform as

$$\begin{aligned}
\Psi_0&=\Psi'_0,\\
\Psi_1&=\Psi'_1+\Psi\Psi'_0,\\
\Psi_2&=\Psi'_2+2\Psi\Psi'_1+\Psi^2\Psi'_0,\\
\Psi_3&=\Psi'_3+3\Psi\Psi'_2+3\Psi^2\Psi'_1+\Psi^3\Psi'_0,\\
\Psi_4&=\Psi'_4+4\Psi\Psi'_3+6\Psi^2\Psi'_2+4\Psi^3\Psi'_1+\Psi^4\Psi'_0.
\end{aligned}\tag{A7}$$

Using the Weyl equation (II.10a)–(II.10b) and the fact that the trace of the neutrino energy-momentum tensor vanishes, the Ricci and Bianchi identities used in this paper can be written in the form

$$D\rho-\bar{\delta}\kappa=\rho^2+\sigma\bar{\sigma}+(\rho+\bar{\rho})\rho-\bar{\kappa}\tau-\kappa(3\alpha+\bar{\tau}-\pi), \tag{R.1}$$

$$D\sigma-\delta\kappa=4\rho\sigma-(4\tau-\bar{\pi}+\bar{\alpha})\kappa+\Psi_0, \tag{R.2}$$

$$D\tau-\Delta\kappa=(\tau+\bar{\pi})\rho+(\bar{\tau}+\pi)\sigma+ \\ (\rho-\bar{\rho})\tau-(3\gamma+\bar{\gamma})\kappa+\Psi_1+\Phi_{01}, \tag{R.3}$$

$$D\alpha-\bar{\delta}\rho=(\bar{\rho}-\rho)\alpha+\tau\bar{\sigma}-\bar{\tau}\rho-\kappa\lambda \\ -\bar{\kappa}\gamma+2\rho\pi+\Phi_{10}, \tag{R.4}$$

$$D\tau-\delta\rho=(\alpha+\pi)\sigma-(\mu+\gamma)\kappa-(\bar{\alpha}-\bar{\pi})\rho+\Psi_1, \tag{R.5}$$

$$D\gamma-\Delta\rho=(\tau+\bar{\pi})\alpha+(\bar{\tau}+\pi)\tau \\ -(\rho+\bar{\rho})\gamma-(\gamma+\bar{\gamma})\rho+\tau\pi-\nu\kappa+\Psi_2+\Phi_{11}, \tag{R.6}$$

$$D\mu-\delta\pi=\lambda\sigma+\pi\bar{\pi}-\rho\mu-\pi(\bar{\alpha}-\tau)-\nu\kappa+\Psi_2, \tag{R.8}$$

$$\delta\rho-\bar{\delta}\sigma=\rho(\bar{\alpha}+\tau)-\sigma(3\alpha-\bar{\tau}) \\ +(\rho-\bar{\rho})\tau+(\mu-\bar{\mu})\kappa-\Psi_1+\Phi_{01}, \tag{R.11}$$

$$\delta\alpha-\bar{\delta}\tau=\rho\mu-\lambda\sigma+\alpha\bar{\alpha}+\tau\bar{\tau}-2\alpha\tau \\ +\gamma(\rho-\bar{\rho})+\rho(\mu-\bar{\mu})-\Psi_2+\Phi_{11}, \tag{R.12}$$

$$\delta\gamma-\Delta\tau=2\mu\tau-\bar{\alpha}\gamma-\sigma\nu-\rho\bar{\nu} \\ -\tau(\gamma-\bar{\gamma})+\alpha\bar{\lambda}+\Phi_{12}, \tag{R.15}$$

$$\delta\tau-\Delta\sigma=\mu\sigma+\bar{\lambda}\rho+(2\tau-\bar{\alpha})\tau-(3\gamma-\bar{\gamma})\sigma-\kappa\bar{\nu}+\Phi_{02}, \tag{R.16}$$

$$\Delta\rho-\bar{\delta}\tau=\nu\kappa-\rho\bar{\mu}–\lambda\sigma-\alpha\tau+(\gamma+\bar{\gamma})\rho-\Psi_2, \tag{R.17}$$

$$\Delta\alpha-\bar{\delta}\gamma=2\rho\nu-2\tau\lambda+(\bar{\gamma}-\bar{\mu})\alpha-\Psi_3, \tag{R.18}$$

$$\bar{\delta}\Psi_0-D\Psi_1+D\Phi_{01}=(4\alpha-\pi)\Psi_0-6\rho\Psi_1 \\ +3\kappa\Psi_2+2\sigma\Phi_{10}-2\kappa\Phi_{11}-\bar{\kappa}\Phi_{02}+2(\rho+\bar{\rho})\Phi_{01}, \tag{B1}$$

$$\begin{aligned}
&3(\bar{\delta}\Psi_1-D\Psi_2)+2(D\Phi_{11}-\delta\Phi_{10})+\bar{\delta}\Phi_{01}\\
&\quad=3\lambda\Psi_0-9\rho\Psi_2+6\kappa\Psi_3+6(\alpha-\pi)\Psi_1\\
&\qquad+2(\alpha+\pi+\bar{\tau})\Phi_{01}+2(\tau-2\bar{\alpha}+\bar{\pi})\Phi_{10}\\
&\qquad+2(2\bar{\rho}-\rho)\Phi_{11}+2\sigma\Phi_{20}-\bar{\sigma}\Phi_{02}\\
&\qquad-2\bar{\kappa}\Phi_{12}-2\kappa\Phi_{21},
\end{aligned}\tag{B3}$$

$$\begin{aligned}
&3(\Delta\Psi_1-\delta\Psi_2)+2(D\Phi_{12}-\delta\Phi_{11})+\bar{\delta}\Phi_{02}-\Delta\Phi_{01}\\
&=3\nu\Psi_0+6(\gamma-\mu)\Psi_1-9\tau\Psi_2+6\sigma\Psi_3-\bar{\nu}\Phi_{00}-2\bar{\lambda}\Phi_{10}\\
&\quad+2(\bar{\mu}-\mu-\gamma)\Phi_{01}+2(\tau+2\bar{\pi})\Phi_{11}-2(\rho+\bar{\rho})\Phi_{12}\\
&\quad+(2\alpha+2\pi-\bar{\tau})\Phi_{02}+2\sigma\Phi_{21}-2\kappa\Phi_{22}.
\end{aligned}\tag{B4}$$

APPENDIX B

All space-time metrics admitting a pure-radiation neutrino field with diverging rays (i.e., $\rho\neq 0$) or a pure-radiation neutrino field with nondiverging rays (i.e., $\rho=0$) and $\Psi_3=0$ have been found explicitly by Collinson and Morris.[11] So, considering the isometry groups of these metrics we may confirm our results of Sec. IV, Case 3 concerning these fields. For all the notations used here and for further details the reader is referred to the paper of the above authors.

1. Neutrino pure radiation fields with diverging rays

In a coordinate system $(x^1,x^2,x^3,x^4)=(u,r,x,y)$ based on the neutrino principal null congruence,[20] the space-time metric admitting these neutrino fields is written in the form

$$g_{\mu\nu}=\begin{pmatrix}-2\mu/r & 1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & -r^2/2 & 0\\ 0 & 0 & 0 & -r^2/2\end{pmatrix}, \tag{B1.1}$$

where the spin coefficient μ is an arbitrary function of u alone. The neutrino flux-vector is

$$l^\mu=Ar^{-2}\delta^\mu_2, \tag{B1.2}$$

where $A\equiv A(u)$ is an arbitrary function of u. According to the form of the function $\mu(u)$ the investigation of the Killing equations yields the following results:

(a) If $\mu\neq 1/(a-2bu)^3$, where a and b are real constants, the metric (B1.1) admits a 3-parameter group of isometries with Killing vector fields

$$n^\mu_{(2)}=(0,0,y,-x),\quad n^\mu_{(3)}=\delta^\mu_3,\quad n^\mu_{(4)}=\delta^\mu_4. \tag{B1.3}$$

By straightforward calculation we find that the Lie derivatives of l^μ with respect to these Killing vector fields vanish:

$$\mathscr{L}_{n(i)}l^\mu=0,\quad i=2,3,4. \tag{B1.4}$$

(b) If $\mu=1/(a-2bu)^3$, $b\neq 0$, the metric (B1.1) admits a 4-parameter group of isometries with Killing vector fields

$$n^\mu_{(1)}=(a/b-2u,\ 2r,\ -2x,\ -2y),\quad n^\mu_{(2)}=(0,0,y,-x), \\ n^\mu_{(3)}=\delta^\mu_3,\quad n^\mu_{(4)}=\delta^\mu_4. \tag{B1.5}$$

Now the Lie derivative of l^μ with respect to $n^\mu_{(1)}$ is nonzero in general:

$$\mathscr{L}_{n(1)}l^\mu=\left[\left(\frac{a}{b}-2u\right)\frac{1}{A}\frac{dA}{du}-6\right]l^\mu. \tag{B1.6}$$

It must be noted that if $\mu=$ constant (i.e., $b=0$) then the neutrino field reduces to a ghost field.

2. Neutrino pure-radiation fields with nondiverging rays and $\Psi_3=0$

In the same coordinate system as previously the space-time metric admitting these fields is written in the form

$$g_{\mu\nu}=\begin{pmatrix}-m(x^2+y^2)-F(u,x,y) & 1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & -\frac{1}{2} & 0\\ 0 & 0 & 0 & -\frac{1}{2}\end{pmatrix}, \tag{B2.1}$$

where $m\equiv m(u)$ is an arbitrary function of u and $F(u,x,y)$ is any function of u,x,y satisfying Laplace's equation $(\partial^2/\partial x^2 + \partial^2/\partial y^2)F = 0$. This metric represents the well known plane-fronted wave. The neutrino flux-vector is

$$l^\mu = A\delta_2^\mu, \tag{B2.2}$$

where $A \equiv A(u)$ is an arbitrary function of u.

In the case where

$$m = \text{positive constant}, \; F(u,x,y) = F(u), \tag{B2.3}$$

this metric admits a 7-parameter group of isometries with Killing vector fields

$$\begin{aligned}
&n^\mu_{(1)} = (1, \tfrac{1}{2}F(u), 0, 0), \quad n^\mu_{(2)} = \delta^\mu_2,\\
&n^\mu_{(3)} = (0, 0, -2y, 2x),\\
&n^\mu_{(4)} = (0, (2m)^{1/2}e^{-(\sqrt{2m})u}x, -2e^{-(\sqrt{2m})u}, 0),\\
&n^\mu_{(5)} = (0, -(2m)^{1/2}e^{(\sqrt{2m})u}x, -2e^{(\sqrt{2m})u}, 0),\\
&n^\mu_{(6)} = (0, (2m)^{1/2}e^{-(\sqrt{2m})}y, 0, -2e^{-(\sqrt{2m})u}),\\
&n^\mu_{(7)} = (0, -(2m)^{1/2}e^{(\sqrt{2m})u}y, 0, -2e^{(\sqrt{2m})u}).
\end{aligned} \tag{B2.4}$$

The Lie derivatives of l^μ with respect to these Killing vector fields are

$$\begin{aligned}
&\mathscr{L}_{n(i)} l^\mu = 0, \quad i = 2, 3, 4, 5, 6, 7,\\
&\mathscr{L}_{n(1)} l^\mu = \frac{1}{A}\frac{dA}{du} l^\mu.
\end{aligned} \tag{B2.5}$$

If with (B2.3) we set

$$A(u) = 1, \tag{B2.6}$$

the neutrino field reduces to the "restricted neutrino field" of Inomata and McKinley.[21] The Lie derivatives with respect to $n^\mu_{(i)}$ of their plane-wave-like solution

$$\xi^A = e^{i\epsilon u}\xi^A_0, \tag{B2.7}$$

where ξ^A_0 is a covariantly constant 2-spinor and ϵ is a real constant, are given by

$$\begin{aligned}
&\mathscr{L}_{n(1)}\xi^A = i\epsilon\xi^A, \quad \mathscr{L}_{n(3)}\xi^A = i\xi^A,\\
&\mathscr{L}_{n(i)}\xi^A = 0, \quad i = 2, 4, 5, 6, 7.
\end{aligned} \tag{B2.8}$$

Also, it must be noted that by virtue of (B2.6) the neutrino flux-vector is equal to $n^\mu_{(2)}$ and therefore constitutes a motion of the neutrino energy-momentum tensor in accordance with the results of Audretsch and Graf.[12]

[1] The main results are included in the following papers: H. Michalski and J. Wainwright, Gen. Relativ. Gravit. **6**, 289–317 (1975); J. R. Ray and E. L. Thompson, J. Math. Phys. **16**, 345–346 (1975); J. Wainwright and P. E. A. Yaremovicz, Gen. Relativ. Gravit. **7**, 345–359 (1976); B. Coll, C. R. Acad. Sci. Paris **280A**, 1773–1776 (1975); M. L. Woolley, J. Phys. A **10**, 2107–2114 (1977); C. Hoenselaers, Prog. Theor. Phys. **59**, 1518–1521 (1978); Z. A. Shteingrad, Sov. Phys. Dokl. **23**, 912–914 (1978).

[2] E. Newman and R. Penrose, J. Math. Phys. **3**, 566–578 (1962).

[3] F. A. E. Pirani, in *Lectures on General Relativity*, 1964 *Brandeis Summer Institute*, Vol. 1, edited by S. Deser and K. Ford (Prentice Hall, Englewood Cliffs, N. J., 1965).

[4] J. B. Griffiths and R. A. Newing, J. Phys. **A3**, 269–273 (1970).

[5] J. B. Griffiths and R. A. Newing, J. Phys. **A4**, 208–213 (1971).

[6] J. Wainwright, J. Math. Phys. **12**, 828–835 (1971).

[7] A Lie derivative of ξ^A with respect to the vector field n^μ satisfying the properties (i) and (ii) is given by $\mathscr{L}_n\xi^A = n^\mu\xi^A{}_{;\mu} + \frac{1}{4}(n_{\mu;\nu} - n_{\nu;\mu})\sigma^{\nu A\dot{X}}\sigma^\mu_{B\dot{X}}\xi^B$. For a review about Lie differentiation of spinors one can consult the following papers: Y. Kosmann, Ann. Math. Pura Appl. **XCI**, 317–395 (1972); P. Spindel, C. Schomblond, and M. Henneaux, Bull. Cl. Sci. Acad. R. Belg. **LXIII**, 179–186 (1977); V. Jhangianni, Found. Phys. **8**, 445–462 (1978); **8**, 593–601 (1978).

[8] K. Yano, *The Theory of Lie-Derivatives and Its Applications* (North-Holland, Amsterdam, 1955).

[9] See Ref. 3, pp. 350–351.

[10] E. Flaherty, "Hermitian and Kalherian Geometry in Relativity," in *Lecture Notes in Physics*, edited by J. Ehlers *et al.* (Springer, New York, 1976), pp. 128 – 132.

[11] C. D. Collinson and P. B. Morris, Int. J. Theor. Phys. **5**, 293–301 (1972).

[12] J. Audretsch and W. Graf, Commun. Math. Phys. **19**, 315–326 (1970).

[13] V. A. Lubimov, E. G. Novikov, V. Z. Nozik, E. F. Tretyakov, and V. S. Kosik, Phys. Lett. **B94**, 266–268 (1980).

[14] J. B. Griffiths, J. Phys. **A12**, 2429–2435 (1979).

[15] C. J. Radford and A. H. Klotz, J. Phys. **A12** 1205–1214 (1979); **12**, 1215–1221 (1979).

[16] M. Henneaux, Gen. Relativ. Gravit. **12**, 137–147 (1980).

[17] See Ref. 2, Eqs. (4.1a) and (2.12).

[18] See Ref. 2, Eqs. (4.4). Also see Ref. 10, Eqs. (VII-23), pp. 132–133.

[19] See Ref. 3, p. 348, Eqs. (4.29) and (4.30).

[20] The neutrino flux-vector is proportional to the gradient of u and r is a parameter along the null geodesics to which the neutrino flux-vector is tangent. The two coordinates x and y label the null geodesics on each hypersurface u = constant.

[21] A. Inomata and W. A. McKinley, Phys. Rev. **140**, B1467–1473 (1965).

The Melnick–Tabensky solutions have high symmetry

W. B. Bonnor

Department of Mathematics, Queen Elizabeth College, University of London, Campden Hill Road, London W.8, England

M. A. H. MacCallum

Department of Applied Mathematics, Queen Mary College, University of London, Mile End Road, London E.1, England

(Received 31 March 1981; accepted for publication 26 June 1981)

Melnick and Tabensky recently gave a class of static perfect fluid solutions of Einstein's equations in which the metric (in comoving coordinates) takes the conformastat form. In this paper we point out that all these solutions have spherical or plane or pseudospherical symmetry.

PACS numbers: 04.20.Cv, 04.20.Jb

Melnick and Tabensky[1] gave a class of static solutions of Einstein's equations for perfect fluid in which the metric takes the conformastat form

$$ds^2 = e^{2\phi}dt^2 - e^{2\psi}(dx^2 + dy^2 + dz^2), \tag{1}$$

where ϕ and ψ are functions of x, y, z. As they remark, their method requires no assumption about symmetry, and indeed their solutions, though evidently possessing axial symmetry, have no other obvious spatial symmetries. Moreover, there are special members of the class which apply to isolated bodies, i.e., have a closed boundary at zero pressure. Thus Melnick and Tabensky's solutions seem to offer the intriguing possibility of static bodies of perfect fluid which are not spherically symmetric. However, in this paper we shall show that the solutions have spherical symmetry (SS), or plane symmetry (PS), or pseudospherical symmetry (PSS).

The class may be written

$$\begin{aligned}&\phi \equiv \phi(u), \quad u = (z+m)^{-1}(x^2+y^2+z^2-q),\\ &e^{-\psi} = e^{\phi}(z+m)\,\alpha(u),\end{aligned} \tag{2}$$

where m, q are constants, and α and ϕ are related by

$$\frac{d^2\alpha}{du^2} = 2\alpha\left(\frac{d\phi}{du}\right)^2. \tag{3}$$

The pressure p and density ρ can be obtained from the expressions

$$\kappa p = \tfrac{1}{3}e^{-2\psi}(2\nabla^2\psi + (\nabla\psi)^2 + 2\nabla^2\phi + 2(\nabla\phi)^2 + 2\nabla\phi\cdot\nabla\psi), \tag{4}$$

$$\kappa\rho = -e^{-2\psi}(2\nabla^2\psi + (\nabla\psi)^2), \tag{5}$$

where κ is the gravitational constant in Einstein's equations (8π in the units of Ref. 1). Equations (4) and (5) correct (3) and (4) of Ref. 1. Both p and ρ are functions of u only, so they satisfy an equation of state of the form $p = f(\rho)$. The 4-velocity of the fluid is

$$u^i = \delta^i_4 e^{-\phi},$$

showing that it is at rest in the coordinate system of (1). As the motion is shear-free and irrotational the solutions are among those discussed by Barnes.[2]

Given any two functions of u, α and ϕ, not necessarily satisfying (3), a metric (1) subject to (2) has at least the four Killing vectors

$$\overset{(1)}{\chi}{}^i = \delta^i_1 y - \delta^i_2 x,$$

$$\begin{aligned}\overset{(2)}{\chi}{}^i &= \delta^i_1(y^2+z^2-x^2+2mz+q) - 2\delta^i_2 xy\\ &\quad - 2\delta^i_3 x(z+m),\end{aligned}$$

$$\begin{aligned}\overset{(3)}{\chi}{}^i &= -2\delta^i_1 xy + \delta^i_2(x^2+z^2-y^2+2mz+q)\\ &\quad - 2\delta^i_3 y(z+m),\end{aligned}$$

$$\overset{(4)}{\chi}{}^i = \delta^i_4,$$

where $(x^1, x^2, x^3, x^4) \equiv (x, y, z, t)$. Since the $\overset{(\mu)}{\chi}{}^i$ $(\mu = 1, 2, 3)$ span a two-plane at each point, the space–time must be locally SS or PS or PSS. One can distinguish the different cases by considering the commutators of the $\overset{(\mu)}{\chi}{}^i$, but in order to facilitate identification of the solutions we shall exhibit explicitly a suitable coordinate transformation of (1).

First write

$$x = \sigma\cos\beta, \quad y = \sigma\sin\beta,$$

and define η by

$$\eta = \sigma + \sigma^{-1}(z^2 + 2mz + q).$$

Introducing now the differential du from (2) we can bring (1), subject to (2), to the form

$$\begin{aligned}ds^2 &= e^{2\phi}dt^2 - e^{-2\phi}\alpha^{-2}X^{-1}\\ &\quad\times[du^2 + X^2Y^{-2}(d\eta^2 + Yd\beta^2)],\end{aligned} \tag{6}$$

where

$$X{:} = (u+2m)^2 + 4(q-m^2),$$
$$Y{:} = \eta^2 + 4(m^2-q).$$

In what follows we shall have to treat separately the three cases

$$m^2 > q, \quad m^2 = q, \quad m^2 < q,$$

and to save writing we shall denote these by (A), (B), and (C), respectively.

Define θ by

$$\eta = \begin{cases} 2(m^2-q)^{1/2}\cot\theta & \text{(A)},\\ \theta^{-1} & \text{(B)},\\ 2(q-m^2)^{1/2}\coth\theta & \text{(C)},\end{cases}$$

which reduces (6) to

$$ds^2 = e^{2\phi}dt^2 - e^{-2\phi}\alpha^{-2}X^{-1} \times [du^2 + K^{-1}X^2(d\theta^2 + f(\theta)d\beta^2)], \qquad (7)$$

where

$$K = \begin{cases} 4|m^2 - q| & \text{(A), (C),} \\ 1 & \text{(B),} \end{cases}$$

and

$$f = \begin{cases} \sin^2\theta & \text{(A),} \\ \theta^2 & \text{(B),} \\ \sinh^2\theta & \text{(C).} \end{cases} \qquad (8)$$

It is obvious from (7) and (8) that the metric has SS, PS, and PSS in cases (A), (B), and (C), respectively.

We can put (7) into a spatially isotropic form if we introduce a different radial coordinate r by

$$\frac{dr}{r} = -\frac{(K)^{1/2}\,du}{(u+2m)^2 + 4(q - m^2)}$$

which gives

$$u + 2m = \begin{cases} (K)^{1/2}\coth(\ln r) & \text{(A),} \\ (\ln r)^{-1} & \text{(B),} \\ (K)^{1/2}\cot(\ln r) & \text{(C).} \end{cases}$$

By this means we arrive finally at the metric

$$ds^2 = e^{2\phi}dt^2 - e^{-2\phi}\alpha^{-2}r^{-2}g^{-2}[dr^2 + r^2(d\theta^2 + f\,d\beta^2)], \qquad (9)$$

where

$$g = \begin{cases} \sinh(\ln r) & \text{(A),} \\ \ln r & \text{(B),} \\ \sin(\ln r) & \text{(C).} \end{cases}$$

Equation (3), which is the only restriction imposed by the field equations on ϕ and α, becomes, in the new coordinates,

$$\alpha'' - 2\alpha\phi'^2 + \alpha'(2g'g^{-1} + r^{-1}) = 0, \qquad (10)$$

where $'$ means d/dr.

The class of solutions of Melnick and Tabensky is therefore equivalent to a known class of static perfect fluid spacetimes admitting a four-parameter group of motions, and included in Barnes.[2] The PS and SS cases in this class are discussed in Secs. 13.6 and 14.1 of Ref. 3. [Note that Eq. (14.15) of Ref. 3, which is equivalent to (10), should say

$$L\frac{d^2G}{dx^2} = 2G\frac{d^2L}{dx^2},$$

where for (9), $x = r^2$, $L = grae^{\phi}$ and $G = Le^{\phi}$.] The vacuum solution given by Melnick and Tabensky[1] is similarly equivalent to the well-known plane symmetric vacuum solution of Taub [Ref. 3, Eq. (13.30)].

[1] J. Melnick and R. Tabensky, J. Math. Phys. **16**, 958 (1975).

[2] A. Barnes, Gen. Relativ. Gravit. **4**, 105 (1973).

[3] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact solutions of Einstein's field equations* (Deutscher Verlag d. W. Berlin and Cambridge U.P., Cambridge, 1980).

Manifestly covariant equations of motion for a particle in an external field

P. H. Lim
Departmento de Fisica, Universidade Federal da Paraiba, C.C.E.N., 58.000-Joao Pessoa-PB, Brazil

(Received 16 September 1980; accepted for publication 26 June 1981)

A relativistic variational principle for a particle in an external field is developed both in flat spacetime and in curved spacetime. In flat spacetime Kalman's equations follow from the variational principle, and their relationship to four-dimensional Euler–Lagrange equations is clarified. It is shown that Kalman's equations are uniquely defined and that they may be recast into a generalized Hamiltonian formalism. The equations of motion arising from the curved-spacetime variational principle are shown to be uniquely defined.

PACS numbers: 04.20.Fy, 03.50. – z

I. INTRODUCTION

In 1961, Kalman[1] presented a set of manifestly covariant equations of motion for a particle in an external field in flat spacetime. The equations were derived from a relativistic generalization of Hamilton's variational principle. Although the procedure parallels the Lagrangian formalism, the resulting equations of motion do not have the form of the covariant Euler–Lagrange equations in which the usual time parameter is replaced by particle proper time. Kalman also showed that the equations of motion cannot be recast into a covariant form exactly analogous to Hamilton's equations. In this work, the above properties of Kalman's equations are examined further and the variational principle is extended to curved spacetime. The relationship between Kalman's equations and four-dimensional Euler–Lagrange equations is clarified, and it is shown that the former may be recast into a generalized Hamiltonian form using Dirac's constraint formalism.[2,3]

It is to be noted that Kalman's equations are of limited applicability; they describe a particle interacting with an external field (i.e., a field independent of the particle position) and hence cannot describe a system of interacting particles. An extension of the formalism to include interacting particles would have to conform with the no-interaction theorem of Currie, Jordan, and Sudarshan[4–8]: A relativistically invariant theory of interacting particles cannot simultaneously be a Hamiltonian theory and allow an invariant description of particle world lines. Although the problem of extending the formalism is not considered here, it is anticipated that Kalman's equations may form a suitable starting point for the construction of a relativistic theory of interacting particles.[9]

In Sec. II, the action integral is parametrized in terms of an arbitrary function. One choice of parameter (particle proper time) leads of Kalman's equations whilst another choice leads to Euler–Lagrange equations. These two sets of equations are, of course, equivalent and this is demonstrated explicitly. It is shown, however, that the two choices of parameter cannot be held simultaneously; hence the Euler–Lagrange equations cannot be parametrized in particle proper time. In Sec. III, it is shown that the latter equations (parametrized in particle proper time) are not even well defined: The source of ambiguity is the relativistic constraint on the particle 4-velocity components. Kalman's equations are shown to be unambiguously defined despite this constraint. In Sec. IV, the constraint is taken into account in the construction of a generalized Hamiltonian formalism. Finally, the variational principle is developed for curved spacetime in Sec. V, and it is shown that the resulting equations of motion are uniquely defined.

II. VARIATIONAL PRINCIPLE IN FLAT SPACETIME

Consider a particle interacting with an external field in flat spacetime. In Ref. 1, the particle trajectory is parametrized in particle proper time. In the following calculations, an arbitrary parametrization is considered. The scalar action functional for the particle-field interaction is taken to have a form analogous to that of nonrelativistic action functionals:[10]

$$S = \int_a^b f(x,\dot{x})\, d\lambda\ , \tag{1}$$

where the integral is over the trajectory $x(\lambda)$. In (1), a and b are arbitrary points on the trajectory and $\dot{x} = dx/d\lambda$. The first variation in S is obtained[11] by comparing (1) with a similar integral over a curve $x'(\lambda')$ lying in the vicinity of the trajectory $x(\lambda)$. Thus,

$$S + \delta S = \int_a^b f(x',\dot{x}')\, d\lambda' , \tag{2}$$

where

$$\begin{aligned} x' &= x + \delta x \\ \lambda' &= \lambda + \delta\lambda ; \end{aligned} \tag{3}$$

the variations δx and $\delta\lambda$ are infinitesimal and the δx are arbitrary but vanish at the end points of integration.[12] From (3), the corresponding variation in $\dot{x}^k$ satisfies

$$\delta\dot{x} \equiv \dot{x}'^k - \dot{x}^k = \frac{d}{d\lambda}(\delta x^k) - \dot{x}^k \frac{d}{d\lambda}(\delta\lambda). \tag{4}$$

Equations (1)–(4) yield the first variation in the action[13]

$$\begin{aligned} \delta S &= \int (\delta f)\, d\lambda + \int f \delta(d\lambda) \\ &= \int \left(\frac{\partial f}{\partial x^k}\delta x^k + \frac{\partial f}{\partial \dot{x}^k}\frac{d}{d\lambda}(\delta x^k) \right) d\lambda \\ &\quad + \int \left(f - \dot{x}^k \frac{\partial f}{\partial \dot{x}^k} \right) d(\delta\lambda). \end{aligned} \tag{5}$$

The required equations of motion are obtained from (5) by setting δS equal to zero. Evidently, the form of the equations depends on the choice of parameter λ. Below, two possible choices are considered. The first is to equate λ with particle proper time, and this leads (as in Ref. 1) to Kalman's equations of motion. Alternatively, λ may be defined in such a way that Euler–Lagrange equations follow from (5).

Consider the choice that λ is the particle proper time, i.e., $d\lambda = ds$, where the line element takes the Minkowski form[14]

$$ds^2 = \eta_{lm}\, dx^l\, dx^m \tag{6}$$

in Lorentzian coordinates. Then, the action integral of (1) takes the form

$$S = \int f(x,U)\, ds, \tag{7}$$

where $f(x,U)$ is a scalar function and U is the particle 4-velocity. From (6),

$$\frac{d}{ds}(\delta s) = U_k\, \frac{d}{ds}(\delta x^k); \tag{8}$$

hence (5), after a partial integration, yields

$$\delta S = \int \left\{ \frac{\partial f}{\partial x^k} - \frac{d}{ds}\left[\frac{\partial f}{\partial U^k} + U_k\left(f - U^l\, \frac{\partial f}{\partial U^l}\right)\right]\right\} \delta x^k\, ds + \left[\frac{\partial f}{\partial U^k} + U_k\left(f - U^l\, \frac{\partial f}{\partial U^l}\right)\right]\delta x^k \Big|_a^b . \tag{9}$$

Setting δS equal to zero then yields Kalman's manifestly covariant equations[15] of Ref. 1:

$$\frac{\partial f}{\partial x^k} = \frac{d}{ds}\left[\frac{\partial f}{\partial U^k} + U_k\left(f - U^l\, \frac{\partial f}{\partial U^l}\right)\right]. \tag{10}$$

In the Appendix, it is shown that the 4-vector in square brackets on the right side of (10) is minus the covariant 4-momentum, and that (10) reduce to Lagrange's equations in a given coordinate system.

In the second choice of parameter, λ is the analog of the time parameter in nonrelativistic action principles: The same λ parametrizes both the trajectory and its neighbor, hence $\lambda = \lambda'$ and $\delta\lambda$ in (3) vanishes. Then, (1) takes the form[16]

$$S = \int l(x,\dot{x})\, d\lambda, \tag{11}$$

and (5) (with $\delta S = 0$) yields Euler–Lagrange equations in the parameter λ:

$$\frac{\partial l}{\partial x^k} = \frac{d}{d\lambda}\left(\frac{\partial l}{\partial \dot{x}^k}\right). \tag{12}$$

It is sometimes claimed[17] that λ in (12) may be chosen to be particle proper time. In that case, (12) would reduce [after equating l and f in (11) and (7)] to the form of (10) without the term $U_k\,(f - U^l\ \partial f/\partial U^l)$, i.e., to Euler–Lagrange equations in the proper time. This claim is erroneous, however, since the condition $\delta s = 0$ is incompatible with (8).

Equations (10) and (12), arising for the same action, must be equivalent. This may be shown explicitly by relating the integrands of (11) and (7):

$$l(x,\dot{x}) = f(x,U)\ \frac{ds}{d\lambda}, \tag{13}$$

where, from (6),

$$\frac{ds}{d\lambda} = (\eta_{lm}\dot{x}^l\dot{x}^m)^{1/2}. \tag{14}$$

Equations (13) and (14) imply that the differentiated expression on the right side of (12) is

$$\frac{\partial l}{\partial \dot{x}^k} = \frac{\partial f}{\partial U^k} + U_k\left(f - U^l\, \frac{\partial f}{\partial U^l}\right). \tag{15}$$

Finally, (13)–(15) reduce (12) to the form of (10).

In conclusion, Kalman's equations (10) are equivalent to (12), but the parameter in the latter cannot be the particle proper time. Indeed, λ in (12) may be chosen to be nonscalar, but then (12) would not be manifestly covariant. These results may prove useful in the construction of a theory of interacting particles which aims to avoid the world line condition in the above-mentioned no-interaction theorem.

III. UNIQUENESS OF KALMAN'S EQUATIONS

The action integral (7) yields Kalman's equations (10), not the Euler–Lagrange equations in particle proper time [which, as noted above, take the form of (10) without the term $U_k(f - U^l\partial f/\partial U^l)$. The latter equations could be rejected solely from inspection of their form: From (6), the components of the 4-velocity are related to one another by the constraint

$$\eta_{lm}\, U^l U^m = 1; \tag{16}$$

hence the partial derivations $(\partial f/\partial U^k)$ are not well defined.

At first sight, it appears that (10) suffer from this defect since they also contain partial derivatives of the form $(\partial f/\partial U^k)$. The constraint (16) could generate ambiguities in (10) in two distinct ways: First, given $f(x,U)$, the derivatives $(\partial f/\partial U^k)$ are ambiguous because one may choose arbitrarily which of the components of U is the dependent variable. Second, (16) allows some arbitrariness in the actual functional dependence of f on U. In the discussion below, these possible sources o^f ambiguity of Eqs. (10) are examined in turn, and it is shown that Kalman's equations are uniquely defined despite the constraint (16).[18]

From (16), one may express any particular 4-velocity component U^p in terms of the other three components (jointly labelled U^q):

$$U^p = U^p(U^q), \quad q \neq p. \tag{17}$$

The function $f(x,U)$ may then be written as a function independent of U^p:

$$f(x,U) = f(x,U^p,U^q) \equiv f'(x,U^q). \tag{18}$$

(The prime indicates a differing dependence on U^p, U^q.) Thus, the constraint (16) induces the identity transformation $f \to f'$ of (18). Under this transformation, the partial derivations $(\partial f'/\partial U^k)$ satisfy

$$\left(\frac{\partial f'}{\partial U^p}\right) = 0\ ,$$

$$\left(\frac{\partial f'}{\partial U^q}\right) = \left(\frac{\partial f}{\partial U^q}\right) - \frac{U_q}{U_p}\left(\frac{\partial f}{\partial U^p}\right), \tag{19}$$

where (16) is used to evaluate $(\partial U^p/\partial U^q)$. Equations (18) and

(19) then yield the invariance of the 4-momentum in (10) under the transformation $f \to f'$:

$$\frac{\partial f}{\partial U^k} + U_k\Big(f - U^l \frac{\partial f}{\partial U^l}\Big) = \frac{\partial f'}{\partial U^k} + U_k\Big(f' - U^l \frac{\partial f'}{\partial U^l}\Big). \tag{20}$$

Thus, (10) are invariant under the transformation of (18), even though the individual partial derivatives $(\partial f / \partial U^k)$ [which transform according to (19)] are not uniquely defined.

The remaining possible source of ambiguity in (10) is the freedom in functional dependence of f on U allowed by another identity transformation $f \to \tilde{f}$, where

$$f(x,U) = \tilde{f}(x,U,\xi),$$
$$\xi = \eta_{lm} U^l U^m = 1. \tag{21}$$

From (21), the derivatives $(\partial \tilde{f}/\partial U^k)$ satisfy

$$\Big(\frac{\partial \tilde{f}}{\partial U^k}\Big) = \Big(\frac{\partial f}{\partial U^k}\Big) + 2U_k\Big(\frac{d\tilde{f}}{d\xi}\Big). \tag{22}$$

Equations (16), (21), and (22) show that the 4-momentum in (10) is invariant under the transformation $f \to \tilde{f}$. Thus, (10) themselves are invariant under this transformation.

The two possible sources of ambiguity in (10) are represented by the identity transformations (18) and (21). Kalman's equations (10) are uniquely defined since they are invariant under each of these transformations. The Euler–Lagrange equations in proper time are ambiguously defined since, from (19) and (22), they are not invariant under either of the transformations. This ambiguity reflects the impossibility of equating the parameter λ in (12) with particle proper time.

Although the Euler–Lagrange equations in particle proper time are incorrect, they can be used in certain cases[17] to yield equations of motion identical with Kalman's equations. Consider, for example, a free particle of mass m. Then,

$$f = -m \tag{23}$$

in (7) and Eqs. (10) yield the constancy of 4-momentum mU. This result may also be obtained from the Euler–Lagrange equations in proper time after an appropriate identity transformation of the type (21). In case of a particle of charge e in an electromagnetic field, f in (7) takes the form

$$f = -m - eA_k(x)U^k. \tag{24}$$

Since the particle-field interaction term in f does not contribute to $U_k(f - U^l \partial f / \partial U^l)$ in (10), it follows that the Euler–Lagrange equations in proper time may, after a suitable transformation of the type (21), yield the correct equations of motion.

IV. GENERALIZED HAMILTONIAN FORMALISM

The particle-field 4-momentum P arising from the action integral (7) and appearing on the right-hand side of (10) satisfies

$$P_k = -\Big[\frac{\partial f}{\partial U^k} + U_k\Big(f - U^l \frac{\partial f}{\partial U^l}\Big)\Big]. \tag{25}$$

P is therefore a function of the coordinates x and the 4-velocity U:

$$P_k = P_k(x,U). \tag{26}$$

The components P_k are not independent of each other, and this may be shown as follows. First, suppose that the relations (26) are invertible for U, i.e., the components U_k may be expressed

$$U_k = U_k(x,P). \tag{27}$$

The constraint (16) then yields the scalar equation

$$\phi(x,P) = 0, \tag{28}$$

hence the components of P are interdependent. If, however, the components of U cannot be expressed as in (27), then the functions P_k in (26) are linearly dependent and this implies[3] that one or more scalar equations of the form (28) hold.

In Ref. 1, Kalman showed that one cannot consistently recast Eqs. (10) into a covariant form exactly analogous to Hamilton's equations, i.e., the equations

$$\frac{\partial H}{\partial x^m} = -\frac{dP_m}{ds} = \frac{\partial f ?}{\partial x^m},$$
$$\frac{\partial H}{\partial P_m} = U^m,$$
$$H \equiv H(x,P), \tag{29}$$

are not self-consistent. This result is not surprising since the components of P are interdependent and hence the second of (29) are ambiguously defined.

If the requirement of an exact analog is dropped, (10) can be recast into a generalized Hamiltonian form[2,3] which takes into account the constraints on P. As noted above, these constraints are expressed by one or more scalar equations of the form (28). For simplicity, it is taken here that there is only one such constraining equation (the generalization to several equations is trivial). From (10), (25), and (16), an infinitesimal variation in $f(x,U)$ satisfies

$$\delta f = -\frac{dP_m}{ds}\delta x^m - P_m \delta U^m. \tag{30}$$

Equations (30) and (28), with a suitable scalar multiplier μ, then yield the manifestly convariant generalized Hamiltonian equations

$$\frac{\partial G}{\partial x^m} = -\frac{dP_m}{ds} + \mu\frac{\partial \phi}{\partial x^m},$$
$$\frac{\partial G}{\partial P_m} = U^m + \mu\frac{\partial \phi}{\partial P_m},$$
$$G \equiv f + P_k U^k = 0. \tag{31}$$

[The vanishing of G in the last of (31) follows from (25) and (16).] These equations may be shown to be well defined, despite the constraint (28) on components of P, by an argument similar to that in Sec. III above.

Equations (31) are designated "generalized Hamiltonian equations" for the following reasons: First, they do not have a form exactly analogous to Hamilton's equations; indeed, the latter, i.e., Eqs. (29), are not self-consistent. Second, G in (31) should not be confused with the Hamiltonian H; the function G is a scalar while H is given by the time component of the 4-momentum P of Eqs. (25) [see Eqs. (A10)].

To illustrate the formalism, consider the following simple examples. First, the function f for a charged particle in an electromagnetic field satisfies (24), and (25) yields

$$P_k = m\, U_k + eA_k.$$

Since this equation is invertible for U, there is only one constraint equation of the form of (28):

$$\phi(x,P) = P^k P_k + e^2 A^k A_k - 2eP^k A_k - m^2 = 0.$$

The second of (31) then shows that $\mu = (2m)^{-1}$ and this result in the first of (31) yields the equations of motion (a comma below denotes partial derivatives)

$$\frac{dP_k}{ds} = eA^l{}_{,k} U_l.$$

[The latter, of course, also follow directly from (10), (24) and (25).] Consider now the case of a scalar interaction

$$f = \psi(x). \tag{32}$$

Equations (25) and (10) then yield

$$P_k = -\psi U_k,$$

$$\psi_{,k} = \frac{d}{ds}(\psi U_k),$$

respectively. The latter equations of motion may also be obtained from (31), for which

$$\phi(x,P) = \psi^2 - P^k P_k = 0,$$

and $\mu = (2\psi)^{-1}$.

V. VARIATIONAL PRINCIPLE IN CURVED SPACETIME

In this section, the variational principle is developed in a covariant form for the action integral (7) in curved spacetime. Suppose the particle interacts with a single tensorial field $B(x)$ (the generalization of the following analysis to interaction with several fields is trivial). Then, the action integral takes the form

$$S = \int f(B(x), U)\, ds, \tag{33}$$

where the line element satisfies

$$ds^2 = g_{lm}\, dx^l\, dx^m, \tag{34}$$

and the g_{lm} are components of the symmetric metric tensor.

As in Sec. II, the coordinates x of the trajectory are related to those of a neighboring curve by the first of (3). Note, however, that the differential expression

$$dx' = dx + d\delta x \tag{35}$$

shows that $d\delta x$ is not a vector, since the vectors dx and dx' are located at different points in curved spacetime. One may define δdx to be the vector difference between dx' at $x + \delta x$ and the vector $dx + \delta_{\parallel}\, dx$, which is the parallel transport of dx from x to $x + \delta x$:

$$\delta\, dx \equiv dx' - (dx + \delta_{\parallel}\, dx). \tag{36}$$

Equations (35) and (36) yield

$$\delta\, dx^m = d\delta x^m + \Gamma^m_{ln}\, dx^l\, \delta x^n, \tag{37}$$

where the Γ^m_{ln} are the affine connection coefficients. The notation of (36) may be extended to any tensor Q; denoting quantities measured at $x + \delta x$ with a prime, the variation

$$\delta Q \equiv Q' - (Q + \delta_{\parallel} Q) \tag{38}$$

has the same tensorial character as does Q itself. (In the case that Q is a scalar, δQ reduces to the flat spacetime variation.) This notation is now employed in the development of the general-relativistic variational principle.

From (33), the first variation in the action is

$$\delta S = \int (\delta f)\, ds + \int f \delta\, ds. \tag{39}$$

Consider the form of the integrands in (39). The first of these satisfies[19]

$$\delta f = \frac{\partial f}{\partial B_{(j)}}(B'_{(j)} - B_{(j)}) + \frac{\partial f}{\partial U^m}(U'^m - U^m). \tag{40}$$

Although δf is a scalar, the right side of (40) is not manifestly scalar since each of the contributions in parenthesis is the difference between a quantity measured on the neighboring curve and its counterpart measured on the trajectory. Clearly, it is necessary to recast (40) into a manifestly scalar form in order to obtain manifestly covariant equations of motion from (39). To accomplish this, note that since f is scalar, its parallel transport from x to $x + \delta x$ vanishes:

$$\delta_{\parallel} f = \frac{\partial f}{\partial B_{(j)}}\, \delta_{\parallel} B_{(j)} + \frac{\partial f}{\partial U^m}\, \delta_{\parallel} U^m = 0. \tag{41}$$

Equations (40) and (41) yield the manifestly scalar expression

$$\delta f = \frac{\partial f}{\partial B_{(j)}} \delta B_{(j)} + \frac{\partial f}{\partial U_m} \delta U^m. \tag{42}$$

From (38), the term $\delta B_{(j)}$ is related to the covariant derivatives of the field $B_{(j)}$:

$$\delta B_{(j)} = B_{(j);m}\, \delta x^m. \tag{43}$$

The term δU^m in (42) is, from (38) and (37),

$$\delta U^m = \frac{d}{ds}\delta x^m - U^m \frac{d\delta s}{ds} + \Gamma^m_{ln} U^l \delta x^n. \tag{44}$$

Equations (42)–(44) express the first integrand of (39) in a manifestly covariant form. In the remaining integrand, the term $\delta\, ds$ is reduced by (34) and (38) to[20]

$$\delta ds = d\delta s = \tfrac{1}{2} g_{lm,k}\, U^l U^m \delta x^k + U_k \frac{d\delta x^k}{ds}. \tag{45}$$

Equations (39) with (42)–(45) reduce, after a partial integration, to

$$\delta S = \int ds \left(\frac{\partial f}{\partial B_{(j)}} B_{(j);k} + P_{k;l} U^l \right) \delta x^k - P_k \delta x^k \Big|_a^b ,$$

$$P_k = -\left[\frac{\partial f}{\partial U^k} + U_k \left(f - U^l \frac{\partial f}{\partial U^l} \right) \right]. \tag{46}$$

Thus, the particle-field momentum components P_k are given by the same expressions in flat spacetime [Eqs. (25)] and in the presence of a gravitational field. The condition that δS in (46) vanishes for the trajectory yields the manifestly covariant equations of motion

$$\frac{\partial f}{\partial B_{(j)}} B_{(j);k} + P_{k;l} U^l = 0. \tag{47}$$

Equations (47), with the last of (46) for P_k, are the generalization of Kalman's equations (10) to curved spacetime for a particle-tensor field interaction. In the case that f depends on several tensor fields $B^s(x)$ the first term in (47) is replaced by a summation over s of the corresponding contributions from each field:

$$\sum_s \frac{\partial f}{\partial B^s_{(j)}} B^s_{(j);k} + P_{k;l} U^l = 0. \tag{48}$$

In Sec. III above, it is shown that in flat spacetime, Kalman's equations are well defined despite the constraint (16) on the 4-velocity components. This result is now extended to curved spacetime, for which the constraint takes the form

$$g_{lm} U^l U^m = 1. \tag{49}$$

As in Sec. III, the constraint may possibly lead to ambiguities in the equations of motion in two distinct ways. The first of these is represented by the identity transformation $f \rightarrow f'$ of (18). Under this transformation, (19) and (20) remain valid and hence (48) are invariant. The remaining possible source of ambiguity is represented by the identity transformation $f \rightarrow \tilde{f}$ where (21) is replaced by

$$f(B^s(x), U) = \tilde{f}(B^s(x), U, \xi),$$
$$\xi = g_{lm} U^l U^m = 1. \tag{50}$$

Equation (22) remains valid, hence P_k in the second of (46) is invariant under the transformation $f \rightarrow \tilde{f}$. The equations of motion (48) then remain invariant since the covariant derivative of the metric tensor is identically zero.

It follows that both the 4-momentum P and the equations of motion (48) are well defined despite the constraint (49).

VI. CONCLUSION

The action integral (7), for a particle in an external field, leads to Kalman's manifestly covariant equations of motion (10) in flat spacetime. It is shown in this work that (10) have the following properties:

(i) They are equivalent to Euler–Lagrange equations, (12), in which the parameter λ cannot be particle proper time.

(ii) They are well defined despite the relativistic constraint (16) on the particle 4-velocity.

(iii) They can be recast into a generalized Hamiltonian formalism.

The variational principle for the action integral (7) is readily generalized to curved spactime, and it leads to equations of motion (48) for particle-tensor field interactions.

In future work, the formalism could be extended in two directions. First, as noted in Sec. II above, Kalman's equations recast into the form of (12) with a nonscalar λ may form a starting point for the construction of a relativistic theory of interacting particles. Finally, it would be interesting to construct a manifestly covariant Hamilton–Jacobi formalism based on Eqs. (31) above, and to study the transition from classical mechanics to quantum mechanics. In this way, it may be possible to deduce a covariant correspondence principle for the transition.

ACKNOWLEDGMENTS

The author thanks A. Jabs, J. Jayaraman, and N. L. Teixeira for their discussions and suggestions.

APPENDIX: PARTICLE-FIELD 4-MOMENTUM

The covariant component P_k of the particle-field 4-momentum corresponding to the action integral (7) is defined[21]

$$P_k = -\frac{\partial S}{\partial x^k}, \tag{A1}$$

where right side of (A1) is evaluated on the particle trajectory. Thus,[1] (A1) and (9) yield

$$P_k = -\left[\frac{\partial f}{\partial U^k} + U_k\left(f - U^l \frac{\partial f}{\partial U^l}\right)\right]. \tag{A2}$$

To demonstrate that the P_k in (A2) constitute the 4-momentum, consider the action integral (7) expressed in a specific Lorentzian coordinate system:

$$S = \int L\, dt, \tag{A3}$$

where L is the Lagrangian. A comparison of the integrands of (7) and (A3) yields

$$f = U^0 L; \tag{A4}$$

hence (A2) takes the form

$$P_k = -U^0\left(\frac{\partial L}{\partial U^k} - U_k U^l \frac{\partial L}{\partial U^l}\right) - L\,\delta^0_k. \tag{A5}$$

[Note that Sec. III shows that the right side of (A5) is well defined despite the constraint (16).]

To evaluate (A5), it is convenient to define the 3-velocity components

$$V^\alpha = \frac{dx^\alpha}{dt} \tag{A6}$$

in the specific coordinate system. From (6), the components of U are related to the 3-velocity components

$$U^0 = (1 + \eta_{\alpha\beta} V^\alpha V^\beta)^{-1/2},$$
$$U^\alpha = V^\alpha U^0 \tag{A7}$$

Thus, the partial derivatives $(\partial U^k / \partial V^\beta)$ satisfy

$$\frac{\partial U^0}{\partial V^\beta} = U^\beta U^{0^2},$$
$$\frac{\partial U^\alpha}{\partial V^\beta} = U^0(\delta^\alpha_\beta + U^\alpha U^\beta). \tag{A8}$$

Equations (A8) and (6) yield the partial derivations of L with respect to the 3-velocity components

$$\frac{\partial L}{\partial V^\beta} = U^0\left(\frac{\partial L}{\partial U^\beta} - U_\beta U^l \frac{\partial L}{\partial U^l}\right),$$
$$V^\beta \frac{\partial L}{\partial V^\beta} = -U^0\left(\frac{\partial L}{\partial U^0} - U_0 U^l \frac{\partial L}{\partial U^l}\right). \tag{A9}$$

Finally, (A9) reduce (A5) to the form

$$P_0 = V^\beta \frac{\partial L}{\partial V^\beta} - L = H,$$
$$P_\alpha = -\frac{\partial L}{\partial V^\alpha} = -p_\alpha, \tag{A10}$$

where H and p_α are the Hamiltonian and α component of the conjugate 3-momentum, respectivley. Thus, the P_k of (A5) are the covariant components of 4-momentum.

In the specific Lorentzian coordinate system employed above, (A4) reduces Kalman's equations (10) to the form

$$\frac{\partial L}{\partial x^k} = -\frac{dP_k}{dt}. \tag{A11}$$

Equations (A10) show that (A11) are Lagrange's equations of motion.

[1] G. Kalman, Phys. Rev. **123**, 384 (1961).

[2] P. A. M. Dirac, Proc. R. Soc. London Ser. A **246**, 326 (1958).

[3] P. A. M. Dirac, Canad. J. Math. **2**, 129 (1950).

[4] D. G. Currie, T. F. Jordan, and E.C.G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963).

[5] J. T. Cannon and T. F. Jordan, J. Math. Phys. **5**, 299 (1964).

[6] H. Leutwyler, Nuovo Cimento **37**, 556 (1965).

[7] E. H. Kerner, J. Math. Phys. **9**, 222 (1968).

[8] E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).

[9] For a recent discussion of the difficulties encountered in attempts to formulate a relativistic theory of interacting particles, see F. Rohrlich, Ann. Phys. **117**, 292 (1979).

[10] In this work, x denotes the set of coordinates (x^0,x^1,x^2,x^3). A similar notation is used for other four-dimensional quantities. Latin and Greek indices run from 0 to 3 and from 1 to 3, respectively.

[11] See, for example, A. Mercier, *Analytical and Canonical Formalism in Physics* (Dover, New York, 1963).

[12] Note that $\delta\lambda$ in (3) is not set to zero at the end points of integration because λ is an arbitrary parameter. If, for example, λ and λ' were chosen to be proper time measured along the trajectory and neighboring (timelike) curve, respectively, then setting $\delta\lambda = 0$ at a would imply that $\delta\lambda \neq 0$ at b.

[13] For simplicity, the limits of integration are henceforth omitted from the calculations.

[14] The signature is $+\ -\ -\ -$ and $c = 1$.

[15] Note that terms with U_k in (10) have a different sign from their counterparts in Ref. 1, Eq. (12). This arises from the use of different signatures [compare Ref. 14 above with Eq. (7) of Ref. 1].

[16] Since both integrals in (7) and (11) refer to the same action S, the integrands have been denoted by different symbols.

[17] See, for example, H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1965).

[18] Note that (10) also result from the action integral (7) when (16) is inserted as an explicit constraint; see A. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (MacMillian, New York, 1964).

[19] The label (j) denotes the set of indices attached to components of B.

[20] Note that (37) and the usual expression for Γ^m_{ln} in terms of the metric tensor components recasts the right side of (45) into the manifestly scalar form $U_k \delta dx^k/ds$.

[21] L. D. Landau and E. M. Lifshitz, *Mechanics* (Addison-Wesley, Reading, Mass., 1960).

Conditional symmetries in parametrized field theories

Karel Kuchař
Department of Physics, University of Utah, Salt Lake City, Utah 84112

(Received 22 December 1981; accepted for publication 19 March 1982)

In parametrized field theories, spacelike hypersurfaces and fields which they carry are evolved by a Hamiltonian which is a linear combination of the super-Hamiltonian and supermomentum constraints. We say that a dynamical variable K generates a conditional symmetry of the Hamiltonian when it is linear both in the hypersurface and the field momenta and its Poisson bracket with the Hamiltonian vanishes by virtue of the constraints. Generators are classified by their dependence on the momenta: P-restricted generators depend only on the hypersurface momenta, π-restricted generators depend only on the field momenta, while mixed generators depend on both kinds of momenta. Conditional symmetries in a parametrized Hamiltonian theory are then linked either with ordinary symmetries (isometries, conformal motions, or homothetic motions) of the spacetime background, or with internal symmetries of the fields. In particular, we prove that a generic field with nonderivative gravitational coupling and a quadratic energy density has a P-restricted conditional symmetry if and only if the spacetime background has a Killing vector, while a field with a trace-free energy–momentum tensor has a P-restricted conditional symmetry if and only if the background has a conformal Killing vector. An algorithm allowing us to enumerate all possible mixed conditional symmetries in a given parametrized field theory is explained on an example of the Klein–Gordon field. These results complement our previous proof that canonical geometrodynamics does not possess any conditional symmetry.

PACS numbers: 04.20.Fy, 11.10.Ef

1. MOTIVATION

In general relativity, one often studies fields which evolve on a given spacetime background. The fields are described by the canonical data $\phi^A(x)$, $\pi_A(x)$, which are defined on a spacelike hypersurface $X^\alpha(x)$. As the hypersurface is deformed in the embedding spacetime, the dynamics of the data is generated by a field Hamiltonian. By a process known as parametrization, it is possible to treat the hypersurface variables $X^\alpha(x)$ as canonical coordinates and to generate the deformation of the hypersurface by a Hamiltonian. The total Hamiltonian of the system is composed from the hypersurface part and the field part.

The resulting formalism closely resembles Hamiltonian geometrodynamics. There, as in parametrized field theories, the total Hamiltonian is a linear combination of a super-Hamiltonian and a supermomentum. The super-Hamiltonian and supermomentum are constrained to vanish. Moreover, their Poisson brackets close in a characteristic way, which is the same for all systems, be they matter fields or geometry. The main difference between geometrodynamics and parametrized field theories is in the role played by hypersurface variables. In geometrodynamics, these variables are inextricably mixed with the dynamical data. In parametrized field theories, they are kept clearly separated from the field variables. This difference is reflected in the structure of the constraints. The geometrodynamical super-Hamiltonian is a hyperbolic function of the momenta, while the super-Hamiltonian of a parametrized field theory is parabolic in the momenta, being linear in the hypersurface momenta and quadratic in the field momenta. The separation of hypersurface variables makes parametrized theories much easier to interpret than geometrodynamics. They are thus an ideal testing ground for concepts proposed in geometrodynamics.

The concept we want to discuss in this paper is that of a symmetry of dynamical evolution. For unconstrained dynamical systems, the definition of symmetry is straightforward. We say that a dynamical variable K generates a symmetry if it is linear in the canonical momenta and has a vanishing Poisson bracket with the Hamiltonian H of the system. However, the presence of constraints creates complications. Symmetry may be conditioned by constraints, because the Poisson bracket $[K, H]$ may vanish only for such values of the canonical variables which satisfy the constraints. A constrained system may have a conditional symmetry even if it does not have any unconditional symmetry.

In parametrized theories and in geometrodynamics, all the dynamics is reducible to constraints. The Hamiltonian which generates the evolution of the system is itself a linear combination of the constraints. We shall study conditional symmetry in this extreme context.

We have already concluded that geometrodynamics does not have any conditional symmetry.[1] It is rather difficult to see what this negative result means from a spacetime viewpoint. In geometrodynamics, the presence or absence of a conditional symmetry is a property of superspace, not a property of spacetimes generated by the evolution of a spatial geometry in superspace. The situation is quite different when viewed within the framework of parametrized field theories. First, these theories may have conditional symmetries. Second, such symmetries are easily linked either with ordinary symmetries of the spacetime background or with internal symmetries of the fields. Our aim is to clarify the meaning and significance of conditional symmetries by spelling out such links in detail.

The study of conditional symmetry is important for canonical quantization. The hyperbolic super-Hamiltonian in geometrodynamics leads to a Klein–Gordon type equation for the state functional. The lack of a conditional symmetry means that superspace equipped by DeWitt's metric does not have a conformal Killing vector which would scale the scalar curvature potential in a prescribed way. The absence of such a vector has serious repercussions. The standard complexification of the space of solutions of a Klein–Gordon equation and the construction of a positive definite inner product fails and a one-system interpretation of quantum geometrodynamics is hard to maintain. In parametrized field theories, the situation is less serious. A parabolic super-Hamiltonian leads to a Schrödinger type equation for quantum fields propagating in a curved spacetime. A positive definite inner product can thus be formally defined even if the theory does not have a conditional symmetry. Still, the symmetry of the background is needed to select a privileged observer and escape thus the well-known ambiguities of an "observer-dependent" quantum field theory. The links between a symmetry of the background and a conditional symmetry of the super-Hamiltonian provide a framework for fixing the observer in the canonical formalism.

2. PARAMETRIZED FIELD THEORIES

We shall briefly explain the basic scheme of parametrized field theories. We state relevant results without proofs and refer to our earlier papers[2] for details. We quote the sections and equations from these papers by prefixing them by the Roman numerals I to IV. A general theory of parametrized fields is rather cumbersome. We thus prefer to restrict ourselves to fields with nonderivative gravitational coupling.

A. Lapse-shift decomposition

We start by cutting the spacetime by an arbitrary spacelike hypersurface $X^\alpha = X^\alpha(x^a)$. The Latin indices always run through the values 1,2,3 and the Greek indices through the values 0,1,2,3. At each point of the hypersurface, we have the basis consisting of the three tangent vectors $X^\alpha_a \equiv X^\alpha{}_{,a}$ to the hypersurface and of the unit normal vector n^α,

$$g_{\alpha\beta}\, X^\alpha_a\, n^\beta = 0, \quad g_{\alpha\beta}\, n^\alpha n^\beta = -1. \tag{2.1}$$

To describe a continuous deformation of the hypersurface in the embedding spacetime, we incorporate it into a one-parameter family of hypersurfaces $X^\alpha = X^\alpha(x^a, t)$. We introduce the deformation vector

$$N^\alpha \equiv \dot{X}^\alpha = \partial X^\alpha(x^a, t)/\partial t \tag{2.2}$$

connecting the points with the same label x^a on two neighboring hypersurfaces. The components N and N^a of the deformation vector with respect to the basis $\{n^\alpha, X^\alpha_a\}$,

$$\begin{aligned} N^\alpha &= N n^\alpha + N^a X^\alpha_a, \\ N &= -N^\alpha n_\alpha, \quad N^a = N^\alpha X^a_\alpha, \end{aligned} \tag{2.3}$$

are called the lapse function and the shift vector.

B. Field projections

To follow the dynamics of an arbitrary tensor field, we project that field perpendicular and parallel to the hypersurface and observe how these projections change when the hypersurface is deformed through spacetime. The projections of a vector field follow the pattern (2.3),

$$\phi^\alpha = \phi^\perp n^\alpha + \phi^a X^\alpha_a, \tag{2.4}$$

$$\begin{aligned} \phi^\perp(x)[X] &= -\phi^\alpha(X(x)) n_\alpha(x)[X], \\ \phi^a(x)[X] &= \phi^\alpha(X(x)) X^a_\alpha(x). \end{aligned} \tag{2.5}$$

They are considered as functions (x) of the labels x^a and functionals $[X]$ of the embedding $X^\alpha(x^a)$. As a rule, we suppress the indices in arguments of functions and functionals.

The projection formulas (2.4)–(2.5) are easily generalized to tensors of an arbitrary rank. Thus, for a second rank tensor $\phi^{\alpha\beta}$,

$$\begin{aligned} \phi^{\alpha\beta} = {} & \phi^{\perp\perp} n^\alpha n^\beta + \phi^{a\perp} X^\alpha_a\, n^\beta \\ & + \phi^{\perp b} n^\alpha X^\beta_b + \phi^{ab} X^\alpha_a\, X^\beta_b, \end{aligned} \tag{2.6}$$

$$\begin{aligned} & \phi^{\perp\perp} = (-1)^2 \phi^{\alpha\beta} n_\alpha n_\beta, \quad \phi^{a\perp} = (-1) \phi^{\alpha\beta} X^a_\alpha\, n_\beta, \\ & \phi^{\perp b} = (-1) \phi^{\alpha\beta} n_\alpha\, X^b_\beta, \quad \phi^{ab} \overset{?}{=} \phi^{\alpha\beta} X^a_\alpha\, X^b_\beta. \end{aligned} \tag{2.7}$$

We shall label by capital Latin indices all possible projections of a tensor field $\phi^{\{\alpha\}} \equiv \phi^{\alpha_1 \cdots \alpha_N}(X)$ or of a collection of such fields. Thus, ϕ^A may mean $\phi^A = \{\phi^\perp, \phi^a; \phi^{\perp\perp}, \phi^{a\perp}, \phi^{\perp b}, \phi^{ab}\}$.

C. Normal and tangential changes

The rate of change $\partial_t \phi^A$ of ϕ^A with the label time t can be decomposed into the normal change $\delta_N \phi^A$ and the tangential change $\delta_{\mathbf{N}} \phi^A$:

$$\dot{\phi}^A \equiv \partial_t \phi^A(x) = \delta_N \phi^A(x) + \delta_{\mathbf{N}} \phi^A(x), \tag{2.8}$$

$$\delta_N \phi^A(x) = \int d^3x'\, N(x') n^{\alpha'}(x')\, \frac{\delta \phi^A(x)}{\delta X^{\alpha'}(x')}, \tag{2.9}$$

$$\delta_{\mathbf{N}} \phi^A(x) = \int d^3x'\, N^{a'}(x') X^{\alpha'}_{a'}(x')\, \frac{\delta \phi^A(x)}{\delta X^{\alpha'}(x')} = L_{\mathbf{N}} \phi^A(x). \tag{2.10}$$

The tangential change is always equal to the Lie derivative $L_{\mathbf{N}}$ of the spatial tensor $\phi^A(x)$ along the shift vector N^a.

There are four different projections of the spacetime covariant derivative $\phi^{\alpha;\beta}$ of a spacetime vector field $\phi^\alpha(x)$. Two of these[3] refer only to a single spacelike hypersurface,

$$\begin{aligned} \phi^{\perp;b} &= \phi^{\perp,b} - K^b_c \phi^c, \\ \phi^{a;b} &= \phi^{a|b} - K^{ab} \phi^\perp. \end{aligned} \tag{2.11}$$

Here, the vertical stroke denotes the spatial covariant derivative with respect to the induced metric g_{ab}, and K_{ab} is the extrinsic curvature of the hypersurface. The remaining two projections,[4]

$$\begin{aligned} N\phi^{\perp;\perp} &= -\delta_N \phi^\perp - \phi^a N_{,a}, \\ N\phi^{a;\perp} &= -\delta_N \phi^a + K^a_b \phi^b N - \phi^\perp N^{,a}, \end{aligned} \tag{2.12}$$

can be used to calculate the change of the field off the hypersurface. Similar formulas can be written for tensor fields of an arbitrary rank.[5] In particular, we get

$$\delta_N\, g_{ab} = -2NK_{ab}, \quad \delta_N\, g^{1/2} = -Ng^{1/2}K \tag{2.13}$$

for the metric tensor.[6] Equations (2.13) help us to pass freely between the covariant and contravariant and tensor and tensor density forms of the projected equations.

D. Reconstruction theorem

We say that the deformation $N^\alpha = Nn^\alpha$ is a hypersurface tilt at a point x if $N(x) = 0$. The tilts leave the spacetime point $X^\alpha = X^\alpha(x)$ fixed.[7] The projections $\phi^\perp, \phi^a$ at x change under hypersurface tilts according to the rules

$$\delta_N \phi^\perp = \phi^a N_{,a}, \quad \delta_N \phi^a = -\phi^\perp N^{,a}. \tag{2.14}$$

Our strategy was to start from a given spacetime field [say, $\phi^\alpha(x)$] and project it $\perp$ and $\|$ to a hypersurface [Eqs. (2.5)]. We shall ask now an inverse question, namely, under what conditions can some given functionals $\phi^\perp(x)[X]$ and $\phi^a(x)[X]$ of $X^\alpha(x)$ be interpreted as the $\perp$ and $\|$ projections (2.5) of a single spacetime vector field $\phi^\alpha(X)$ restricted to the hypersurface $X^\alpha = X^\alpha(x)$. The answer to this question is given by Eqs. (2.10) and (2.14): The functionals $\phi^\perp(x)[X]$ and $\phi^a(x)[X]$ can be reassembled into a spacetime vector field (2.4) if and only if they behave properly under hypersurface shifts [Eq. (2.10)] and hypersurface tilts [Eq. (2.14)]. We shall call this statement *the reconstruction theorem*. Its generalization to arbitrary tensor fields is obvious.

E. Killing fields

Besides dynamical fields $\phi^{\{\alpha\}}$, the Killing vector fields $k^\alpha(X)$ play a prominent role in any study of symmetry. Equations (2.11) and (2.12) help us to project the Killing tensor $k^{(\alpha;\beta)} \equiv k^{\alpha;\beta} + k^{\beta;\alpha}$,

$$k^{(a;b)} = k^{(a|b)} - 2K^{ab}k^\perp, \tag{2.15}$$

$$Nk^{(a;\perp)} = -\delta_N k^a + k^{\perp,a}N - k^\perp N^{,a}, \tag{2.16}$$

$$Nk^{(\perp;\perp)} = -2\delta_N k^\perp - 2k^a N_{,a}. \tag{2.17}$$

They also help us to obtain a projected version of other differential operators. When dealing with conserved currents, we shall need to project the divergence equation[8]:

$$|^4g| K^\alpha{}_{;\alpha} = \delta_N(g^{1/2}K^\perp) + (Ng^{1/2}K^a)_{,a}. \tag{2.18}$$

F. Hamiltonian field theories

The dynamics of tensor fields follows from the field action

$$S[\phi^{\{\alpha\}}] = \int d^4X\, L(\phi^{\{\alpha\}}, \phi^{\{\alpha\}}{}_{,\beta}, g_{\alpha\beta}). \tag{2.19}$$

The field Lagrangian L is a scalar density constructed from the field $\phi^{\{\alpha\}}$, its first derivatives $\phi^{\{\alpha\}}{}_{,\beta}$, and the metric tensor $g_{\alpha\beta}$. For fields with nonderivative gravitational coupling, the derivatives of the metric tensor do not enter the Lagrangian. The field $\phi^{\{\alpha\}}(X)$ is varied to yield the field equations while the metric $g_{\alpha\beta}(X)$ is kept as a prescribed function of X^α. By varying the metric, we obtain the energy–momentum tensor of the field[9]

$$|^4g|^{1/2} T^{\alpha\beta} = 2\frac{\partial L}{\partial g_{\alpha\beta}}. \tag{2.20}$$

The Hamiltonian form of the action is derived by the projection process followed by the Legendre dual transformation. First, the Lagrangian is expressed as a function of the projected variables N, N^a, g_{ab}, ϕ^A and of the derivatives $\dot\phi^A$ along a given one-parameter family $X^\alpha = X^\alpha(x^a, t)$ of spacelike hypersurfaces.[10] Differentiating the hypersurface Lagrangian with respect to $\dot\phi^A$, we obtain the hypersurface momenta π_A. The action is then cast into a Hamiltonian form[11]

$$S[\phi^A, \pi_A] = \int dt \int d^3x(\pi_A\dot\phi^A - NH^\phi - N^aH^\phi_a). \tag{2.21}$$

Here, the lapse and the shift functions are treated as given functions of x^a and t, and the metric $g_{ab}(x)[X_t]$ as a given functional of $X^\alpha(x, t)$,

$$g_{ab}(x)[X_t] = g_{\alpha\beta}(X)X^\alpha_a X^\beta_b|_{X=X(x,t)}. \tag{2.22}$$

The field Hamiltonian

$$H^\phi_N + H^\phi_{\mathbf{N}} = \int d^3x(N(x)H^\phi(x) + N^a(x)H^\phi_a(x)) \tag{2.23}$$

is a linear combination of the field energy density $H^\phi(x)$ and the field momentum density $H^\phi_a(x)$,

$$H^\phi = g^{1/2}T_{\perp\perp}, \quad H^\phi_a = -g^{1/2}T_{\perp a}. \tag{2.24}$$

Both densities are measured by an observer moving $\perp$ to the hypersurface. The field energy is constructed from the variables ϕ_A, π^A, and $g_{ab}(x)$, the field momentum only from the canonical variables ϕ_A and π^A. Neither of these expressions contains N and N^a. The field Hamiltonian (2.23) is obtained by smearing $H^\phi(x)$ by $N(x)$ and $H^\phi_a(x)$ by $N^a(x)$.

The energy density H^ϕ contains all the information about the stress tensor T^{ab} (Ref. 12):

$$g^{1/2}T^{ab} = -2\frac{\partial H^\phi}{\partial g_{ab}}. \tag{2.25}$$

This equation follows from Eq. (2.20); notice that the only term in the action (2.21) which depends on g_{ab} is H^ϕ.

The field evolves according to the Hamilton equations

$$\begin{aligned} \dot\phi^A(x) &= [\phi^A(x), H_N + H_{\mathbf{N}}], \\ \dot\pi_A(x) &= [\pi_A(x), H_N + H_{\mathbf{N}}]. \end{aligned} \tag{2.26}$$

Recalling Eqs. (2.9) and (2.23), we can split Eqs. (2.26) into the lapse and the shift parts,

$$\delta_N\phi^A(x) = [\phi^A(x), H_N], \quad \delta_N\pi_A(x) = [\pi_A(x), H_N], \tag{2.27}$$

and

$$\delta_{\mathbf{N}}\phi^A(x) = [\phi^A(x), H_{\mathbf{N}}], \quad \delta_{\mathbf{N}}\pi_A(x) = [\pi_A(x), H_{\mathbf{N}}]. \tag{2.28}$$

Because the left-hand sides of Eqs. (2.28) must reproduce the spatial Lie derivatives of the canonical variables, Eqs. (2.28) determine the field momentum:

$$H_{\mathbf{N}} = \int d^3x\, \pi_A(x) L_{\mathbf{N}}\phi^A(x) \left[= -\int d^3x\, \phi^A(x) L_{\mathbf{N}}\pi_A(x)\right]. \tag{2.29}$$

As an example, we write the field momentum of a scalar field and a vector field[13]:

$$H^\phi_a = \pi\phi_{,a}, \tag{2.30}$$

$$H^\phi_a = \pi^\perp\phi_{\perp,a} + \pi_b\phi^b{}_{,a} + (\pi_a\phi^b)_{,b}. \tag{2.31}$$

G. Model field theories

We shall give now a few typical examples of field theories with nonderivative gravitational coupling. First, study a scalar field with the Lagrangian

$$L = |{}^4g|^{1/2}(-\tfrac{1}{2}U(\phi)\, g^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta} - W(\phi)), \tag{2.32}$$

where $U \neq 0$ and W are two arbitrary functions of ϕ. The expression (2.32) is the most general scalar density which can be formed from the variables $g_{\alpha\beta}$, ϕ, and $\phi_{,\alpha}$ so that it is at most quadratic in $\phi_{,\alpha}$. From the Lagrangian (2.32), we get

$$H^{\phi} = \tfrac{1}{2}\, U^{-1} g^{-1/2}\pi^2 + \tfrac{1}{2}\, U g^{1/2} g^{ab}\phi_{,a}\phi_{,b} + g^{1/2}W. \tag{2.33}$$

For a Klein–Gordon field, $U = 1$ and $W = \frac{1}{2}m^2$. Equation (2.33) then reduces to

$$H^{\phi} = \tfrac{1}{2}\, g^{-1/2}\pi^2 + V,$$
$$V \equiv \tfrac{1}{2}\, g^{1/2}(g^{ab}\phi_{,a}\phi_{,b} + m^2\phi^2). \tag{2.34}$$

From Eq. (2.25) we obtain the stress tensor of the Klein–Gordon field,

$$g^{1/2}T_{ab} = g^{1/2}\phi_{,a}\phi_{,b} + (\tfrac{1}{2}\, g^{-1/2}\pi^2 - V)\, g_{ab}. \tag{2.35}$$

For a complex Klein–Gordon field, the field Lagrangian has the form

$$L = -|{}^4g|^{1/2}(g^{\alpha\beta}\phi^*{}_{,\alpha}\phi_{,\beta} + m^2\phi^*\phi); \tag{2.36}$$

here, ϕ and ϕ^* are varied as independent variables. We get

$$H^{\phi} = g^{-1/2}\pi^*\pi + g^{1/2}(g^{ab}\phi^*{}_{,a}\phi_{,b} + m^2\phi^*\phi). \tag{2.37}$$

Unlike the real field, the complex field has a conserved current

$$K^{\alpha} \equiv \tfrac{1}{2}i(\phi^*\phi^{,\alpha} - \phi^{*,\alpha}\phi), \quad K^{\alpha}{}_{;\alpha} = 0. \tag{2.38}$$

Its projection can be expressed in terms of the canonical variables. In particular,

$$g^{1/2}K^{\perp} = \tfrac{1}{2}i(\phi\pi - \phi^*\pi^*). \tag{2.39}$$

As our next example, take a massive vector field with the Lagrangian[14]

$$L = -|{}^4g|^{1/2}(\tfrac{1}{4}\, g^{\alpha\gamma}g^{\beta\delta}\phi_{[\alpha,\beta\,]}\phi_{[\gamma,\delta\,]} + \tfrac{1}{2}m^2 g^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta}). \tag{2.40}$$

This leads to the hypersurface action

$$S\,[\phi_a, \pi^a] = \int dt \int d^3x(\pi^a\dot{\phi}_a - NH^{\phi} - N^aH^{\phi}_a), \tag{2.41}$$

$$H^{\phi} = \tfrac{1}{2}\, g^{-1/2}g_{ab}\pi^a\pi^b + \tfrac{1}{2}m^{-2}g^{-1/2}(\pi^a{}_{,a})^2$$
$$+ \tfrac{1}{4}\, g^{1/2}g^{ac}g^{bd}\phi_{[a,b\,]}\phi_{[c,d\,]} + \tfrac{1}{2}m^2 g^{1/2}g^{ab}\phi_a\phi_b, \tag{2.42}$$

$$H^{\phi}_a = -\phi_a\pi^b{}_{,b} - \phi_{[a,b\,]}\pi^b. \tag{2.43}$$

Note that the projections $\phi_{\perp}, \pi^{\perp}$ were eliminated from the action (2.41) by using the relation

$$m^2\phi_{\perp} + g^{-1/2}\pi^a{}_{|a} = 0. \tag{2.44}$$

For $m = 0$, the Lagrangian (2.40) describes Maxwell's electrodynamics. Here, Eq. (2.44) reduces to a supplementary constraint

$$\pi^a{}_{,a} = 0 \tag{2.45}$$

and the scalar potential $\phi_{\perp}$ can no longer be eliminated from the Hamiltonian. We have

$$H^{\phi} = \phi_{\perp}\pi^a{}_{,a} + \tfrac{1}{2}\, g^{-1/2}g_{ab}\pi^a\pi^b$$
$$+ \tfrac{1}{4}\, g^{1/2}g^{ac}g^{bd}\phi_{[a,b\,]}\phi_{[c,d\,]} \tag{2.46}$$

and $\phi_{\perp}$ enters the action as a Lagrange multiplier. The momentum π^a has the meaning of the electric field strength measured by an observer moving perpendicular to the hypersurface.

H. Parametrized Hamiltonian field theories

We are going now to treat the hypersurface variables as canonical coordinates. The deformation vector N^{α} is defined by Eq. (2.2). We can reproduce this definition from the action

$$S\,[X^{\alpha}, P_{\alpha}, N^{\alpha}] = \int dt \int d^3x(P_{\alpha}\dot{X}^{\alpha} - N^{\alpha}P_{\alpha}). \tag{2.47}$$

Varying S with respect to the momentum P_{α}, we recover Eq. (2.2). Varying it with respect to the multiplier N^{α} and the hypersurface variable X^{α} we obtain the equations

$$P_{\alpha} = 0, \quad \dot{P}_{\alpha} = 0, \tag{2.48}$$

which tell us that the hypersurface momentum is trivial and remains trivial in the dynamical evolution.

Instead of varying the deformation vector N^{α}, we can replace it by the lapse function N and the shift vector N^a from Eq. (2.3). The action (2.47) then assumes the form

$$S\,[X^{\alpha}, P_{\alpha}; N, N^a] = \int dt \int d^3x(P_{\alpha}\dot{X}^{\alpha} - NP - N^aP_a), \tag{2.49}$$

where P and P_a are given by the expressions

$$P = -P_{\perp} = n^{\alpha}[X\,]P_{\alpha}, \quad P_a = X^{\alpha}_a P_{\alpha}. \tag{2.50}$$

They are to be considered as functionals of X^{α} and P_{α}. The variation of P_{α} in the action (2.49) yields directly the lapse-shift decomposition (2.3).

In the field action (2.21), the lapse and shift functions are externally prescribed and are not to be varied. When we develop the canonical data by Hamilton's equations (2.26), we should remember as an independent fact that X^{α} changes by Eqs. (2.2) and (2.3), inducing thus a change of the spatial metric (2.22). By adjoining the hypersurface action (2.49) to the field action (2.21), these changes are accounted for within a canonical formalism. This process is known as parametrization of the field theory. The parametrized action has the form

$$S\,[\phi^A, \pi_A; X^{\alpha}, P_{\alpha}; N, N^a]$$
$$= \int dt \int d^3x(P_{\alpha}\dot{X}^{\alpha} + \pi_A\dot{\phi}^A - NH - N^aH_a), \tag{2.51}$$

where

$$H = P + H^{\phi}, \quad H_a = P_a + H^{\phi}_a \tag{2.52}$$

are called the super-Hamiltonian and supermomentum of the parametrized theory. The metric g_{ab} in H^{ϕ} is to be considered as a functional of $X^{\alpha}(x)$ in accordance with Eq. (2.22). The variation of P_{α} still gives us the lapse-shift decomposition. The variation of the field variables ϕ^A and π_A leads to the old field equations (2.26). The hypersurface momentum P_{α}, however, ceases to be trivial. By varying the multipliers

N, N^a, we get the constraints

$$H = 0 = H_a. \tag{2.53}$$

The constraints can easily be solved with respect to the hypersurface momentum P_α,

$$P_\alpha = -P n_\alpha + P_a X^a_\alpha,$$
$$P = -H^\phi, \quad P_a = -H_a. \tag{2.54}$$

We see that $-P$ is to be interpreted as the energy density and $-P_a$ as the momentum density of the field. The remaining Hamilton equations, obtained by varying X^α, tell us how these densities change from one hypersurface to another.

The change of an arbitrary dynamical variable K constructed from the canonical variables ϕ^A, π_A, X^α, P_α is given by its Poisson bracket with the Hamiltonian

$$H_N + H_{\mathbf{N}} = \int d^3x (N(x) H(x) + N^a(x) H_a(x)). \tag{2.55}$$

In brief,

$$\dot{K} = [K, H_N + H_{\mathbf{N}}]. \tag{2.56}$$

Equation (2.56) summarizes the content of Hamilton's equations of the parametrized theory.

If the Poisson bracket (2.56) vanishes for all N and $\mathbf{N}$, K has the same value on every spacelike hypersurface, i.e., it is conserved. In fact, for K to be conserved, $[K, H_N + H_{\mathbf{N}}]$ does not need to vanish identically in the canonical variables X^α, P_α, ϕ^A, π^A, but only for such values of the variables which satisfy the constraints (2.53). If a dynamical variable F vanishes only modulo the constraints, we say with Dirac[15] that it vanishes weakly, and write $F \approx 0$. On the other hand, if a dynamical variable F vanishes identically in the canonical variables, we say that it vanishes strongly, and write $F = 0$. The weakly vanishing Poisson bracket (2.56) is a basic ingredient in our definition of a conditional symmetry.

I. Closure relations

The constraints (2.53) must be preserved in time, which means that the Poisson bracket

$$[H_M + H_{\mathbf{M}}, H_N + H_{\mathbf{N}}] \tag{2.57}$$

must weakly vanish for arbitrary smearing functions M, $\mathbf{M}$ and N, $\mathbf{N}$. This is ensured by the closure relations[16]

$$[H_M, H_N] = H_{(M,N)}, \tag{2.58}$$
$$[H_{\mathbf{M}}, H_N] = H_{\mathbf{M}\cdot\partial N}, \tag{2.59}$$
$$[H_{\mathbf{M}}, H_{\mathbf{N}}] = H_{[\mathbf{M},\mathbf{N}]}, \tag{2.60}$$

where

$$(M, N)^a \equiv M N^{,a} - N M^{,a}, \tag{2.61}$$
$$\mathbf{M}\cdot\partial N \equiv M^a N_{,a}, \tag{2.62}$$
$$[\mathbf{M}\cdot\mathbf{N}] = L_{\mathbf{M}} \mathbf{N} \tag{2.63}$$

define a composition of the smearing functions. When some of these functions are themselves dynamical variables, e.g., when $M(x) = M(x)[X, P, \phi, \pi]$ and $M^a(x) = M^a(x) \times [X, P, \phi, \pi]$, Eqs. (2.58)–(2.60) acquire additional terms:

$$[H_M, H_N] = H_{[M, H_N]} + H_{(M,N)}, \tag{2.64}$$
$$[H_{\mathbf{M}}, H_N] = H_{\mathbf{M}\cdot\partial N} + H_{[\mathbf{M}, H_N]}, \tag{2.65}$$
$$[H_{\mathbf{M}}, H_{\mathbf{N}}] = H_{[\mathbf{M}, H_{\mathbf{N}}]} + H_{[\mathbf{M},\mathbf{N}]}, \tag{2.66}$$
$$[N_M, H_{\mathbf{N}}] = H_{-\mathbf{N}\cdot\partial M} + H_{[M, H_{\mathbf{N}}]}. \tag{2.67}$$

The right-hand sides of Eqs. (2.58)–(2.60) or Eqs. (2.64)–(2.66) are linear combinations of the constraints, which ensures that the Poisson bracket (2.57) weakly vanishes. The corresponding conserved quantities are of course trivial. They coincide with the constraints and therefore weakly vanish.

In an extreme case, there is no field to propagate on the spacetime background. The constraint functions (2.52) then reduce to the expressions (2.50). These expressions, P and P_a, thus satisfy the same closure relations (2.58)–(2.67) as the original functions H and H_a.

The closure relations (2.58)–(2.67) in parametrized field theories are exactly the same as the corresponding relations in Hamiltonian geometrodynamics.[1] This makes parametrized field theories so suitable as models for geometrodynamics.

3. CONDITIONAL SYMMETRIES IN PARAMETRIZED FIELD THEORIES

We say that a dynamical variable $K[X^\alpha, \phi^A; P_\alpha, \pi_A]$ generates a conditional symmetry in a parametrized field theory if it is linear in the canonical momenta,

$$\begin{aligned} K = K_{(k)} + K_{(h)} &= \int d^3x \, (k^\alpha(x)[X,\phi] P_\alpha(x) \\ &\quad + h^A(x)[X,\phi] \pi_A(x)) \\ &= \int d^3x \, (k^\perp(x)[X,\phi] P(x) + k^a(x)[X,\phi] P_a(x) \\ &\quad + h^A(x)[X,\phi] \pi_A(x)), \end{aligned} \tag{3.1}$$

and its Poisson bracket with the Hamiltonian $H_N + H_{\mathbf{N}}$ vanishes for such values of the canonical variables $X^\alpha, \phi^A, P_\alpha, \pi_A$ which satisfy the constraints:

$$[K, H_N + H_{\mathbf{N}}] \approx 0. \tag{3.2}$$

Because the lapse function and the shift vector are arbitrary, Eq. (3.2) means that the variable K is conserved under an arbitrary deformation of the hypersurface. Taking into account this arbitrariness, we can replace Eq. (3.2) by an infinite set of equations,

$$[K, H(x)] \approx 0 \tag{3.3}$$

and

$$[K, H_a(x)] \approx 0, \tag{3.4}$$

four equations for each point of the hypersurface.

There are two kinds of constraints we can encounter in a parametrized field theory. First, there are always the super-Hamiltonian and supermomentum constraints which are introduced by the process of parametrization. Second, if we are dealing with gauge theories, there are still supplementary constraints, linear in the field momenta. Our main motivation for studying parametrized field theories is to draw parallels to geometrodynamics. There are no supplementary constraints in vacuum geometrodynamics. To concentrate on the essentials, we shall limit our further discussion to parametrized theories without supplementary constraints.

Broadly speaking, we want to find all generators (3.1) which satisfy the weak equations (3.3) and (3.4). For this purpose, it is advantageous to replace the weak equations by an equivalent set of strong equations. One way of doing this is to adjoin the constraints (2.51)–(2.52) to the equations by means of Lagrange multipliers. This is the only practical method of analyzing the weak equations if the constraints cannot be explicitly solved, as it is the case in geometrodynamics. However, in parametrized theories the constraints can be solved with respect to the hypersurface momenta P_α. This suggests an alternative way of replacing the weak equations (3.3) and (3.4) by strong equations: After evaluating the Poisson brackets as function(al)s of the canonical variables, we replace the hypersurface momenta by their expressions (2.54) in terms of the remaining canonical variables ϕ^A, π_A, and X^α. For theories without supplementary constraints, these data are completely arbitrary and the Poisson brackets must thus strongly vanish in these variables. This method makes the analysis of conditional symmetries in parametrized field theories considerably simpler than in geometrodynamics.

The clear separation of the hypersurface momenta P_α from the field momenta π_A also leads to a useful classification of the generators. We say that a generator K is P-restricted if it does not depend on the field momenta π_A, i.e., if $h^A = 0$; we say that it is π-restricted if it does not depend on the hypersurface momenta P_α, i.e., if $k^\alpha = 0$; and we call it mixed if both kinds of momenta are present in the expansion (3.1). We shall show that the P-restricted generators are associated with symmetries of the spacetime background and the π-restricted generators with internal symmetries of the field. A mixed generator can either be a sum of two separately conserved generators, one of which is P-restricted and the other one is π-restricted or, under circumstances which we shall explain later, they may again correspond to a spacetime symmetry.

4. INVARIANCE OF GENERATORS UNDER SPATIAL DIFFEOMORPHISMS

Any two dynamical variables K and $\bar K$ which coincide on the constraint surface,

$$\bar K \approx K, \tag{4.1}$$

are for all physical purposes equivalent to each other. Moreover, by virtue of the closing relations (2.64)–(2.67), when K is conditionally conserved, $\bar K$ is also conditionally conserved,

$$\dot K \approx 0 \Rightarrow \dot{\bar K} \approx 0. \tag{4.2}$$

When K is linear in the momenta and conditionally conserved, we can adjoin to it the supermomentum constraint H_a,

$$\bar K = K + \int d^3x\, \mu^a(x)[X,\phi\,]H_a(x), \tag{4.3}$$

without disturbing either the linearity or the conservation. The generators of conditional symmetries thus fall into equivalence classes (4.3) modulo the supermomentum constraint. We can consider Eq. (4.3) as a gauge transformation on the generators produced by the gauge functional $\mu^a(x)[X,\phi\,]$.

We can use the gauge transformation to eliminate the hypersurface momenta P_a from the generator (3.1). It suffices to take $\mu^a = -k^a$; then,

$$\bar k^a = 0 \tag{4.4}$$

and $\bar K$ reduces to

$$\bar K = \int d^3x (k(x)P(x) + \bar h^A(x)\pi_A(x)). \tag{4.5}$$

We can thus always represent the equivalence class (4.3) by that generator (4.5) which satisfies Eq. (4.4).

More important, we can prove that the generator (4.5) must satisfy the strong equation

$$[\bar K, H_{\rm N}\,] = 0, \tag{4.6}$$

while the original generator satisfied only a weak equation (3.4). The Poisson bracket in Eq. (4.6) can be evaluated separately for the P part and the π part of the generator. The first Poisson bracket,

$$\left[\int d^3x\, k(x)P(x), H_{\rm N}\right] = \int d^3x \{k(x)[P(x), H_{\rm N}\,] + [k(x), H_{\rm N}\,]P(x)\}, \tag{4.7}$$

can be written as

$$\left[\int d^3x\, k(x)P(x), H_{\rm N}\right] = \int d^3x(-L_{\rm N}\, k(x) + [k(x), H_{\rm N}\,])P(x) \tag{4.8}$$

once we realize that $[P(x), H^\phi_{\rm N}\,] = 0$ and apply the field-free limit of Eq. (2.59) to the bracket $[P(x), P_{\rm N}\,]$. The second Poisson bracket,

$$\left[\int d^3x\, \bar h^A(x)\pi_A(x), H_{\rm N}\right] = \int d^3x \{\bar h^A(x)[\pi_A(x), H_{\rm N}\,] + [\bar h^A(x), H_{\rm N}\,]\pi_A(x)\}, \tag{4.9}$$

allows a similar rearrangement. Because of Eq. (2.29) we can integrate by parts,

$$\int d^3x\, \bar h^A(x)[\pi_A(x), H_{\rm N}\,] = \int d^3x\, \bar h^A(x) L_{\rm N}\pi_A(x) = -\int d^3x\, L_{\rm N}\, \bar h^a(x)\cdot\pi_A(x), \tag{4.10}$$

and write

$$\left[\int d^3x\, \bar h^A(x)\pi_A(x), H_{\rm N}\right] = \int d^3x(-L_{\rm N}\bar h^A(x) + [\bar h^A(x), H_{\rm N}\,])\pi_A(x). \tag{4.11}$$

Start now from the weak form of Eq. (4.6),

$$[K, H_{\rm N}\,] = \int d^3x\{(-L_{\rm N}\, k(x) + [k(x), H_{\rm N}\,])P(x) + (-L_N\, \bar h^A(x) + [\bar h^A(x), H_{\rm N}\,])\pi_A(x)\} \approx 0. \tag{4.12}$$

Following the method explained in Sec. 3, we replace $P(x)$ in Eq. (4.12) by $-H^{\phi}(x)[X,\phi,\pi]$ and require that the resulting expression vanish strongly in the remaining variables. Typically, H^{ϕ} is a nondegenerate quadratic function of the field momenta $\pi_A(x)$ [cf. Eq. (8.3)]. The coefficients of $H^{\phi}(x)$ and $\pi_A(x)$ must therefore vanish separately,

$$-L_{\mathbf{N}}\, k(x) + [k(x), H_{\mathbf{N}}] = 0 \tag{4.13}$$

and

$$-L_{\mathbf{N}}\, \bar{h}^A(x) + [\bar{h}^A(x), H_{\mathbf{N}}] = 0, \tag{4.14}$$

for our expression to vanish identically in the momentum variables π_A. Substituting these equations back into the original expression (4.12), we conclude that the Poisson bracket $[K, H_{\mathbf{N}}]$ strongly vanishes, Eq. (4.6).

Supermomentum constraints generate the transformation of dynamical variables under spatial diffeomorphisms, Eqs. (2.28)–(2.29). The weak equation (3.4) implies that K is invariant under spatial diffeomorphisms only for those values of the canonical variables which satisfy the constraints. On the other hand, the strong equation (4.6) means that $\bar{K}$ is an invariant throughout the whole phase space. The strong equation (4.13) then tells us that $k(x)$ is a spatial scalar and the strong equation (4.14) tells us that $\bar{h}^A$ is a spatial tensor of the same rank as π_A. This conveniently simplifies our further considerations.

It is worthwhile to note where the argument fails when we try to repeat it for the original form (3.1) of the generator. Starting from this form, we pick up an additional term

$$[P_{\mathbf{k}}, H_{\mathbf{N}}] = P_{\mathbf{M}},$$

$$\mathbf{M}(x) = -L_{\mathbf{N}}\, \mathbf{k}(x) + [\mathbf{k}(x), H_{\mathbf{N}}] \tag{4.15}$$

in the Poisson bracket $[K, H_{\mathbf{N}}]$, so that our old equation (4.12) reads

$$\begin{aligned}[K, H_{\mathbf{N}}] = \int d^3x\{ & (-L_{\mathbf{N}}\, k(x) + [k(x), H_{\mathbf{N}}])P(x) \\ & + (-L_{\mathbf{N}}\, h^A(x) + [h^A(x), H_{\mathbf{N}}])\pi_A(x) \\ & + M^a(x)P_a(x)\} \approx 0.\end{aligned} \tag{4.16}$$

The momentum $P_a(x)$ can be eliminated from Eq. (4.16) by using the supermomentum constraint,

$$\begin{aligned}\int d^3x\, M^a(x)P_a(x) &\approx -\int d^3x\, M^a(x)H^{\phi}_a(x) \\ &= \int d^3x\, L_{\mathbf{M}}\phi^A(x)\cdot\pi_A(x),\end{aligned} \tag{4.17}$$

yielding a term linear in the field momentum. Repeating the reasoning which led us from Eq. (4.12) to Eqs. (4.13)–(4.14), we see that Eq. (4.13) still holds, but Eq. (4.14) gets replaced by

$$-L_{\mathbf{M}}\phi^A + (-L_{\mathbf{N}}\, h^A + [h^A, H_{\mathbf{N}}]) = 0. \tag{4.18}$$

From this equation we are unable to conclude that $\mathbf{M}$ and $(-L_{\mathbf{N}}\, h^A + [h^A, H_{\mathbf{N}}])$ vanish separately and so we cannot maintain that the Poisson bracket (4.16) strongly vanishes.

Of course, $\bar{K}$ is not the only representative of the equivalence class (4.3) which is a spatial invariant, Eq. (4.6). If $k^a(x)[X,\phi]$ is any spatial vector constructed from the canonical coordinates X and ϕ and if we put

$$\bar{\bar{K}} = \bar{K} + H_{\mathbf{k}}, \tag{4.19}$$

$\bar{\bar{K}}$ is also a spatial invariant,

$$[\bar{\bar{K}}, H_{\mathbf{N}}] = [H_{\mathbf{k}}, H_{\mathbf{N}}] = H_{-L_{\mathbf{N}}\mathbf{k} + [\mathbf{k}, H_{\mathbf{N}}]} = 0. \tag{4.20}$$

There are thus spatially invariant generators containing the hypersurface momentum P_a in their expansion (3.1).

5. SYMMETRY OF THE BACKGROUND AS A CONDITIONAL SYMMETRY

Take an arbitrary field that propagates in a spacetime with a Killing vector $k^{\alpha}(x)$,

$$k_{(\alpha;\beta)} = 0. \tag{5.1}$$

The projection

$$K^{\alpha}(X) \equiv T^{\alpha\beta}(X)k_{\beta}(X) \tag{5.2}$$

of the energy–momentum tensor $T^{\alpha\beta}$ into the Killing vector k^{α} satisfies the equation of continuity

$$K^{\alpha}{}_{;\alpha} = 0. \tag{5.3}$$

Similarly, when the field has a trace-free energy–momentum tensor,

$$T^{\alpha}_{\alpha} = 0, \tag{5.4}$$

and propagates in a spacetime which has a conformal Killing vector,

$$k_{(\alpha;\beta)} = \Lambda(X)\, g_{\alpha\beta}, \tag{5.5}$$

the vector field (5.2) again satisfies the equation of continuity (5.3).

We project Eq. (5.3) along an arbitrary spacelike hypersurface, Eq. (2.18), and integrate it over d^3x. If the hypersurface is compact or the field vanishes sufficiently fast at infinity, the spatial divergence drops out and the quantity

$$K \equiv \int d^3x\, g^{1/2}K_{\perp} \tag{5.6}$$

is conserved under normal deformations of the hypersurface:

$$\delta_N K = 0. \tag{5.7}$$

Using Eq. (2.24) and the constraint (2.52), $g^{1/2}K_{\perp}$ can be expressed as a linear function of the hypersurface momenta,

$$\begin{aligned}g^{1/2}K_{\perp} &\equiv -g^{1/2}n_{\alpha}T^{\alpha\beta}k_{\beta} \\ &= (T_{\perp\perp}n_{\beta} + T_{\perp b}\, X^b_{\beta})k^{\beta} \\ &= (H^{\phi}n_{\beta} - H^{\phi}_b\, X^b_{\beta})k^{\beta} \\ &\approx (-Pn_{\beta} + P_a\, X^a_{\beta})k^{\beta} = k^{\beta}P_{\beta}.\end{aligned} \tag{5.8}$$

This suggests that

$$K \equiv \int d^3x\, K^{\alpha}(X(x))P_{\alpha}(x) \tag{5.9}$$

generates a symmetry of the parametrized field theory. We shall show that this symmetry is unconditional if the background has a true Killing vector (5.1) and conditional if the background has a conformal Killing vector (5.5).

Take an arbitrary vector field $k^{\alpha}(x)$, not necessarily a Killing field (5.1) or (5.5), and calculate the Poisson bracket of the dynamical variable (5.9) with H_N. By the field-free limit of the closure relations (2.64) and (2.65),

$$\begin{aligned}[K, P_N] &= \left[P_{k^\perp} + P_{\mathbf{k}}, P_N\right] \\ &= (P_{[k^\perp, P_N]} + P_{(k^\perp, N)}) \\ &\quad + (P_{[\mathbf{k}, P_N]} + P_{\mathbf{k}\cdot\partial N}) \\ &= P_{\delta_N k^\perp + \mathbf{k}\cdot\partial N} + P_{\delta_N \mathbf{k} + (k^\perp, N)}. \end{aligned} \tag{5.10}$$

From the projection equation (2.17) we get

$$\delta_N k^\perp + \mathbf{k}\cdot\partial N = -NK^{\perp;\perp} \tag{5.11}$$

and from the projection equation (2.16) we get

$$(\delta_N \mathbf{k} + (k^\perp, N))^a = -Nk^{(a;\perp)}. \tag{5.12}$$

Therefore,

$$[K, P_N] = -P_{Nk^{\perp;\perp}} - P_{Nk^{(\perp;a)}} \tag{5.13}$$

or

$$[K, P(x)] = -\tfrac{1}{2}k^{(\perp;\perp)}(x)P(x) - k^{(\perp;a)}(x)P_a(x). \tag{5.14}$$

Next, we evaluate the Poisson bracket $[K, H^\phi_N]$. The hypersurface variables enter into H^ϕ_N entirely through the metric $g_{ab}(x)[X]$. Therefore,

$$[K, H^\phi_N] = \int d^3x [K, g_{ab}(x)] \frac{\delta H^\phi_N}{\delta g_{ab}(x)}. \tag{5.15}$$

The variational derivative $\delta H^\phi_N / \delta g_{ab}(x)$ yields the stress tensor by Eq. (2.25). The Poisson bracket $[\, g_{ab}(x), K\,]$ is the change of the metric induced by the deformation k^α of the hypersurface,

$$\begin{aligned}[\, g_{ab}(x), K\,] &= \left[\, g_{ab}(x), P_{k^\perp} + P_{\mathbf{k}}\,\right] \\ &= -2k^\perp K_{ab} + k_{(a|b)} = k_{(a;b)}. \end{aligned} \tag{5.16}$$

Equation (5.15) thus gives

$$[K, H^\phi_N] = \tfrac{1}{2}\int d^3x\, N g^{1/2} T^{ab} k_{(a;b)} \tag{5.17}$$

or

$$[K, H^\phi(x)] = \tfrac{1}{2} g^{1/2} T^{ab} k_{(a;b)}. \tag{5.18}$$

We can put now the two pieces, Eqs. (5.14) and (5.18), together and obtain thus an important identity

$$\begin{aligned}[K, H(x)] &= -\tfrac{1}{2}k^{(\perp;\perp)}H(x) - k^{(\perp;a)}H_a(x) \\ &\quad + \tfrac{1}{2} g^{1/2} k_{(\alpha;\beta)} T^{\alpha\beta}. \end{aligned} \tag{5.19}$$

In the process, we have reassembled the projections T_{ab}, $T^{\perp a}$, and $T^{\perp\perp}$ into the spacetime energy–momentum tensor $T^{\alpha\beta}$,

$$\begin{aligned}&\tfrac{1}{2} g^{1/2} T^{ab} k_{(a;b)} + k^{(\perp;a)} H^\phi_a + k^{\perp;\perp} H^\phi \\ &\quad = \tfrac{1}{2} g^{1/2} k_{(\alpha;\beta)} T^{\alpha\beta}. \end{aligned} \tag{5.20}$$

Of course, the dynamical variable (5.9) is an invariant under spatial transformations,

$$[K, H_a(x)] = 0. \tag{5.21}$$

The strong equation (5.21) can be verified by a direct evaluation of the Poisson bracket.

From Eqs. (5.19) and (5.21) we are able to draw the desired conclusions:

Theorem: If the background has a Killing vector $k^\alpha(x)$, Eq. (5.1), any parametrized field theory on that background has the unconditional symmetry (5.9).

Theorem: If the field has a trace-free energy–momentum tensor, Eq. (5.4), and the background has a conformal Killing vector, Eq. (5.5), the parametrized field theory has the conditional symmetry (5.9) with

$$[K, H(x)] = \tfrac{1}{2}\Lambda(x) H(x). \tag{5.22}$$

A special but interesting situation arises for a massless Klein–Gordon field. In this case, the energy–momentum tensor is not trace-free, but its trace reduces to a pure divergence modulo the field equation $\Box\phi = 0$:

$$T^\alpha_\alpha = -g^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta} = -(g^{\alpha\beta}\phi\phi_{,\beta})_{;\alpha} + \phi\,\Box\phi. \tag{5.23}$$

This leads to a conservation law if the field propagates in a spacetime which admits a homothetic motion [i.e., which has a conformal Killing vector field (5.5) with a constant $\Lambda(X) \equiv \Lambda = \text{const}$]. Indeed,

$$(T^{\alpha\beta}k_\beta)_{;\alpha} = \tfrac{1}{2}\Lambda T^{\alpha\beta} g_{\alpha\beta} = -\tfrac{1}{2}(\Lambda g^{\alpha\beta}\phi\phi_{,\beta})_{;\alpha} \tag{5.24}$$

and so the vector

$$K^\alpha \equiv T^{\alpha\beta}k_\beta + \tfrac{1}{2}\Lambda g^{\alpha\beta}\phi\phi_{,\beta} \tag{5.25}$$

satisfies the equation of continuity (5.3). As a result, the variable K defined by Eq. (5.6) is conserved, Eq. (5.7).

We can express K as a linear functional of the momenta $\{P_\alpha, \pi\}$ by using the rearrangement (5.8) and introducing $\pi = -g^{1/2}\phi_{,\perp}$:

$$K = \int d^3x (k^\alpha(X(x)) P_\alpha(x) - \tfrac{1}{2}\Lambda\phi(x)\pi(x)). \tag{5.26}$$

It is easy to check that this dynamical variable obeys Eq. (5.22) (with a constant Λ): The P part of the generator (5.26) is subject to the identity (5.19), while a direct evaluation yields

$$\begin{aligned}\left[\int d^3x' \phi(x')\pi(x'), H(x)\right] &= g^{-1/2}\pi^2 - g^{1/2} g^{ab}\phi_{,a}\phi_{,b} \\ &= g^{1/2}(T^a_a - T_{\perp\perp}) = g^{1/2} T^\alpha_\alpha. \end{aligned} \tag{5.27}$$

It should be emphasized that neither the P-restricted part nor the π-restricted part of the mixed generator (5.26) are conserved separately, but they are both needed for the mutual cancellation of the trace term T^α_α. To summarize,

Theorem: If a massless Klein–Gordon field propagates in a spacetime which admits a homothetic motion [Eq. (5.5), with $\Lambda = \text{const}$], the mixed dynamical variable (5.26) generates a conditional symmetry, Eq. (5.22).

6. CONDITIONAL SYMMETRIES AND INVARIANCE OF THE ACTION

The meaning of the P-restricted conditional symmetries can also be grasped through their connection with the invariance of the parametrized action (2.51) under transformations induced by spacetime diffeomorphisms.

To see how the spacetime diffeomorphisms enter into the game, recall that a given vector field $k^\alpha(x)$ produces a one-parameter group of diffeomorphisms

$$X^\alpha = X^\alpha(X^{\beta'}, \tau) \tag{6.1}$$

of the spacetime background by the equation

$$\frac{dX^{\alpha}(\tau)}{d\tau} = k^{\alpha}(X(\tau)), \quad X^{\alpha}(0) = X^{\alpha'}. \tag{6.2}$$

On the other hand, any covector P_{α} is pulled back by the diffeomorphisms (6.1),

$$P_{\alpha'} = X^{\beta}_{\alpha'}(X',\tau)P_{\beta}(\tau), \tag{6.3}$$

and thus

$$\frac{dP_{\alpha}(\tau)}{d\tau} = -k^{\beta}{}_{,\alpha}(X(\tau))P_{\beta}(\tau). \tag{6.4}$$

Equations (6.2) and (6.3) can be interpreted as a one-parameter group of canonical transformations

$$\frac{dX^{\alpha}}{d\tau} = [X^{\alpha}, K], \quad \frac{dP_{\alpha}}{d\tau} = [P_{\alpha}, K] \tag{6.5}$$

generated in an 8-dimensional phase space $\{X^{\alpha},P_{\alpha}\}$ by the dynamical variable

$$K = k^{\alpha}(X)P_{\alpha}. \tag{6.6}$$

The hypersurface variables $\{X^{\alpha}(x), P_{\alpha}(x)\}$ form an $8\infty^3$ dimensional phase space. The diffeomorphisms (6.1) act on the hypersurfaces $X^{\alpha} = X^{\alpha}(x)$ by dragging them along the flowlines $k^{\alpha}(x)$ and on the hypersurface momenta $P_{\alpha}(x)$ by pulling them back according to Eq. (6.3):

$$X^{\alpha}(x,\tau) = X^{\alpha}(X^{\beta'}(x),\tau),$$
$$P_{\alpha'} = X^{\beta}_{\alpha'}(X^{\gamma'}(x),\tau)P_{\beta}(x,\tau). \tag{6.7}$$

This action can again be interpreted as a one-parameter group of canonical transformations

$$\frac{\partial X^{\alpha}(x,\tau)}{\partial\tau} = [X^{\alpha}(x,\tau), K],$$
$$\frac{\partial P_{\alpha}(x,\tau)}{\partial\tau} = [P_{\alpha}(x,\tau), K], \tag{6.8}$$

this time generated by the dynamical variable

$$K = \int d^3x\, k^{\alpha}(X(x))P_{\alpha}(x). \tag{6.9}$$

The parametrized canonical action functional (2.51) changes under the canonical transformations (6.9) at the rate

$$\frac{\partial S}{\partial\tau} = [S, K]. \tag{6.10}$$

However, $\int d^3x(P_{\alpha}\dot{X}^{\alpha} + \pi_A\dot{\phi}^A)$ is a canonical invariant and the multipliers $N(x)$, $N^{a}(x)$ do not depend on the canonical variables. Moreover, K is invariant under spatial transformations, Eq. (5.21). As a result,

$$\frac{\partial S}{\partial\tau} = \int dt \int d^3x\, N(x)[K, H(x)]. \tag{6.11}$$

Equation (6.11) holds for an arbitrary vector field $k^{\alpha}(X)$. When this field happens to be a Killing vector field of the metric $g_{\alpha\beta}(X)$, the Poisson bracket $[K, H(x)]$ vanishes by virtue of Eq. (5.19) and the action S stays unchanged. Therefore, if the background has a Killing vector, the parametrized canonical action remains invariant under a one-parameter group of canonical transformations (6.8), (6.9) induced by the diffeomorphisms (6.2) of the spacetime background. During this transformation, the field variables ϕ^A,π_A and the multipliers N, N^a are kept fixed. The fixation of multipliers is consistent with their intended meaning, because the lapse function and the shift vector remain unchanged if the hypersurfaces which they connect are dragged along a Killing vector field. In this way, the unconditional symmetry (5.9) is linked with the invariance of the action.

A more interesting situation arises when the background has a conformal Killing vector (5.5) while the field has a trace-free energy–momentum tensor, Eq. (5.4). Under such circumstances, Eq. (5.22) implies that the action is only conditionally invariant:

$$\frac{\partial S}{\partial\tau} = \int dt \int d^3x\, \tfrac{1}{2} NAH \approx 0. \tag{6.12}$$

We can paraphrase this fact by saying that the canonical transformation does not effect the action if we take into account an equation obtained by varying the action with respect to a multiplier, $N(x)$, which itself is not a canonical variable.

We can reinterpret this invariance as an unconditional invariance if, together with the canonical transformation (6.8)–(6.9) along a conformal Killing field, we keep rescaling the lapse function:

$$\frac{\partial N(x,\tau)}{\partial\tau} = \tfrac{1}{2} AN. \tag{6.13}$$

Under the extended transformation (6.8)–(6.9) and (6.13), the parametrized action behaves as an unconditional invariant,

$$\frac{\partial S}{\partial\tau} = -\int dt \int d^3x\Big(\frac{\partial N}{\partial\tau} H + N\frac{\partial H}{\partial\tau}\Big) = 0. \tag{6.14}$$

The origin of the extended invariance becomes obvious when we recall that the tracelessness of the energy–momentum tensor on which the whole argument is based follows from the invariance of the field action under conformal transformations of the spacetime metric,

$$g_{\alpha\beta}(X) \rightarrow e^{\lambda(X)}g_{\alpha\beta}(X). \tag{6.15}$$

We shall first restate this result in terms of the parametrized action. The conformal transformation (6.15) induces not only the scaling of the spatial metric

$$g_{ab}(x) \rightarrow e^{\lambda(x)}g_{ab}(x), \quad \lambda(x) \equiv \lambda(X(x)), \tag{6.16}$$

but also that of the unit normal,

$$n^{\alpha}(x) \rightarrow e^{-(1/2)\lambda(x)}n^{\alpha}(x), \quad n_{\alpha}(x) \rightarrow e^{(1/2)\lambda(x)}n_{\alpha}(x). \tag{6.17}$$

We want to identify the multiplier N with the lapse function $N = \dot{X}^{\alpha}n_{\alpha}$ and so we should prescribe its scaling behavior as

$$N(x) \rightarrow e^{(1/2)\lambda(x)}N(x). \tag{6.18}$$

The infinitesimal version of Eqs. (6.16)–(6.18) is

$$\delta g_{ab}(x) = \Lambda(x)g_{ab}(x), \tag{6.19}$$

$$\delta n^{\alpha}(x) = -\tfrac{1}{2}\Lambda(x)n^{\alpha}(x), \quad \delta n_{\alpha}(x) = \tfrac{1}{2}\Lambda(x)n_{\alpha}(x), \tag{6.20}$$

$$\delta N(x) = \tfrac{1}{2}\Lambda(x)N(x), \tag{6.21}$$

where $\Lambda(x) \equiv \delta\lambda(x)$. We immediately see that tracelessness of the energy–momentum tensor follows from the invariance of the parametrized action under the scaling (6.19)–(6.21):

$$0=\delta S = -\int dt \int d^3x[N\delta H + H\delta N]$$
$$= -\int dt \int d^3x\Big[N\Big(P_\alpha \delta n^\alpha + \frac{\partial H^\phi}{\partial g_{ab}}\delta g_{ab}\Big) + H\delta N\Big]$$
$$= \int dt \int d^3x\, \tfrac{1}{2} N\Lambda\, (P + g^{1/2}T^{ab}g_{ab} - H)$$
$$= \int dt \int d^3x\, \tfrac{1}{2} N\Lambda\, g^{1/2}(T^a_a - T_{\perp\perp}). \tag{6.22}$$

At this point, we can deduce the conditional symmetry (5.22) from the scaling invariance of the parametrized action. Indeed, if the background has a conformal Killing vector, the scaling can be achieved by going along its flowlines. Then,

$$\delta H = \frac{\partial H}{\partial \tau}\cdot \partial \tau = [H, K\,]\partial\tau,$$
$$\delta N = \tfrac{1}{2}\Lambda N \partial\tau, \tag{6.23}$$

and the scaling invariance (6.22) of the action takes the form

$$0 = \frac{\delta S}{\delta \tau} = -\int dt \int d^3x\, N([H, K\,] + \tfrac{1}{2}\Lambda H). \tag{6.24}$$

Because N is arbitrary, Eq. (5.22) follows. We can state this connection as a theorem:

Theorem: If the parametrized action is invariant under the scaling transformation (6.19)–(6.21) and the background has a conformal Killing vector (5.5), the action is invariant under the extended transformation (6.8), (6.9), and (6.13). The dynamical variable (5.9) then generates a conditional symmetry, Eq. (5.22).

The conformal transformation (6.15) leaves the field variables unscaled. One can wonder what happens when a simultaneous scaling of the metric and of the field variables is needed to keep the action invariant. A typical example is a scalar field conformally coupled to the background.[17] However, the conformal coupling is derivative and as such it falls outside the framework of our discussion. There does not seem to be an example of the simultaneous scale invariance for a nonderivatively coupled field. However, if we restrict ourselves to a constant scaling, there is an interesting case to be considered. It is the minimally coupled massless Klein–Gordon field.

It is obvious that the Lagrangian $L = -\frac{1}{2}|{}^4g|^{1/2}g^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta}$ of this field stays unchanged if we scale both the metric $g_{\alpha\beta}$ and the field ϕ by a constant factor,

$$g_{\alpha\beta}(X\,) \rightarrow e^{\lambda}g_{\alpha\beta}(X\,), \quad \phi\,(X\,) \rightarrow e^{-(1/2)\lambda}\phi\,(X\,), \quad \lambda = \text{const.} \tag{6.25}$$

Let us see what this scaling means for the projected variables. Of course, Eqs. (6.16)–(6.18) still hold for λ = const, but they must now be complemented by the scaling

$$\phi\,(x) \rightarrow e^{-(1/2)\lambda}\phi\,(x), \quad \pi(x) \rightarrow e^{(1/2)\lambda}\pi(x) \tag{6.26}$$

of the field variables. The scaling of the momentum $\pi(x)$ is consistent with the Hamilton equation $\pi = -g^{1/2}\phi_{,\perp}$. An infinitesimal version of Eq. (6.26) is

$$\delta\phi\,(x) = -\tfrac{1}{2}\Lambda\phi\,(x), \quad \delta\pi(x) = \tfrac{1}{2}\Lambda\pi(x), \tag{6.27}$$

where $\Lambda \equiv \delta\lambda$ is a constant. The transformation (6.26) or (6.27) is a canonical transformation of the field variables $\{\phi,\pi\}$ generated by the functional

$$K_\pi = -\tfrac{1}{2}\Lambda \int d^3x\, \phi\,(x)\pi(x). \tag{6.28}$$

It is easy to check that the parametrized action (2.51) of the scalar field is invariant under the constant scaling (6.16)–(6.18), (6.26). In particular, the constraint functions (2.34) and (2.30) scale as

$$H \rightarrow e^{-(1/2)\lambda}H, \quad H_a \rightarrow H_a, \tag{6.29}$$

and so the scaling (6.18) of N exactly compensates the scaling (6.29) of H.

The existence of the conditional symmetry (5.6) is a direct consequence of this scale invariance. If the background admits a homothetic motion k^α, the scaling (6.16)–(6.18) can be achieved by going along its flowlines. The scaling of the hypersurface variables is then generated by the functional

$$K_P = \int d^3x\, k^\alpha(X(x))P_\alpha(x), \tag{6.30}$$

the scaling of the field variables by the functional (6.28), and the scaling of all canonical variables together by the functional (5.26). The scaling of the lapse function is externally prescribed by Eq. (6.13) with a constant Λ. Because the parametrized action remains unchanged under this combined transformation, Eq. (6.24) again holds, now for the K given by Eq. (5.26) and for a constant Λ. We thus see that the conditional symmetry (5.26) for a massless scalar field propagating on a background which admits a homothetic motion follows from the invariance of the parametrized field action under the combined (constant) scaling (6.16)–(6.18) and (6.26) of the hypersurface variables, field variables, and the lapse function. This line of argument clearly shows that the mixed character of the generator comes from the necessity to scale both the field variables and the metric in order to keep the action invariant.

7. INTERNAL SYMMETRY AS A CONDITIONAL SYMMETRY

We have just seen how conditional symmetries follow from an invariance of the parametrized action. The transformations we have considered were induced by spacetime diffeomorphisms and thus necessarily effected the hypersurface variables. The generators of such transformations were consequently constructed from these variables, Eq. (6.9).

Another important class of transformations which leaves the action invariant does not act on the hypersurface variables at all, but effects only the variables in the field fibers. These transformations correspond to internal symmetries of the field. In the parametrized canonical formalism, the generators of such transformations can again be interpreted as generators of conditional symmetries. Because the transformations effect only the field variables, their generators are π-restricted.

Internal symmetries occur when a Lie group G of transformations acting on a collection of fields leaves the field action invariant. By Noether's theorem, any one-parameter subgroup of G leads to a conserved current. In the canonical formalism, the projections ϕ^A of the fields are identified with field coordinates. The one-parameter subgroup induces a one-parameter group of transformations

$$\phi^{A'} = \phi^{A'}(\phi^B, \tau) \tag{7.1}$$

on ϕ^A's. Equation (7.1) does not contain any reference to the hypersurface variables. Its infinitesimal version, analogous to Eq. (6.2), is

$$\frac{d\phi^A}{d\tau} = h^A(\phi^B). \tag{7.2}$$

Typically, Eq. (7.1) is a representation rather than a realization of the group. In this case, the coefficients $h^A(\phi^B)$ in Eq. (7.2) are linear functions of ϕ^B.

The canonical momenta π_A conjugate to the canonical coordinates ϕ^A are pulled back by the transformation (7.1) and so

$$\frac{d\pi_A}{d\tau} = -h^B{}_{,A}\pi_B. \tag{7.3}$$

Equations (7.2) and (7.3) define a one-parameter group of canonical transformations generated by the dynamical variable

$$K = \int d^3x\, h^A(x)(\phi^B)\pi_A(x). \tag{7.4}$$

If the field action is left invariant by the group G, the parametrized action (2.51) is left invariant by the canonical transformations (7.2), (7.3):

$$\frac{\partial S}{\partial \tau} = -\int dt \int d^3x\, N\,[H(x), K\,] = 0. \tag{7.5}$$

The generator (7.4) is a spatial invariant and so we do not need to worry about the term $[H_\alpha(x), K\,] = 0$. Because K does not contain any hypersurface variables, we can put $[H(x), K\,] = [H^\phi(x), K\,]$ in Eq. (7.5). Because everything in Eq. (7.5) is expressed exclusively in terms of the field variables, the super-Hamiltonian and supermomentum constraints cannot help us and Eq. (7.5) must be a strong equation. (We assume that there are no other constraints, like those encountered in gauge theories.) The invariance (7.5) of the action then means that K generates an unconditional symmetry

$$[K, H(x)] = [K, H^\phi(x)] = 0. \tag{7.6}$$

As an example, take a charged scalar field $\phi^A = \{\phi, \phi^*\}$, $\pi_A = \{\pi, \pi^*\}$ with the energy density (2.37). This density is invariant under the phase transformations

$$\phi' = e^{-(1/2)i\eta\tau}\phi, \quad \pi' = e^{(1/2)i\eta\tau}\pi, \quad \text{and c.c.} \tag{7.7}$$

In this case,

$$\dot\phi = \tfrac{1}{2}\, i\eta\phi, \quad \dot\pi = -\tfrac{1}{2}\, i\eta\pi, \quad \text{and c.c.,} \tag{7.8}$$

$$h^A(x) = \{\tfrac{1}{2}\, i\eta\phi(x), -\tfrac{1}{2}\, i\eta\phi^*(x)\}, \tag{7.9}$$

and the generator (7.4) reduces to the expression

$$K = \eta \int d^3x\, \tfrac{1}{2}\, i(\phi\pi - \phi^*\pi^*), \tag{7.10}$$

which we have already met in Eq. (2.39).

An even simpler example is provided by the transformation

$$\phi' = \phi - \mu\tau, \quad \mu = \text{const}, \tag{7.11}$$

which leaves invariant the action of a massless Klein–Gordon field. Here, $h(x) = \mu$ and the conserved generator is

$$K = \mu \int d^3x\, \pi(x). \tag{7.12}$$

Let us finally briefly mention the complications encountered in gauge theories. There, the Lie group G acting on a collection of "carrier fields" is turned into an infinitely dimensional gauge group by letting the group parameters depend on position. The action functional of the carrier fields ceases to be invariant under the gauge group. To restore the invariance, one introduces another collection of fields, called "compensating fields," and couples them to the carrier fields by the Yang–Mills algorithm. The gauge invariance of the total action leads to supplementary constraints on the canonical variables. So, in our first example (7.7), η is made position dependent, the compensating field is the electromagnetic potential ϕ_α which transforms in the familiar way

$$\frac{d}{d\tau}\phi_\alpha(x) = \phi_\alpha(x) + \eta_{,\alpha}(x) \tag{7.13}$$

under the gauge transformation, the total action has the electromagnetic field minimally coupled to the charged scalar field, and the supplementary constraint is the divergence equation (2.45).

Complications arise because supplementary constraints can condition symmetries similarly as the super-Hamiltonian and supermomentum constraints do. However, we do not intend to discuss these complications in the present paper. Once again, our main purpose is to understand the lack of conditional symmetry in geometrodynamics by studying a simpler case of parametrized fields. The super-Hamiltonian and supermomentum constraints are common to geometrodynamics and parametrized field theories. On the other hand, there are no supplementary constraints in vacuum geometrodynamics.

8. GENERAL THEOREMS ABOUT CONDITIONAL SYMMETRIES

We have seen how a Killing symmetry of the background or an internal symmetry of the field leads to a conditional symmetry within the canonical formalism. The generator K has the form

$$K \equiv \int d^3x\, k^\alpha(X(x))P_\alpha(x) \tag{8.1}$$

for the Killing symmetry and the form

$$K \equiv \int d^3x\, h^A(\phi(x))\pi_A(x) \tag{8.2}$$

for the internal symmetry. The generator (8.1) does not depend on the field variables ϕ^A, π_A, while the generator (8.2) does not depend on the hypersurface variables X^α, P_α. Moreover, the coefficient $k^\alpha(X(x))$ is not an arbitrary functional of the hypersurface coordinates, but it is a restriction of a spacetime vector field to a spacelike hypersurface.

Let us pose now an inverse problem. Can we prove, for a given field theory, that the generator K necessarily reduces to one of the previous types, or maybe to their superposition, such as we met for the massless Klein–Gordon field? This is by no means obvious, because the generic form of K, Eq. (3.1), allows much more flexibility in the coefficients $k^{\perp}, k_a$, and h^A. Unfortunately, it is difficult to carry the argument for all field theories at once, without relying on the specific features of the field super-Hamiltonian. Various factors, like the presence of additional constraints or coupling of several fields, makes the general argument extremely cumbersome. However, we can proceed quite far on the general level if we limit our attention to fields whose energy density is a (local) quadratic function of the momenta without a linear term,

$$\begin{aligned} H^{\phi}(x) &= T(x) + V(x)[X,\phi\,], \\ T(x) &= \tfrac{1}{2}\, G^{AB}(x)[X,\phi\,]\pi_A(x)\pi_B(x) \\ &\quad + \tfrac{1}{2}\, G^{Aa\,Bb}(x)[X,\phi\,]\pi_{A,a}(x)\pi_{B,b}(x). \end{aligned} \tag{8.3}$$

We assume that the field variables ϕ^A, π_A are unconstrained and that the "supermetric" $\{G^{AB}, G^{Aa\,Bb}\}$ is nondegenerate. The latter requirement means that

$$\begin{aligned} &\int d^3x \frac{\delta T_N}{\delta\pi_A(x)} \psi_A(x)[X,\phi\,] = 0 \\ &\forall N,\pi_A \;\Rightarrow\; \psi_A(x) = 0. \end{aligned} \tag{8.4}$$

The Klein–Gordon field theory or the massive vector field theory which we have discussed in Sec. 2 are of this type.

To see under what conditions the generator (3.1) reduces to a combination of the generators (8.1) and (8.2), let us study the Poisson bracket

$$\begin{aligned} [K, H_N] &= [K_{(k)},P_N] + [K_{(k)},H^{\phi}_N] \\ &\quad + [K_{(h)},P_N] + [K_{(h)},H^{\phi}_N]. \end{aligned} \tag{8.5}$$

With a little bit of caution we can use the results of Sec. 5. Equation (5.10) still holds if we interpret δ_N as the partial derivative $\eth_N$ with respect to the hypersurface coordinates; i.e., $\eth_N\, F(x)[X,\phi\,]$ is the normal change of the variable $F(x)[X,\phi\,]$ obtained by varying X explicitly but not implicitly in $\phi\,[X\,]$. This yields

$$\begin{aligned} [K_{(k)},P_N] &= \int d^3x\{(\eth_N\, k^{\perp}(x) + k^aN_{,a})P \\ &\quad + (\eth_N\, k^a(x) + k^{\perp}N^{,a} - Nk^{\perp,a})P_a\}. \end{aligned} \tag{8.6}$$

Further,

$$\begin{aligned} [K_{(k)}, H^{\phi}_N] &= \int d^3x[K, g_{ab}(x)]\, \frac{\delta H^{\phi}_N}{\delta g_{ab}(x)} \\ &\quad + \int d^3x \int d^3x' \frac{\delta k^{\alpha}(x)[X,\phi\,]}{\delta\phi^A(x')}\, \frac{\delta T^{\phi}_N}{\delta\pi_A(x')} P_{\alpha}(x). \end{aligned} \tag{8.7}$$

Again, the first term can be handled as in Sec. 5, with the result (5.17). The only difference is that the expression $k_{(a;b)}$ cannot be interpreted as a projected covariant derivative of a spacetime vector field, but it must be defined through Eq. (5.16) in terms of $k_{\perp}$ and k_a. This gives

$$\begin{aligned} [K_{(k)}, H^{\phi}_N] &= \frac{1}{2}\int d^3x\, Ng^{1/2}T^{ab}k_{(a;b)} \\ &\quad + \int d^3x \int d^3x' \frac{\delta T_N}{\delta\pi_A(x')}\left(\frac{\delta k^{\perp}(x)}{\delta\phi^A(x')} P(x) \right. \\ &\quad \left. + \frac{\delta k^a(x)}{\delta\phi^A(x')} P_a(x)\right). \end{aligned} \tag{8.8}$$

The third Poisson bracket in Eq. (8.5) is

$$[K_{(h)}, P_N] = \int d^3x\, \eth_N\, h^A(x)\cdot\pi_A(x). \tag{8.9}$$

Finally,

$$\begin{aligned} [K_{(h)}, H^{\phi}_N] &= \int d^3x\{\, [h^A(x), H^{\phi}_N\,]\pi_A(x) \\ &\quad + h^A(x)[\pi_A(x), H^{\phi}_N\,]\,\}. \end{aligned} \tag{8.10}$$

We do not need to calculate it in detail. It is enough to notice that under our assumption (8.3) about the form of the energy density, $[K_{(h)}, H^{\phi}_N\,]$ is a quadratic function of the momenta without a linear term. The same thing is true about the stress tensor T^{ab}, which is connected with H^{ϕ} by Eq. (2.25). The momentum density H^{ϕ}_a is always a linear homogeneous function of the canonical momenta.

Keeping this in mind, we return back to the condition $[K, H_N\,] \approx 0$. This weak equation means that the sum of the terms (8.6), (8.8), (8.9), and (8.10) must vanish for all values of the variables ϕ^A, π_A, X^{α}, and P_{α} which satisfy the constraints. Alternatively, we can solve the constraints for the hypersurface momenta, Eq. (2.54), substitute these expressions into the sum, and maintain that the sum vanishes identically in the variables ϕ^A, π_A, and X^{α}. The only term containing the momenta in the third order comes from the expression (8.8),

$$\int d^3x \int d^3x' \frac{\delta k^{\perp}(x)}{\delta\phi^A(x')}\, \frac{\delta T_N}{\delta\pi_A(x')}\, T(x). \tag{8.11}$$

This term must vanish identically in N and π_A. For nondegenerate supermetrics, this implies

$$\delta k^{\perp}(x)/\delta\phi^A(x') = 0. \tag{8.12}$$

We thus see that the coefficient $k^{\perp}$ can depend only on the hypersurface variables X^{α}, not on the fields ϕ^A.

To complete the argument, it is advantageous to replace the generator K by an equivalent generator with $k^a = 0$ (Sec. 4). We then select those terms in the sum which are proportional to the momenta and require that they vanish,

$$\int d^3x\{(Nk^{\perp,a} - k^{\perp}N^{,a})H^{\phi}_a + \eth_N\, h^A(x)\cdot\pi_a(x)\} = 0. \tag{8.13}$$

Let us analyze this equation first for a single scalar field. Then, $H^{\phi}_a = \pi\phi_{,a}$ and $\eth_N\, h^A(x)\cdot\pi_A(x)$ reduces to $\eth_N\, h\,(x)\cdot\pi(x)$. Because $\pi(x)$ is arbitrary, we get

$$\eth_N\, h\,(x) = -(Nk^{\perp,a} - k^{\perp}N^{,a})\phi_{,a}. \tag{8.14}$$

An $h\,(x)$ which satisfies this equation can contain only two types of terms: (I) those which are linear in $\phi_{,a}$, (II) those which do not depend on X. In other words,

$$h\,(x) = -k^a(x)[X\,]\phi_{,a}(x) + h\,(x)[\phi\,]. \tag{8.15}$$

Substituting this form of $h(x)$ back into Eq. (8.14), we learn that the coefficient $k^a(x)$ is subject to the condition

$$\delta_N k^a(x)[X] = Nk^{\perp,a} - k^{\perp}N^{,a} \equiv N^a. \tag{8.16}$$

Next, pass to a vector field $\phi^A \equiv \{\phi_\perp, \phi_a\}$, $\pi_A = \{\pi^\perp, \pi^a\}$. The supermomentum then consists of two parts, one corresponding to the spatial scalar $\phi_\perp$, the other one to the spatial vector ϕ_a. In the smeared form,

$$H^\phi_{\mathbf{N}} = \int d^3x \{\pi^\perp(x) L_{\mathbf{N}}\phi_\perp(x) + \pi^a(x) L_{\mathbf{N}}\phi_a(x)\}. \tag{8.17}$$

Equation (8.13) then reads

$$H^\phi_{\mathbf{N}} + \int d^3x (\boldsymbol{\delta}_N h_\perp \cdot \pi^\perp + \boldsymbol{\delta}_N h_a \cdot \pi^a) = 0, \tag{8.18}$$

and it splits into two parts,

$$\boldsymbol{\delta}_N h_\perp + L_{\mathbf{N}}\phi_\perp = 0, \tag{8.19}$$

$$\boldsymbol{\delta}_N h_a + L_{\mathbf{N}}\phi_a = 0. \tag{8.20}$$

The first equation is our old equation (8.14) for the scalar field $\phi_\perp$. It implies that

$$h_\perp = -k^a[X]\phi_{\perp,a} + h[\phi_\perp, \phi_a], \tag{8.21}$$

where $k^a[X]$ must satisfy Eq. (8.16),

$$\delta_N k^a[X] = N^a. \tag{8.22}$$

This enables us to rearrange the term

$$L_{\mathbf{N}}\phi_a = L_{\delta_N \mathbf{k}}\phi_a = \boldsymbol{\delta}_N L_{\mathbf{k}}\phi_a \tag{8.23}$$

and rewrite the vectorial equation (8.20) in the form

$$\boldsymbol{\delta}_N(h_a + L_{\mathbf{k}}\phi_a) = 0. \tag{8.24}$$

It is easy to see that Eq. (8.24) has the general solution

$$H_a = -L_{\mathbf{k}}\phi + h_a[\phi_\perp, \phi_a]. \tag{8.25}$$

Equations (8.12), (8.21), and (8.25) show that the generator K for the vector field must have the form

$$K = \int d^3x (k^\perp[X]P(x) - \pi^\perp L_{\mathbf{k}}\phi_\perp - \pi^a L_{\mathbf{k}}\phi_a + h_A(x)[\phi^B]\pi^A(x)). \tag{8.26}$$

Moreover, the Lie derivative terms in Eq. (8.26) yield the smeared momentum $-H_{\mathbf{k}}$. The generator (8.26) is thus weakly equivalent to the generator

$$K = \int d^3x (k^\perp(x)[X]P(x) + k^a(x)[X]P_a(x) + h_A(x)[\phi^B]\pi^A(x)). \tag{8.27}$$

Apparently, our procedure can easily be generalized to arbitrary tensor fields or collections of such fields. Summarizing this part of our argument, we can conclude:

Theorem: In field theories with a quadratic energy density (8.3), any generator K of a conditional symmetry is always weakly equivalent to a generator of the form (8.27), in which the coefficients $k^\perp$, k^a do not depend on the field variables, and the coefficients h_A do not depend on the hypersurface variables. Moreover, $k^\perp$ and k^a are connected by Eq. (8.16).

It is hard to draw further conclusions about mixed generators in an unspecified field theory. However, we are able to proceed either if we limit our attention to the P-restricted generators, or if we fix the structure of the field theory. We shall follow these two routes in succession. First, we derive some general results for the P-restricted generators. Second, as a characteristic example, we show what is the most general mixed generator in the Klein–Gordon field theory.

For a P-restricted generator, the coefficient h_A in Eq. (8.26) must vanish. We collect the surviving terms in the Poisson brackets (8.6) and (8.8), eliminating again the hypersurface momenta through the constraints (2.54). We end with the requirement

$$[K_{(k)}, H_N] \approx \int d^3x\, g^{1/2} \cdot \{-(\delta_N k^\perp(x) + k^a N_{,a})T_{\perp\perp} + (\delta_N k^a(x) + k^\perp N^{,a} - Nk^{\perp,a})T_{\perp a} + \tfrac{1}{2} Nk^{(a;b)}T_{ab}\} = 0 \tag{8.28}$$

on the restricted generator.

We say that the field is generic if the projections $T_{\perp\perp}(x)$, $T_{\perp a}(x)$, and $T_{ab}(x)$ of its energy–momentum tensor can be arbitrarily varied along a hypersurface. For example, both the massive and the massless Klein–Gordon fields are generic in this sense because, by assigning the canonical variables $\phi(x)$, $\pi(x)$ arbitrarily along a hypersurface, we can give the projections (2.34), (2.35), (2.30) any chosen values. For a generic field, Eq. (8.28) implies that the coefficients of $T_{\perp\perp}$, $T_{\perp a}$, and T_{ab} must vanish independently. The first two equations,

$$\delta_N k^\perp = -k^a N_{,a}, \tag{8.29}$$

$$\delta_N k^a = -k^\perp N^{,a} + Nk^{\perp,a}, \tag{8.30}$$

show that the coefficients $k^\perp$ and k^a behave as projections of a spacetime vector field under hypersurface tilts. We have already chosen $k^\perp$ as a spatial scalar and k^a as a spatial vector and we are thus able to conclude, by the reconstruction theorem of Sec. 2, that $k^\alpha = k^\perp n^\alpha + k^a X^\alpha_a$ is the restriction of a spacetime vector field to the hypersurface:

$$k^\alpha(x)[X] = k^\alpha(X(x)). \tag{8.31}$$

Taking into account not only the hypersurface tilts, but also hypersurface translations, we can now relate the normal changes $\delta_N k^\perp$ and $\delta_N k^a$ to the projections $k^{\alpha;\beta}$ by Eqs. (2.16) and (2.17). On the other hand, we know that these changes must be given by Eqs. (8.29) and (8.30). This allows us to conclude that

$$k_{\perp;\perp} = 0 = k_{(\perp;a)}. \tag{8.32}$$

Also, once we know that $k_\perp$ and k_a are projections of a spacetime vector field (8.31), we are able to interpret $k_{(a;b)}$, which up to now was a mere abbreviation for the expression (5.16), as the ab projection of the covariant derivative $k_{(\alpha;\beta)}$. The remaining equation,

$$k_{(a;b)} = 0, \tag{8.33}$$

following from Eq. (8.28) by varying T^{ab}, ensures that this projection vanishes. Taken together, Eqs. (8.32) and (8.33) tell us that k^α is a Killing vector field. This yields a theorem,

Theorem: A generic field with nonderivative gravitational coupling and a quadratic energy density (8.3) has a P-restricted conditional symmetry if and only if the spacetime background has a Killing vector.

As a matter of fact, we know from Sec. 5 that such a symmetry necessarily reduces to an unconditional symmetry.

The whole argument can be generalized from generic fields to the fields restricted by the trace condition

$$T^{\alpha}_{\alpha} \equiv T_{ab}\, g^{ab} - T_{\perp\perp} = 0. \tag{8.34}$$

Let Eq. (8.34) be the only restriction on the energy–momentum tensor. Then, $T_{\perp\perp}$, $T_{\perp a}$, and the trace-free part $\theta_{ab} \equiv T_{ab} - \frac{1}{3}\, T_{cd}\, g^{cd} g_{ab}$ of the stress tensor can be assigned arbitrarily, while the trace $T_{cd}\, g^{cd}$ is obtained from Eq. (8.34). Varying $T_{\perp a}$, we recover our old equation (8.30). The coefficient of θ_{ab} in Eq. (8.28) must be proportional to the metric, which leads to the equation

$$k^{(a;b)} = \Lambda\, g^{ab} \quad (\text{with } \Lambda = \tfrac{2}{3}\, k^{c}{}_{;c}). \tag{8.35}$$

The variation of $T_{\perp\perp}$ then yields

$$\delta_N\, k^{\perp} + k^{a} N_{,a} - \tfrac{1}{2} \Lambda N = 0. \tag{8.36}$$

Once more, Eqs. (8.30) and (8.36) imply that $k^{\perp}$ and k^{a} have the correct behavior under hypersurface tilts. We can thus conclude that they are actually projections of a spacetime vector field (8.31) and that $k_{(a;b)}$ is the ab projection of the covariant derivative $k_{(\alpha;\beta)}$. Comparing Eqs. (8.30) and (8.36) with Eqs. (2.16) and (2.17), we see that

$$\begin{aligned} &k^{(a;b)} = \Lambda\, g^{ab}, \\ &k_{(a;\perp)} = 0, \\ &k_{\perp;\perp} = -\tfrac{1}{2} \Lambda. \end{aligned} \tag{8.37}$$

Equations (8.37) are all possible projections of the conformal Killing vector field equation

$$k_{(\alpha;\beta)} = \Lambda\, g_{\alpha\beta}. \tag{8.38}$$

We have thus arrived at the following:

Theorem: Let the energy–momentum tensor of a field with nonderivative gravitational coupling and a quadratic energy density (8.3) be freely specifiable on a spacelike hypersurface, except for the trace condition (8.34). Then, the theory has a P-restricted conditional symmetry if and only if the background has a conformal Killing vector (8.38).

This time we know that this symmetry is truly conditional and cannot be reduced to an unconditional symmetry.

9. CONDITIONAL SYMMETRIES IN THE KLEIN–GORDON FIELD THEORY

The theorems of the last section deal with the P-restricted generators. Let us now follow the second route, leaving the generators unrestricted, but fixing the structure of the field theory. We choose the simplest possible model, namely, a single real Klein–Gordon field. Our aim is to enumerate all possible conditional symmetries such a theory can possess.

The energy density (2.34) of the Klein–Gordon field is of the type (8.3). The Poisson bracket $[K_{(k)}, H_N]$ is thus given by the expression (8.28). In fact, our general theorem for mixed generators shows that the coefficient of $T_{\perp a}$ must vanish, Eq. (8.16). We can thus express the Poisson bracket $[K_{(k)}, H_N]$ explicitly as a function of the field variables by substituting into Eq. (8.28) the expressions (2.34) and (2.35) for the energy density $T_{\perp\perp}$ and the stress tensor T_{ab} of the Klein–Gordon field.

The mixed generator brings in two additional Poisson brackets, Eqs. (8.9) and (8.10). However, again by our general theorem, the coefficients $h^{A}(x)$ cannot depend on X^{α} and so the Poisson bracket (8.9) vanishes. For the Klein–Gordon field, the Poisson bracket (8.10) takes on the form

$$\begin{aligned}[K_{(h)}, H^{\phi}_{N}] = &\int d^3x \int d^3x'\, B(x,x')\pi(x)\pi(x') \\ &- \int d^3x \int d^3x' \,\frac{N(x')h(x)\delta V(x')}{\delta\phi(x)},\end{aligned} \tag{9.1}$$

where $B(x, x')$ denotes the symmetric biscalar

$$\begin{aligned} B(x,x') = \tfrac{1}{2} N(x')\, g^{-1/2}(x')\delta h(x)/\delta\phi(x') \\ + (x \leftrightarrow x'). \end{aligned} \tag{9.2}$$

Putting the terms (8.28) and (9.1) together, we end with the conservation condition

$$\begin{aligned}[K, H_N] \approx &- \int d^3x\, A(x)\pi^2(x) \\ &+ \int d^3x \int d^3x'\, B(x,x')\pi(x)\pi(x') \\ &+ \int d^3x [-(\delta_N\, k^{\perp} + k^{a} N_{,a} + N k^{c}{}_{;c})V \\ &+ N g^{1/2} k^{a;b} \phi_{,a}\phi_{,b}] \\ &- \int d^3x \int d^3x'\, \frac{N(x')h(x)\delta V(x')}{\delta\phi(x)} = 0\end{aligned} \tag{9.3}$$

on the mixed generator. Here,

$$A(x) \equiv \tfrac{1}{2} g^{-1/2}(\delta_N\, k^{\perp} + k^{a} N_{,a} - N k^{c}{}_{;c}). \tag{9.4}$$

The main complication introduced by the mixed generator is the presence of the biscalar term (9.2) in Eq. (9.3). Our first step is to determine the biscalar and through it the form of the functional $h(x)[\phi]$.

The quadratic term and the absolute term in Eq. (9.3) must vanish separately. Substituting $B(x, x') = A(x)\delta(x, x') + C(x, x')$ into the quadratic term, we learn that

$$\int d^3x \int d^3x'\, C(x,x')\pi(x)\pi(x') = 0. \tag{9.5}$$

As a consequence, $C(x, x')$ must vanish and

$$B(x,x') = A(x)\delta(x', x). \tag{9.6}$$

To see what are the consequences of Eq. (9.6), replace h by a new variable

$$\bar{h} \equiv h - \frac{A}{N} g^{1/2}\phi. \tag{9.7}$$

Equation (9.6) then reads

$$\begin{aligned} N(x')\, g^{-1/2}(x')\delta\bar{h}(x)/\delta\phi(x') \\ + N(x)\, g^{-1/2}(x)\delta\bar{h}(x')/\delta\phi(x) = 0. \end{aligned} \tag{9.8}$$

It is true for an arbitrary $N(x)$ if and only if $\bar{h}(x)$ does not depend on $\phi(x')$, $\bar{h}(x) = \bar{h}(x)[X]$. On the other hand, from the last section we know that $h(x)$ can depend only on $\phi(x')$ but not on $X^{\alpha}(x')$. The consistency of Eq. (9.7) thus requires that

$$\frac{A}{N} g^{1/2} = -\tfrac{1}{2}\Lambda = \text{const}, \quad \bar{h} = \mu = \text{const}. \tag{9.9}$$

Recalling the meaning of A, we see that

$$\delta_N k^{\perp} + k^{a} N_{,a} - N(k^{c}{}_{;c} - \Lambda) = 0. \tag{9.10}$$

This equation, together with Eq. (8.19), implies that the coefficients $k^{\perp}$ and k^{a} behave as projections of a spacetime vector field under hypersurface tilts. Once more we invoke the reconstruction theorem of Sec. 2 and conclude that such a field exists, Eq. (8.31). The comparison of the generic equations (2.16) and (2.17) with the specific equations (8.16) and (9.10) then reveals that

$$k_{\perp;\perp} = -k^{c}{}_{;c} + \Lambda \tag{9.11}$$

and

$$k_{(a;\perp)} = 0. \tag{9.12}$$

At this point, we can also interpret $k_{(a;b)}$ as the ab projection of the spacetime covariant derivative $k_{(\alpha;\beta)}$.

To complete the argument, we turn to the absolute term in Eq. (9.3). Using Eqs. (9.7), (9.9), and (9.10), we reduce it to the form

$$\int d^3x\, N\left[2V(\Lambda - k^{a}{}_{;a}) + \tfrac{1}{3} g^{1/2} k^{a}{}_{;a} \phi_{,b}\phi^{,b} + g^{1/2} k^{a;b}\theta_{ab} - \mu m^2 g^{1/2}\phi\right] = 0, \tag{9.13}$$

where $\theta_{ab} = \phi_{,a}\phi_{,b} - \frac{1}{3}\phi_{,c}\phi^{,c} g_{ab}$ is again the trace-free part of the stress tensor T_{ab}.

Because $N(x)$ is arbitrary, the expression in the square brackets must vanish. We must now distinguish two cases: (I) $m = 0$, (II) $m \neq 0$.

(I) $m = 0$. Equation (9.13) reduces to

$$2(\Lambda - \tfrac{2}{3} k^{c}{}_{;c})V + g^{1/2} k^{a;b}\theta_{ab} = 0. \tag{9.14}$$

At each point, $V = \frac{1}{2} g^{1/2} g^{ab}\phi_{,a}\phi_{,b}$ and the trace-free θ_{ab} can be considered as independent variables. Therefore,

$$\Lambda - \tfrac{2}{3} k^{c}{}_{;c} = 0, \quad k_{(a;b)} = \bar{\Lambda}\, g_{ab}. \tag{9.15}$$

Contracting the second equation and comparing it with the first one, we learn that $\bar{\Lambda} = \Lambda$. Equations (9.11), (9.12), and (9.15) then yield all possible projections

$$k_{(a;b)} = \Lambda\, g_{ab}, \quad k_{(a;\perp)} = 0, \quad k_{(\perp;\perp)} = -\Lambda \tag{9.16}$$

of the spacetime equation

$$k_{(\alpha;\beta)} = \Lambda\, g_{\alpha\beta}. \tag{9.17}$$

We know that Λ is a constant and so k^{α} is a homothetic motion. We can summarize our results in a theorem:

Theorem: The massless Klein–Gordon field always has the conditional symmetry

$$K = \mu \int d^3x\, \pi(x). \tag{9.18}$$

Moreover, it has the conditional symmetry

$$K \equiv \int d^3x\left[k^{\alpha}(X(x))P_{\alpha}(x) - \tfrac{1}{2}\Lambda\phi(x)\pi(x)\right] \tag{9.19}$$

if and only if the background admits a homothetic motion k^{α}, Eq. (9.17).

(II) $m \neq 0$. At a given point, the quantities V, θ_{ab}, and ϕ can be considered as independent variables. All equations derived in Case (I) still hold, but Eq. (9.13) yields now an additional condition,

$$m^2 g^{1/2}(\tfrac{1}{2}\Lambda\phi^2 + \mu\phi) = 0. \tag{9.20}$$

Because ϕ is arbitrary,

$$\Lambda = 0 = \mu.$$

The generator (9.18) disappears, while the generator (9.19) becomes P-restricted and k^{α} turns into a true Killing vector. Hence:

Theorem: The massive Klein–Gordon field has a conditional symmetry if and only if the background has a Killing vector.

We have thus proved that the Klein–Gordon field does not have any other conditional symmetries except those which we have already found in Secs. 5 and 7.

[1] K. Kuchař, J. Math. Phys. **22**, 2640 (1981).

[2] K. Kuchař, J. Math. Phys. **17**, 777 (1976); **17**, 792 (1976); **17**, 801 (1976); **18**, 1589 (1977). These papers are referred to as I–IV, respectively.

[3] See II, Eq. (2.5).

[4] See II, Eq. (2.6).

[5] See II, Sec. 2.

[6] See II, Eqs. (6.9) and (9.5).

[7] See I, Sec. 10 and II, Sec. 3.

[8] See II, Sec. 9, especially Eq. (9.6).

[9] See, e.g., III, Sec. 3.

[10] See III, Secs. 4 and 5.

[11] See III, Secs. 5 and 6.

[12] See III, Sec. 8, especially Eqs. (8.13), (8.17), and (8.23).

[13] See II, Sec. 5.

[14] See III, Sec. 11C.

[15] P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva U. P., New York, 1964).

[16] See, e.g., III, Sec. 12 and IV, Sec. 4.

[17] Cf., e.g., C. G. Callan, S. Coleman, and R. Jackiw, Ann. Phys. (N.Y.) **59**, 42 (1970).

Time-dependent embeddings for Schwarzschild-like solutions to the gravitational field equations

Christopher F. Chyba[a)]
Department of Physics, Swarthmore College, Swarthmore, Pennsylvania 19081

(Received 21 September 1981; accepted for publication 14 May 1982)

An explicit formula for embedding the Schwarzschild solution in a three-dimensional flat space with indefinite metric for arbitrary Kruskal timelike coordinate v is presented. The time development of the Schwarzschild solution can then be represented by a succession of spacelike surfaces, each corresponding to a different value of v. It is seen that the standard representation of the Schwarzschild metric, the Flamm paraboloid, is in fact the $v = 0$ special case of a similar time-dependent embedding in a three-dimensional Euclidean space with positive definite metric. However, this embedding is inadequate in that it is not defined for most values of v. Thus, the embedding in a space with indefinite metric is to be preferred. The results for the Schwarzschild case are found to be readily extended to all metrics of a certain class, and a general embedding formula for arbitrary v results. Embeddings for the Schwarzschild, de Sitter, and Reissner–Nordström metrics are then special cases of this general form. It is seen that all such solutions behave similarly as v gets large. This suggests an alternate interpretation of the oscillatory character of the Reissner–Nordström "wormhole."

PACS numbers: 04.50. + h

I. INTRODUCTION

The Schwarzschild line element for a body of mass m (Ref. 1) is given by

$$ds^2 = -\Phi\, dt^2 + \Phi^{-1}\, dr^2 + r^2\, d\Omega^2, \tag{1}$$

where

$$\Phi = 1 - 2m/r \tag{2}$$

and

$$d\Omega^2 = d\theta^2 + \sin^2\theta\, d\phi^2 \tag{3}$$

is the metric of a unit sphere. Various methods have been employed to visualize the geometry of spacetime which arises from this solution. One aproach has been to embed the entire four-dimensional manifold in a flat space of higher dimension. Kasner[2] has shown that, excluding the trivial pseudo-Euclidean case, no four-dimensional manifold satisfying $R_{\mu\nu} = 0$ can be embedded in a five-dimensional flat space. However, Kasner,[3] and later Fronsdal,[4] have embedded (1) in a six-dimensional space. The geometry of the 4-manifold can then be pictured by taking subspaces of the higher-dimensional flat space.

A simpler approach is that first used by Flamm,[5] which takes advantage of the spherical symmetry of the Schwarzschild solution. Taking a constant-time slice of the $\theta = \pi/2$ plane yields the two-dimensional line element

$$ds^2 = \Phi^{-1}\, dr^2 + r^2\, d\phi^2, \tag{4}$$

which is then embedded by equating it to the metric of a three-dimensional Euclidean space[6] (positive definite metric):

$$ds^2 = dz^2 + dr^2 + r^2 d\phi^2. \tag{5}$$

Solving for dz^2 gives

$$dz^2 = (\Phi^{-1} - 1)\, dr^2, \tag{6}$$

which upon integration yields the well-known two-sheeted Flamm paraboloid:

$$z(r) = [8m(r - 2m)]^{1/2}. \tag{7}$$

This equation corresponds to a surface with the topology of an Einstein–Rosen bridge,[7] or "wormhole," connecting two asymptotically flat universes. The "throat" of the bridge has a narrowest region in the $z = 0$ plane, where the two universes join along a circle of circumference $4\pi m$, or, taking into account the θ-coordinate, along a sphere of surface area $16\pi m^2$.

The Reissner–Nordström solution for a body of mass m and electric charge q is given by an expression similar to (1):

$$ds^2 = -\Phi\, dt^2 + \Phi^{-1}\, dr^2 + r^2\, d\Omega^2, \tag{8}$$

where

$$\Phi = 1 - 2m/r + q^2/r^2 \tag{9}$$

and $d\Omega^2$ is as before. An identical procedure to that outlined above, with $m > |q|$, gives the embedding formula[8]:

$$z(r) = \int \left[\frac{1-\Phi}{\Phi}\right]^{1/2} dr = \int \left[\frac{2mr - q^2}{(r - r_+)(r - r_-)}\right]^{1/2} dr, \tag{10}$$

where $r_\pm = m \pm (m^2 - q^2)^{1/2}$.

Both these embeddings suffer from an inability to provide any geometrodynamic information, that is, neither can indicate how the curved space develops in time. Yet both the Schwarzschild and Reissner–Nordström solutions are known to exhibit quite dramatic time evolution. Kruskal diagrams[9] indicate that the Schwarzschild "throat" pinches off in a finite time[10] and the Reissner–Nordström "throat" oscillates between a minimum and maximum circumference of $2\pi r_-$ and $2\pi r_+$.[8]

In this paper, we develop a method for embedding any solution of the form

$$ds^2 = -\Phi\, dt^2 + \Phi^{-1}\, dr^2 + r^2\, d\Omega^2, \quad \Phi = \Phi(r) \tag{11}$$

[a)] Present address: Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England.

0022-2488/82/091662-06$02.50

at an arbitrary, but explicit, Kruskal-like time coordinate v. That is, we are able to portray precisely, rather than merely qualitatively, embeddings which include the effectively *time-dependent* nature of certain black-hole type solutions. The time development of the solution can then be represented as a succession of spacelike surfaces, each surface corresponding to a different value of v. These surfaces are only defined for all v if the flat embedding space is endowed with an indefinite metric. It will be seen that the standard Schwarzschild and Reissner–Nordström embeddings discussed above are actually special cases, at time $v = 0$, of the embeddings which result from a similar procedure in which a flat space with positive definite metric is used. Such an embedding is found to be undefined (becomes imaginary) for most values of v. We suggest it is physically more appropriate, in representing solutions to the field equations, to use embeddings that avoid such behavior.

In Sec. II, we present two methods for obtaining such an embedding for the Schwarzschild metric (1). A succession of surfaces at different v is given, and the $v = 0$ surface is compared to the standard Flamm embedding. In Sec. III, with a slight extension of the general Kruskal-like transformations of Graves and Brill,[8] we generalize one of the methods of Sec. II to any metric of the form (11). In Sec. IV, we consider several special cases of this general form, including the Schwarzschild and Reissner–Nordström metrics. It is seen that all solutions of the form (11) must exhibit similar behavior as v goes to $\pm\infty$. Consideration of the dissimilar time evolutions of the Schwarzschild and Reissner–Nordström solutions, in the light of this result, suggests an alternate view of the oscillatory behavior of the Reissner–Nordström "wormhole." Rather than crediting the pulsation in time to the separate and opposing actions of gravitational pull and Maxwell pressure,[8] it is simpler to take the view that the portrayal of the full manifold which results from solving the equations $R_{\mu\nu} = -8\pi T_{\mu\nu}$ for a spherical mass endowed with charge requires a timelike coordinate v that is itself oscillatory.

II. EMBEDDING THE SCHWARZSCHILD METRIC AT ARBITRARY *v*

The well-known Kruskal transformation[9] giving the maximal analytic extension of the Schwarzschild solution is

$$\begin{Bmatrix} u \\ v \end{Bmatrix} = (1 - r/2m)^{1/2}\exp(r/4m)\begin{Bmatrix} \sinh(t/4m) \\ \cosh(t/4m) \end{Bmatrix} \tag{12}$$

for $r < 2m$, and

$$\begin{Bmatrix} u \\ v \end{Bmatrix} = [(r/2m) - 1]^{1/2}\exp(r/4m)\begin{Bmatrix} \cosh(t/4m) \\ \sinh(t/4m) \end{Bmatrix} \tag{13}$$

for $r > 2m$. The line element, now free of the coordinate ("pseudo") singularity at $r = 2m$, becomes

$$ds^2 = f^2(-dv^2 + du^2) + r^2\,d\Omega^2\,, \tag{14}$$

where

$$f^2 = (32m^3/r)\exp(-r/2m)\,, \tag{15}$$

and $d\Omega^2$ is as before.

Our goal is to embed the $\theta = \pi/2$ plane of (1),

$$ds^2 = -\Phi\,dt^2 + \Phi^{-1}\,dr^2 + r^2 d\phi^2\,, \tag{16}$$

into a flat space with metric given by

$$ds^2 = -dr^2 + dz^2 + r^2\,d\phi^2\,. \tag{17}$$

The choice of this particular metric will be discussed shortly. We eliminate dt^2 from (16) in such a way that the time dependence of the metric remains explicit. This is done by solving Eqs. (12) and (13) for t as a function of $v =$ const, differentiating, and squaring the result.[11] We obtain

$$dt^2 = \frac{v^2(r/2m)\,dr^2}{[v^2 - (1 - r/2m)\exp(r/2m)]\,[1 - r/2m]^2} \tag{18}$$

for r both greater and less than $2m$. Equation (16) then becomes

$$ds^2 = \left[\frac{(r/2m)\exp(r/2m)}{v^2 - (1 - r/2m)\exp(r/2m)}\right]dr^2 + r^2\,d\phi^2. \tag{19}$$

Equating Eqs. (19) and (17) gives the embedding formula:

$$dz = \left[\frac{(r/2m)\exp(r/2m)}{v^2 - (1 - r/2m)\exp(r/2m)} + 1\right]^{1/2} dr\,. \tag{20}$$

The same equation, with a useful intermediate result, is more easily obtained by setting $v =$ const. in Eq. (14). Equations (12) and (13) give

$$u^2 - v^2 = -(1 - r/2m)\exp(r/2m) \tag{21}$$

or

$$u = [v^2 - (1 - r/2m)\exp(r/2m)]^{1/2} \tag{22}$$

for all r. We therefore have the requirement that

$$v^2\exp(-r/2m) \geqslant (1 - r/2m)\,. \tag{23}$$

This inequality, which is independent of the signature of the space in which we embed our metric, is a particulary compact representation of the time evolution of the Schwarzschild solution, as shown in Fig. 1.

Differentiating (22) with $v =$ const, we obtain

$$du = (r/8m^2)\exp(r/2m)[v^2 - (1 - r/2m) \times\exp(r/2m)]^{-1/2}\,dr\,. \tag{24}$$

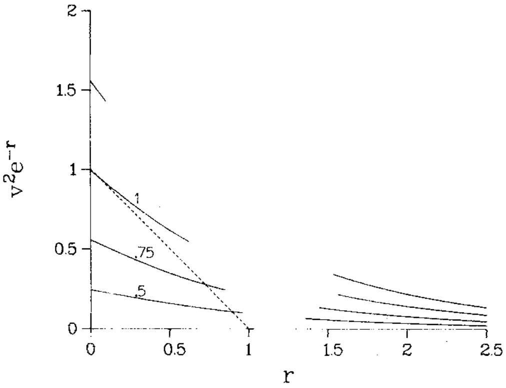


FIG. 1. The inequality $v^2e^{-r}{>}1 - r$ (we have set $2m = 1$), a necessary condition for the Schwarzschild solution in Kruskal coordinates to be embedded for arbitrary v, is a particularly simple representation of the solution's development in time. The embedding is defined only when v^2e^{-r} (solid lines) is greater than $1 - r$ (dashed line). The number attached to each curve indicates the corresponding value of $|v|$. At $|v| = 0$, the "throat" has minimum radius 1; as $|v|$ increases, increasingly smaller values of r are allowed (the "throat" contracts). Finally, at $|v| = \pm 1$, r can equal zero (the "throat" pinches off).

Substituting this expression into (14) and equating to (17) yields the embedding formula (20) immediately.

Had we used the positive definite metric (5) for our flat embedding space, rather than the indefinite metric (17), we would have obtained

$$dz = \left[\frac{(r/2m)\exp(r/2m)}{v^2 - (1 - r/2m)\exp(r/2m)} - 1\right]^{1/2} dr \tag{25}$$

as our embedding formula. At the Kruskal time $v = 0$, this reduces to

$$dz = \left[\frac{(r/2m)}{(r/2m) - 1} - 1\right]^{1/2} dr = (\Phi^{-1} - 1)^{1/2}\, dr, \tag{26}$$

which is just the Flamm embedding (7). We therefore see that the Flamm paraboloid is a special case of the time-dependent embedding (25). However, it is clear that the square root in this equation becomes imaginary for many realizable values of r and v. First write Eq. (25) in the form

$$dz = \left[\frac{\exp(r/2m) - v^2}{v^2 - (1 - r/2m)\exp(r/2m)}\right]^{1/2} dr\,. \tag{27}$$

Equation (23) guarantees that the denominator of this expression is positive. Equation (27) will therefore be undefined (have imaginary square root) whenever

$$v^2 > \exp(r/2m)\,. \tag{28}$$

Such a result is unsatisfactory; we expect a physically acceptable representation of our curved space to be well defined for all-time v. This suggests that (20) is a more appropriate choice than (25), which in turn indicates that the $v = 0$ special case of (20) is a more appropriate embedding than the Flamm paraboloid.

This result is not surprising. We should expect the Schwarzschild line element to require a space of indefinite metric to be embedded for all v. In order to embed an n-dimensional surface given by

$$ds^2 = \sum_{u,v=0}^{n-1} g_{\mu\nu}\, dx_\mu\, dx_\nu \tag{29}$$

in an m-dimensional flat space of arbitrary signature, with metric

$$ds^2 = \sum_{i=0}^{m-1} \alpha_i\, df_i^2\,, \tag{30}$$

where $f_i = f_i(x)$ and $\alpha_i = \pm 1$, we must have

$$\begin{aligned} ds^2 &= \sum_{u,v=0}^{n-1} g_{\mu\nu}\, dx_\mu\, dx_\nu = \sum_{i=0}^{m-1} \alpha_i\, df_i^2 \\ &= \sum_{i=0}^{m-1} \sum_{u,v=0}^{n-1} \alpha_i \frac{\partial f_i}{\partial x_\mu} \frac{\partial f_i}{\partial x_\nu}\, dx_\mu\, dx_\nu\,, \end{aligned} \tag{31}$$

whence,

$$g_{\mu\nu} = \sum_{i=0}^{m-1} \alpha_i \frac{\partial f_i}{\partial x_\mu} \frac{\partial f_i}{\partial x_\nu}\,. \tag{32}$$

Symmetry of the metric tensor $g_{\mu\nu}$ in this equation gives $\frac{1}{2}n(n+1)$ first-order partial differential equations in the m unknowns $f_i(x)$. If there are no inconsistencies in the equations, we have the standard result that any n-dimensional manifold can always be embedded in a flat space of dimension $m \geqslant \frac{1}{2}n(n+1)$.[12] In the case of the Schwarzschild metric, we have $g_{00} = -1/g_{11}$ and the equations are not consistent. Equation (32) yields

$$g_{00} = -\Phi = \sum_{i=0}^{m-1} \alpha_i \frac{\partial f_i}{\partial x_0} \frac{\partial f_i}{\partial x_0}, \tag{33}$$

and

$$g_{11} = \Phi^{-1} = \sum_{i=0}^{m-1} \alpha_i \frac{\partial f_i}{\partial x_1} \frac{\partial f_i}{\partial x_1}, \tag{34}$$

which give

$$\sum_{i=0}^{m-1} \alpha_i \left(\frac{\partial f_i}{\partial x_0}\right)^2 = -\left[\sum_{i=0}^{m-1} \alpha_i \left(\frac{\partial f_i}{\partial x_1}\right)^2\right]^{-1}\,. \tag{35}$$

However, given a positive definite metric ($\alpha_i = 1$ for all i),

$$\sum_{i=0}^{m-1} \alpha_i \left(\frac{\partial f_i}{\partial x_\nu}\right)^2 = \sum_{i=0}^{m-1} \left(\frac{\partial f_i}{\partial x_\nu}\right)^2 \geqslant 0 \tag{36}$$

for any ν. Equation (35) therefore shows the impossibility of embedding the entire Schwarzschild manifold in a positive definite Euclidean space. The case $v = 0$ is, of course, an exception to this result. If $v = 0$, then $r > 2m$, and (13) shows that $t = 0$ identically for any allowable r. The Schwarzschild metric is then no longer indefinite (since $dt = 0$), and for this special case the entire manifold can thus be embedded.[13]

To show the time evolution of the Schwarzschild solution using embedding diagrams, we choose different constant values of v in (20). For a given v, the equation can then be integrated numerically to give a spacelike two-dimensional surface. It is clear from (20) that the time evolution of the manifold is symmetric in v about the value $v = 0$, and that, as v goes to $\pm\infty$, $z(r) = r$. Embeddings for illustrative values of v are shown in Figs. 2 and 3. Of particular interest is the $v = 0$ embedding, corresponding to the maximum size of the Schwarzschild "throat." At $v = 0$, (20) can be integrated exactly to give

$$\begin{aligned} z = \sqrt{2} \int \left[\frac{r - m}{r - 2m}\right]^{1/2} dr &= [2(r - m)(r - 2m)]^{1/2} \\ &+ \frac{\sqrt{2}}{2}\, m \log\left[\frac{(r - 2m)^{1/2} + (r - m)^{1/2}}{(r - m)^{1/2} - (r - 2m)^{1/2}}\right]. \end{aligned} \tag{37}$$

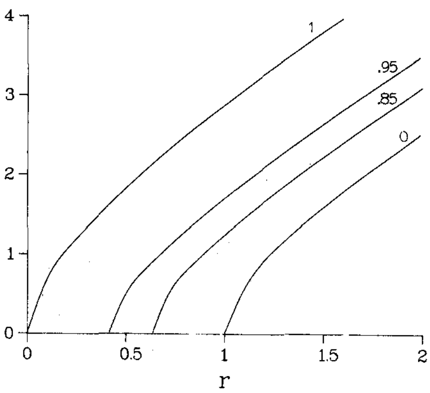


FIG. 2. Equation (20) gives an embedding of the Schwarzschild solution for any Kruskal-time v. Substituting into (20) a constant value of $|v|$, the equation can be numerically integrated to give $z = z(r)$. Here we show the embedding corresponding to $|v| = 0$ (maximum size of "throat"), $|v| = 0.85, 0.95$ ("throat" contracts), and $|v| = 1$ ("throat" pinches off). To obtain the entire two-sheeted embeddings, the curves must be rotated about the z axis, and reflected across the $z = 0$ plane. We have set $2m = 1$.

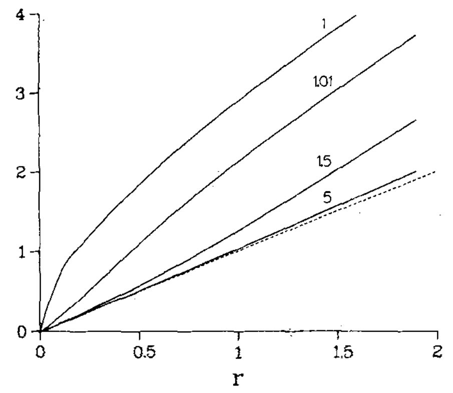


FIG. 3. Identical to Fig. 2, for the cases $|v| = 1, 1.01, 1.5, 5$. The Schwarzschild "throat" approaches the line $z = r$ as $|v|$ grows large.

This new $v = 0$ embedding is compared to the standard Flamm embedding in Fig. 4. It is seen that the behavior of the new embedding is qualitatively similar to that of Flamm: the "throat" has a narrowest radius of $r = 2m$ in the $z = 0$ plane, and the surface is asymptotically flat at large r.

III. THE GENERAL CASE

Graves and Brill[8] have given a general Kruskal-like transformation to remove pseudosingularities from metrics of the form (11), of which the Schwarzschild, de Sitter, and Reissner–Nordström metrics are special cases. It is assumed that $\phi(r)$ has zeroes or poles (the pseudosingularities) which are to be eliminated by transforming r and t to new coordinates $u(r,t)$ and $v(r,t)$, in terms of which light continues to travel along lines of slope ± 1. In such coordinates, the metric (11) takes the form

$$ds^2 = f^2(u,v)(du^2 - dv^2) + r^2(u,v)\, d\Omega^2\,, \tag{38}$$

where

$$f^2(u,v) = \Phi(r)\exp(-2\gamma r^*)/4A^2\gamma^2 \tag{39}$$

and

$$r^* = \int dr/\Phi(r)\,. \tag{40}$$

A is an arbitrary scale factor and γ is a constant chosen so that (39) is regular at the pseudosingularity (if more than one such singularity exists, several coordinate patches may be required). The coordinate transformation itself is given as

$$\begin{Bmatrix} u(r,t) \\ v(r,t) \end{Bmatrix} = 2A\exp(\gamma r^*)\begin{Bmatrix} \cosh(\gamma t) \\ \sinh(\gamma t) \end{Bmatrix} \tag{41}$$

with the inverse transformation given implicitly by

$$u^2 - v^2 = 4A^2\exp(2\gamma r^*)\,, \tag{42}$$

$$t = (1/2\gamma)\tanh^{-1}[2uv/(u^2 - v^2)]\,. \tag{43}$$

Equation (42) gives

$$u = [v^2 + 4A^2\exp(2\gamma r^*)]^{1/2}\,. \tag{44}$$

Differentiating this equation and substituting into (38), with $v = \text{const.}$, yields

$$ds^2 = 4A^2\Phi^{-1}(r)\exp(2\gamma r^*) \times [v^2 + 4A^2\exp(2\gamma r^*)]^{-1}dr^2 + r^2\, d\Omega^2\,. \tag{45}$$

Equating (45) and (17) then gives

$$dz = \left[\frac{4A^2\exp(2\gamma r^*)}{\Phi(r)[v^2 + 4A^2\exp(2\gamma r^*)]} + 1\right]^{1/2} dr\,. \tag{46}$$

We therefore have a general procedure for embedding any metric of the form (11) at arbitrary time v. Finally, we note that (44) provides the general requirement

$$v^2 \geqslant -4A^2\exp(2\gamma r^*)\,. \tag{47}$$

IV. APPLICATIONS

For the Schwarzschild metric, Graves and Brill put

$$\gamma = 1/4m, \quad A = \tfrac{1}{2}, \quad \Phi = (1 - 2m/r),$$
$$r^* = r + 2m\log(r - 2m)\,. \tag{48}$$

These values give the transformation equations

$$\begin{Bmatrix} u \\ v \end{Bmatrix} = (r - 2m)^{1/2}\exp(r/4m)\begin{Bmatrix} \cosh(t/4m) \\ \sinh(t/4m) \end{Bmatrix}. \tag{49}$$

Clearly, however, these equations are not valid when $r < 2m$. We therefore choose $r^* = r + 2m\log|r - 2m|$ in general, and, in addition to (49), take

$$\begin{Bmatrix} u \\ v \end{Bmatrix} = 2A\exp(\gamma r^*)\begin{Bmatrix} \sinh(\gamma t) \\ \cosh(\gamma t) \end{Bmatrix} \tag{50}$$

for the Schwarzschild metric in the case $r < 2m$. The inverse transformation (42)—and hence our embedding formula—now remains unique regardless of the value of r. Finally, to bring our results completely in line with the transformation of Kruskal, we take $A = 1/(8m)^{1/2}$. Substitution of these values into (42) reveals that the Schwarzschild embedding of Sec. II is a special case of the general procedure presented in Sec. III.

As a second example, consider the metric of the de Sitter universe in the static frame.[14] We have

$$\Phi = 1 - r^2/R^2\,, \tag{51}$$

where $0 < r < R$. We restrict our discussion of this metric to the inequality (47), which, with

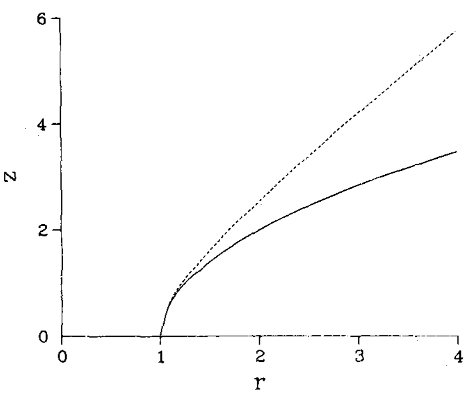


FIG. 4. The well-known Flamm paraboloid (7) is the $|v| = 0$ special case of an arbitrary v embedding into a space with positive definite metric (25), and is given by the solid line. The $|v| = 0$ special case of an embedding in a space of indefinite metric, (37), behaves similarly (dashed line); its minimum radius is 1, and it is asymptotically flat for large r. Both curves are to be rotated about the z axis and reflected through the $z = 0$ plane to give the full two-dimensional embedding.

$$r^* = \frac{R}{2}\log\left(\frac{R+r}{R-r}\right); \quad \gamma = -\frac{1}{R}; A = 1\,, \tag{52}$$

becomes

$$v^2(R+r) \geqslant 4(r-R)\,. \tag{53}$$

This inequality indicates that r cannot become infinite unless $|v| \geqslant 2$, in agreement with the usual result.[15]

Finally, we consider the Reissner–Nordström metric (8), restricting ourselves to the case in which the mass exceeds the value associated by general relativity with the charge

$$m > |q|\,, \tag{54}$$

where both are in units of centimeters. While such a restriction avoids so-called "naked" singularities,[16] the physical significance of this metric remains unclear. Misner and Wheeler[17] have shown the condition (54) to be incompatible with a nonclassical description of charge and mass. In addition, it has recently been shown that a gravitational collapse to the Reissner–Nordström singularity is impossible for a broad class of boundary-surface histories.[18]

With the condition (54), the metric has two pseudosingularities at

$$r_\pm = m \pm (m^2 - q^2)^{1/2}. \tag{55}$$

Two coordinate patches (i,j) are thus required in the neighborhoods of r_+ and r_-. Graves and Brill give

$$r^* = r + \left(\frac{r_+^2}{r_+ - r_-}\right)\log(r - r_+) - \left(\frac{r_-^2}{r_+ - r_-}\right)\log(r - r_-) \tag{56}$$

and

$$\gamma_i = (r_i - r_j)/2r_i^2\,, \tag{57}$$

which yield the transformation

$$\begin{Bmatrix} u_i \\ v_i \end{Bmatrix} = 2A\,(r - r_i)^{1/2}(r - r_j)^{\alpha_j}\exp(\gamma_i\, r)\begin{Bmatrix} \cosh \gamma_i\, t \\ \sinh \gamma_i\, t \end{Bmatrix}, \tag{58}$$

where

$$\alpha_j = -\tfrac{1}{2}(r_j/r_i)^2\,, \tag{59}$$

with $(i,j) = (+,-)$ or $(-,+)$. As in the Schwarzschild case, however, these equations need to be generalized for values of r other than $r > r_+ > r_-$. Our criterion is that the inverse transformation (42) remains unique for each coordinate patch. Thus, our transformations becomes

$$r^* = r + \left(\frac{r_+^2}{r_+ - r_-}\right)\log|r - r_+| - \left(\frac{r_-^2}{r_+ - r_-}\right)\log|r - r_-| \tag{60}$$

and

$$\begin{Bmatrix} u_i \\ v_i \end{Bmatrix} = 2A\,|r - r_i|^{1/2}|r - r_j|^{\alpha_j}\exp(\gamma_i\, r)\beta_i\,, \tag{61}$$

where α_j is as before and

$$\beta_i = \begin{Bmatrix} \cosh \gamma_i\, t \\ \sinh \gamma_i\, t \end{Bmatrix} \quad \text{or} \quad \begin{Bmatrix} \sinh \gamma_i\, t \\ \cosh \gamma_i\, t \end{Bmatrix}, \tag{62}$$

depending on the sign of $|r - r_i||r - r_j|^{2\alpha_j}$ relative to $(r - r_i)(r - r_j)^{2\alpha_j}$.

We can substitute these values into (46) to obtain an embedding for the Reissner–Nordström solution at arbitrary v. In particular, the $v = 0$ embedding

$$z(r) = \int \left[\frac{1 + \Phi(r)}{\Phi(r)}\right]^{1/2} dr \tag{63}$$

differs from the Flamm-like embedding (10), which results from a flat space with positive definite metric.

Rather than utilizing two coordinate patches, we could, at least formally, follow a similar embedding procedure in the extended Reissner–Nordström manifold. Here the metric (8) may be written[19] in the form

$$ds^2 = F^2(-d\psi^2 + d\xi^2) + r^2\, d\Omega^2\,, \tag{64}$$

where

$$F = F(\psi,\xi); \quad r = r(\psi,\xi) \tag{65}$$

and $d\Omega$ is the usual spherical surface element. However, the complicated nature of the transformations (65) indicates that it is in practice simpler to use the series of coordinate patches (i,j) given by Graves and Brill.

Finally, we wish to consider the time development of the manifold. Consider the general embedding equation (46), of which that of the Reissner–Nordström metric is a special case. It is clear from (46) that, as we let the absolute value of v in our constant time embeddings grow large, the equation goes to

$$z(r) = \int dr = r\,, \tag{66}$$

which is pinched off at $r = 0$ in the $z = 0$ plane. In particular, this same behavior holds for the Reissner–Nordström embedding. Yet it is known that the radius of the "throat" for this metric must pulsate periodically in time. This pulsation has been credited to a "cushioning" by Maxwell pressure of the electric field through the "throat."[8] From a consideration of the embedding formula, however, in which the effect of the presence of electric charge is taken into account by the values assigned r^* and γ, it seems the "throat" must pinch off as in the Schwarzschild case. This does not take place because $|v|$ never goes to infinity; for an observer on the "throat" ($u = 0$ in the first patch), v reaches a maximum value of

$$v^2 = 4A^2\exp(2\gamma_+ r_c)(r_+ - r_c)(r_c - r_-)^{2\alpha}\,, \tag{67}$$

where

$$\alpha_- = -\frac{1}{2}\left(\frac{r_-}{r_+}\right)^2, \quad \gamma_+ = \left(\frac{r_+ - r_-}{2r_+^2}\right), \tag{68}$$

and

$$r_+ > r_c > r_-\,. \tag{69}$$

At this value of $r = r_c$, the observer crosses into the second patch. Upon return to a patch identical to the first, the observer moves only between two finite values of v, again departing the patch at a time v given by (67). That is, $|v|$ never approaches infinity, but rather, oscillates between finite values. We adopt the view that the Reissner–Nordström "throat" pulsates because the timelike coordinate needed to describe both patches of the manifold which results from a spherically symmetric mass and charge distribution must itself be oscillatory.

ACKNOWLEDGMENTS

We are very grateful to Dr. John R. Boccio for his consistent encouragement, assistance, and a series of exciting discussions on the subject of this paper. The bulk of this research was conducted at Swarthmore College under a grant from the National Science Foundation (NSF-URP Grant #SPI-7926963).

[1] We use units in which $c = G = 1$. In Gaussian units, these values give the charge of a proton as 1.381×10^{-39} kilometers.

[2] E. Kasner, Am. J. Math. **43**, 126 (1921).

[3] E. Kasner, Am. J. Math. **43**, 130 (1921).

[4] C. Fronsdal, Phys. Rev. **116**, 778 (1959).

[5] L. Flamm, Phys. Z. **17**, 448 (1916); see also H. Weyl, *Space–Time–Matter* (Dover, New York, 1952).

[6] With some modification, Kasner's proof in Ref. 2 is readily generalizable to the impossiblity of embedding a nontrivial solution of $R^{\delta}_{\ \alpha\rho\delta} = 0$ for an n-dimensional space into an $(n + 1)$-dimensional flat space. The embedding of the $\theta = \pi/2, dt = 0$ (i.e., two-dimensional) slice of the Schwarzschild solution into a three-dimensional Euclidean space does not contradict this result, as $ds^2 = \Phi^{-1} dr^2 + d\phi^2$ does not satisfy $R^{\delta}_{\ \alpha\rho\delta} = 0$ in two dimensions.

[7] A. Einstein and N. Rosen, Phys. Rev. **48**, 73 (1935).

[8] J. Graves and D. Brill, Phys. Rev. **120**, 1507 (1960).

[9] M. Kruskal, Phys. Rev. **119**, 1743 (1960).

[10] R. Fuller and J. Wheeler, Phys. Rev. **128**, 919 (1962).

[11] Such an approach was first attempted by J. Aronowitz (unpublished), Swarthmore College, 1979, NSF-URP Grant #SPI-7827548.

[12] T. Levi-Civita, *The Absolute Differential Calculus* (Blackie, London, 1927), p. 122.

[13] We are indebted to Dr. J. R. Boccio for suggesting this possibility.

[14] E. Schrödinger, *Expanding Universes* (Cambridge U.P., Cambridge, 1956).

[15] Graves and Brill (Ref. 8) present a Kruskal-like diagram giving this result. However, they lose a factor of 2 in their application of the general transformation (41) to the de Sitter case (52). Dropping this factor, (53) becomes $v^2(R + r) \geqslant (r - R)$, which gives $|v| \geqslant 1$ as the requirement for $r = \pm \infty$, in agreement with their diagram.

[16] S. Hawking and G. Ellis, *The Large Scale Strucutre of Space-Time* (Cambride U. P., Cambridge, 1973).

[17] C. Misner and J. Wheeler, Ann. Phys. **2**, 525 (1957).

[18] K. Lake and L. Nelson, Phys. Rev. D **22**, 1266 (1980).

[19] B. Carter, Phys. Rev. **141**, 1242 (1966); J. Finley, J. Math. Phys. **15**, 1698 (1974); R. St. John and J. Finley, J. Math. Phys. **15**, 147 (1974).

Spatial topology and Yang–Mills vacua

C. J. Isham and G. Kunstatter[a)]
Blackett Laboratory, Imperial College, London SW7 2BZ, England

(Received 15 May 1981; accepted for publication 25 June 1981)

We study the canonical vacuum structure of Yang–Mills theories defined on an arbitrary (nonsimply connected) three-space. We find that the presence of flat Yang–Mills connections with a nontrivial (discrete) holonomy group has profound consequences at the quantum level. In particular, such connections may lead to either an increase or a decrease in the number of quantum vacuum sectors. Our method consists of finding a representation for the space of classical zero-energy field configurations in terms of a function space $\mathscr{D}$. A simple assumption concerning the physical equivalence of these classical configurations then permits a formal classification of the quantum vacua by the zeroth homotopy set $\pi_0(\mathscr{D})$. Significant progress is made in the analysis of $\pi_0(\mathscr{D})$ for arbitrary three-spaces and gauge groups, and several specific questions concerning the vacuum states and their diagonalization are answered.

PACS numbers: 04.50. + h, 11.10.Np

1. INTRODUCTION

In spite of an ever-increasing level of activity, the quantization of the gravitational field remains an unsolved problem. Currently, much effort is devoted to showing that the $N = 8$ supergravity theory is finite, order by order in perturbation theory.[1] However, there has always been a school of thought which maintains that progress in quantum gravity can only be achieved through a deeper understanding of the underlying conceptual and technical structures. In particular, attention has been focused on the possibility that, at the Planck length $(G\hbar/C^3)^{1/2} \approx 10^{-33}$ cm, the topological properties of space and time may differ greatly from those that are implicit in conventional perturbative quantum gravity. Even if supergravity theory *is* found to be finite, a number of deep and fascinating questions will still exist concerning the role of spacetime topology.

This interest in spacetime topology has inspired many studies of quantum fields propagating on a spacetime manifold M whose metric is fixed and unquantized. The aim is to abstract the effects which are a direct result of the topological properties of M from those arising from the background metric. The present paper lies within this category and contains some initial results of an investigation into the vacuum structure of a canonically quantized Yang–Mills field ω_μ that is defined on an arbitrary, orientable, C^∞ three-manifold Σ. It is technically convenient to assume that Σ is compact—a condition that could arise in practice by imposing vanishing boundary conditions on gauge fields originally defined on a noncompact space. According to the Poincaré conjecture almost all three-manifolds are not simply connected and, as we shall demonstrate, this leads to a vacuum structure which differs significantly from that arising in the conventional flat-space theory where $\mathbb{R}^3$ is compactified to S^3.

We employ the timelike gauge $\omega_0 = 0$ and look for classical Yang–Mills fields $\omega_i(\mathbf{x})(i = 1,2,3)$ which satisfy the zero energy condition $F_{ij} = 0$ (vanishing field strength) with the usual assumption that a quantum vacuum functional $\psi[\omega]$ is peaked around such a configuration.[2] On S^3, the only solutions to $F_{ij} = 0$ are of the form

$$\omega_i(\mathbf{x}) = \Omega(\mathbf{x})\partial_i\Omega(\mathbf{x})^{-1}, \qquad (1.1)$$

where Ω is a gauge function (i.e., a map from S^3 into the symmetry group G). Clearly $\omega_i(x)$ is a gauge transform of the zero Yang–Mills field.

When $\pi_1(\Sigma)\neq 0$ two new features arise which are most clearly seen by adopting the standard mathematical picture of the Yang–Mills field as a connection in a principal G bundle. Now principal G bundles over S^n are classified by $\pi_{n-1}(G)$.[3] However, $\pi_2(G) = 0$ for any Lie group and hence, in the usual flat space S^3 picture, the bundle associated with the *canonical* fields is automatically trivial. Nontrivial bundles only arise in space *and* time considerations of vacuum tunneling. The situation is different for a general three-manifold Σ (since G bundles are now classified by elements[4] of the cohomology group $H^2(\Sigma;\pi_1(G))\approx H_1(\Sigma;\pi_1(G))$ and some of these bundles may admit flat connections (i.e., $F_{ij} = 0$). It is clearly impossible to construct spacetime fields $\omega_i(\mathbf{x},t)$ which interpolate between two connections in *different* bundles and hence tunneling between bundle sectors cannot occur. In this sense the "topological charges" in $H^2(\Sigma;\pi_1(G))$ can be regarded as labels for a type of superselection sector. We will only consider the case where the G bundles over Σ are trivial and will defer discussion of the general situation to a future publication.

The second and major effect of $\pi_1(\Sigma)\neq 0$ is the possible existence of zero energy Yang–Mills fields which are only locally like (1.1) and must be gauge patched globally. These arise when the holonomy group (a discrete subgroup of G) of the connection is nonvanishing.[5] Since $F_{ij} = 0$, parallel transport around a curve in Σ depends only on its homotopy class and hence leads to a homomorphism from $\pi_1(\Sigma)$ into G.[6] We shall see in Sec. 2 that the converse is also true and any such homomorphism induces a flat connection in a G bundle over Σ.

The effect of nonvanishing holonomy becomes apparent when we consider the problem of identifying classical

[a)]Supported in part by the Natural Sciences and Engineering Research Council of Canada.

vacua. Two solutions ω and $\tilde{\omega}$ of $F_{ij} = 0$ can be regarded for quantization purposes as being physically equivalent if there exists a one-parameter family of flat connections $\omega^{(\lambda)}$ with $\omega^{(0)} = \omega$ and $\omega^{(1)} = \tilde{\omega}$. Under these conditions ω and $\tilde{\omega}$ may be joined by a path in time with an arbitrarily small action. Equivalently, an adiabatic movement from one configuration to the other may be initiated by imparting an arbitrarily small amount of energy to the system. The quantum vacuum states are to be labeled[2] by the equivalence classes of such solutions and the purpose of the present work is to identify these classes for arbitrary Σ and G. When $\Sigma = S^3$, only the pure gauge configurations (1) exist and two connections are physically equivalent if and only if the corresponding gauge functions Ω are homotopic. This leads to the usual labeling of the quantum vacua by the homotopy classes of functions from S^3 into G, i.e., by $\pi_3(G)$.[2] The homotopy classes $[\Sigma,G]$ of gauge functions from a general Σ into G may be classified[7] by elements of $H^1(\Sigma;\pi_1(G))$ and $H^3(\Sigma;\pi_3(G))$ which we shall refer to as the primary and secondary cohomology classes or "winding numbers." However, labeling the quantum states by $[\Sigma,G]$ would be incorrect as it ignores the existence of connections with nonvanishing holonomy. Such connections may give rise to an interpolating path of zero energy Yang–Mills potentials which renders two connections physically equivalent even though their primary and/or secondary winding numbers are different. This collapsing of vacuum sectors is easily illustrated when $G = \mathrm{U1}$ since *any* pair of flat connections ω and $\tilde{\omega}$ may be linked by the (flat) affine sum $\lambda\omega + (1-\lambda)\tilde{\omega}$ (for a concrete example see Sec. 2. B). Alternatively the presence of holonomy can *increase* the number of sectors. For example, if $\Sigma = \mathbb{R}P^3$ and $G = \mathrm{SU2}$, then $\pi_1(\Sigma) = Z_2 = \{e,a\}$ and connections with the holonomy group Z_2 exist which cannot be linked by a zero energy path to the pure gauge configurations. In this case (Sec. 4.A) each holonomy sector is associated with a countable set of winding numbers and hence the total set of vacuum sectors has twice the number of elements that would arise if holonomy were overlooked.

It is clear from the definition of physical equivalence that the main task is to construct a mathematical representation of the space of zero energy connections and find the set of arc connected components. This zeroth homotopy set will then become the labeling set for the quantum vacuum sectors. We shall exhibit such a representation, $\mathscr{D}$, in Secs. 2 and 3 and show that this space is a principal bundle over a certain subset of the set of homomorphisms from $\pi_1(\Sigma)$ into G with fiber the topological group of gauge functions. Valuable information on $\pi_0(\mathscr{D})$ can be extracted from the homotopy sequence of this bundle (Sec. 4.A) which is used in Sec. 5, to derive specific results. The problem of diagonalizing the resulting states and tunneling amplitudes is discussed in Sec. 4.B.

All spaces will be assumed to possess a preferred base point and, unless stated to the contrary, all maps will be base-point preserving. The set of such maps between two spaces X and Y is denoted Y^X and the set of base point preserving homotopy classes will be written $[X,Y]$. The set of r-times differentiable maps between two C^∞-manifolds M and N is written $C^r(M,N)$.

The preliminary results of our investigation were summarized in Ref. 8.

2. FLAT CONNECTIONS

A. Construction of flat connections

We start by reviewing briefly a standard technique[9,10] for constructing principal G bundles with flat connections. Let $\hat{\Sigma}$ denote the universal covering space of Σ and let $\mathrm{Hom}(\pi_1(\Sigma),G)$ be the set of homomorphisms from $\pi_1(\Sigma)$ into G. Then $\hat{\Sigma}$ is a principal $\pi_1(\Sigma)$-bundle over Σ and, given any h in $\mathrm{Hom}(\pi_1(\Sigma),G)$, leads naturally to a principal G-bundle. More precisely we define $\hat{\Sigma}\times_h G$ to be the set of equivalence classes under the $\pi_1(\Sigma)$ action,

$$\gamma{:}\hat{\Sigma}\times G\rightarrow\hat{\Sigma}\times G,$$

$$(y,g)\rightsquigarrow(y\gamma,h(\gamma^{-1})g), \tag{2.1}$$

where $y\gamma$ denotes the usual action of $\gamma\in\pi_1(\Sigma)$ on $\hat{\Sigma}$. Then the C^∞ manifold $\hat{\Sigma}\times_h G$ is a principal G bundle over Σ under the group action

$$g'{:}\hat{\Sigma}\times_h G\rightarrow\hat{\Sigma}\times_h G,$$

$$[y,g]\rightsquigarrow[y,gg'], \tag{2.2}$$

with projection map $\tilde{r}$ defined by $\tilde{r}[y,g] := r(y)$, where r is the natural C^∞ map from $\hat{\Sigma}$ onto Σ.

Let $t \rightsquigarrow x_t$ be a differentiable loop in Σ starting at the base point $x_0\in\Sigma$ and let $t\rightsquigarrow y_t$ be the covering path in $\hat{\Sigma}$ starting at the base point $y_0\in\hat{\Sigma}$. Then the horizontal lift of x_t in $\hat{\Sigma}\times_h G$, passing through the point $[y_0,g]$, is defined to be

$$t\rightsquigarrow[y_t,g]. \tag{2.3}$$

Now $y_1 = y_0\gamma$, where $\gamma\in\pi_1(\Sigma)$ is the homotopy class of the loop $t\rightsquigarrow x_t$, and hence

$$[y_1,g] = [y_0\gamma,g] = [y_0,h(\gamma)g] = [y_0,g]g^{-1}h(\gamma)g, \tag{2.4}$$

which shows that the holonomy group of the point $[y_0,g]$ is $Adg^{-1}h(\pi_1(\Sigma))$.

Different homomorphisms may induce inequivalent G bundles but in the present paper we will only consider the product case. Thus let $\mathscr{R}$ denote the set of all elements h of $\mathrm{Hom}(\pi_1(\Sigma),G)$ such that $\hat{\Sigma}\times_h G$ is isomorphic to the trivial bundle $\Sigma\times G$. For any such homomorphism there is a trivializing bundle map:

$$\hat{\Sigma}\times_h G\rightarrow\Sigma\times G,$$

$$[y,g]\rightsquigarrow(r(y),D(y)g), \tag{2.5}$$

with

$$D(y\gamma) = D(y)h(\gamma)\quad\text{for all }\gamma\text{ in }\pi_1(\Sigma). \tag{2.6}$$

The condition (2.6) on the map D from $\hat{\Sigma}$ into G ensures that (2.5) is independent of the choice of representative elements y and g in the equivalence class $[y,g]$. In terms of the fixed product structure $\Sigma\times G$, the horizontal lift (2.3) becomes $t\rightsquigarrow(x_t,D(y_t)g)$, which is to be compared with the lift $t\rightsquigarrow(x_t,g)$ corresponding to the trivial connection in $\Sigma\times G$ with a vanishing Yang–Mills field. Using standard results in Ref. 6 it follows that the Yang–Mills field on $\Sigma\times G$ associated with the homomorphism h, is the Lie-algebra-valued one-form on Σ,

$$\omega_i(x) = D(y)\partial_i D(y)^{-1},\quad r(y) = x, \tag{2.7}$$

where ∂_i is defined by lifting the local coordinates in Σ to $\hat{\Sigma}$. By virtue of (2.6), the right-hand side of (2.7) is independent of the point y in the fiber in $\hat{\Sigma}$ lying over $x \in \Sigma$. Equivalently, ω may be defined as the projection into Σ of the $\pi_1(\Sigma)$-invariant, Lie-algebra-valued one-form on $\hat{\Sigma}$, $D^{-1*}(\theta)$, where θ is the Cartan–Maurer form on G.

Since every flat connection on $\Sigma \times G$ can be obtained in this way[9] we see that the set of zero energy C^∞ Yang–Mills fields on the trivial G bundle is in bijective correspondence with the set

$$\mathscr{D}^{(\infty)} := \{D \in C^\infty(\hat{\Sigma}, G) \mid \exists h \in \mathscr{R} \text{ such that } D(y\gamma) = D(y)h(\gamma) \forall \gamma \in \pi_1(\Sigma)\}. \tag{2.8}$$

Note that gauge transformations appear in the form

$$D(y) \rightsquigarrow \Omega(r(y))D(y). \tag{2.9}$$

If $\omega_i^{(0)} = D^{(0)}\partial_i D^{(0)-1}$ and $\omega_i^{(1)} = D^{(1)}\partial_i D^{(1)-1}$ then a one-parameter family $\omega_i^{(\lambda)}$ of interpolating flat connections may be written as

$$\omega_i^{(\lambda)}(x) = C_\lambda(y)\partial_i C_\lambda(y)^{-1} \quad \text{with} \quad C_\lambda(y\gamma) = C_\lambda(y)h_\lambda(\gamma), \tag{2.10}$$

where $C_0 = D^{(0)}$, $C_1 = D^{(1)}$, and $\lambda \rightsquigarrow h_\lambda$ is a path in $\mathscr{R}$ whose end points are the homomorphisms producing the functions $D^{(0)}$ and $D^{(1)}$.

B. A U1 example

To illustrate these ideas consider a U1 example with $\Sigma = S^1$ (hence $\hat{\Sigma} = \mathbb{R}$). Let $D^{(1)}(y) = e^{i2\pi ny}$ and $D^{(2)}(y) = e^{i2\pi my}$. Then, if S^1 is parametrized by an angle θ lying between 0 and 1, the period of the covering space $\mathbb{R}$ is one and $D^{(1)}$ and $D^{(2)}$ correspond to pure gauge functions with primary winding numbers n and m and potentials $\omega_\theta^{(1)} = -i2\pi n$ and $\omega_\theta^{(2)} = -i2\pi m$, respectively. The affine sum $\omega^{(\lambda)} = \lambda\omega_\theta^{(1)} + (1-\lambda)\omega_\theta^{(2)}$ is associated with the function $C_\lambda(y) = \exp i2\pi y\{\lambda n + (1-\lambda)m\}$, which satisfies

$$C_\lambda(y+1) = C_\lambda(y)\exp\{i2\pi\lambda(n-m)\} \tag{2.11}$$

and the holonomy group of $\omega^{(\lambda)}$ is Z if λ is irrational and $Z_{(q,n-m)}$ if $\lambda = p/q$ with p and q having no common divisors. Note that although C_λ interpolates smoothly between $D^{(1)}$ and $D^{(2)}$ the holonomy group changes discontinuously.

3. TOPOLOGICAL PROPERTIES OF $\mathscr{D}$

A. The topology on $\mathscr{D}$

In Sec. 2 we constructed C^∞ connections using C^∞ maps from $\hat{\Sigma}$ into G. However, it is difficult to decide *a priori* what degree of differentiability is really appropriate. In order that F_{ij} can be constructed, the functions D from $\hat{\Sigma}$ to G must be at least twice differentiable and it is natural to extend the treatment of Sec. 2 to C^r fields and define

$$\mathscr{D}^{(r)} := \{D \in C^{(r)}(\hat{\Sigma}, G) \mid \exists h \in \mathscr{R} \text{ such that } D(y\gamma) = D(y)h(\gamma) \forall \gamma \in \pi_1(\Sigma)\}. \tag{3.1}$$

Thus in principle we might obtain different vacuum-state classifications by $\pi_0(\mathscr{D}^{(r)})$ for different r's.

This problem cannot be resolved without first specifying the function-space topologies. Again there is no obvious unique choice but there are two natural requirements. The gauge transformation map (cf. 2.9) $C^r(\Sigma,\mathrm{G}) \times \mathscr{D}^{(r)} \rightarrow \mathscr{D}^{(r)}$, $(\Omega, D) \rightsquigarrow \Omega \cdot rD$, should be continuous and if D_n is a convergent sequence (or net) of $C^{(r)}$ functions, then $D_n \partial_i D_n^{-1}$ should also converge. These criteria are met easily by employing the compact open C^r topologies on $C^r(\Sigma, G)$ and $C^r(\hat{\Sigma}, G)$ and by giving $\mathscr{D}^{(r)}$ the subspace topology. In the Appendix we prove the useful result that, for all r, $\pi_0(\mathscr{D}^{(r)}) = \pi_0(\mathscr{D}^{(0)})$. This enables us to concentrate on the space $\mathscr{D} \equiv \mathscr{D}^{(0)}$ which is readily subject to an algebraic-topological analysis.

B. The bundle of flat connections

The continuous action in Sec. 3.A of G^Σ on $\mathscr{D}$ is free and suggests that $\mathscr{D}$ might be a principal G bundle over the quotient space $\mathscr{D}/G^\Sigma$. Moreover $\mathscr{D}/G^\Sigma$ is clearly in bijective correspondence with $\mathscr{R}$ via the projection map which associates with each $D \in \mathscr{D}$ the homomorphism h for which $D(y\gamma) = D(y)h(\gamma)$. If $\mathscr{D}$ were indeed a bundle over $\mathscr{R}$ the associated homotopy exact sequence would provide potent information on $\pi_0(\mathscr{D})$.

The first step is to put a suitable topology on $\mathrm{Hom}(\pi_1(\Sigma), G)$. Since $\pi_1(\Sigma)$ is a discrete group, the natural choice is the point open topology in which the open sets are arbitrary unions and finite intersections of all sets of the form

$$N_{\gamma,O} = \{h \in \mathrm{Hom}(\pi_1(\Sigma), G) \mid h(\gamma) \subset O\}, \tag{3.2}$$

where $\gamma \in \pi_1(\Sigma)$ and O is an open subset of the Lie group G. Weil has observed that $\mathrm{Hom}(\pi_1(\Sigma), G)$ is a real analytic variety.[11] This is shown by considering a presentation of $\pi_1(\Sigma)$ in terms of n generators and m relations (n and m are finite integers since Σ is compact). If F is the corresponding free group then $\mathrm{Hom}(F,G)$ is a product $X_{i=1}^n G$ of n copies of G and $\mathrm{Hom}(\pi_1(\Sigma), G)$ is the real analytic subset of $X_{i=1}^n G$ on which the m relations are satisfied. Analytic sets are locally arcwise connected and hence the path components of $\mathrm{Hom}(\pi_1(\Sigma), G)$ coincide with the topological components. All homomorphisms in a path component induce isomorphic principal G bundles $\hat{\Sigma} \times_h G$ and in particular the set $\mathscr{R}$ of homomorphisms which induce the product bundle is a union of components of $\mathrm{Hom}(\pi_1(\Sigma), G)$.

Let q be the projection from $\mathscr{D}$ onto $\mathscr{R}$ which maps D into the defining homomorphism. Since D is base-point preserving, $D(y_0) = \mathbb{1}$, and hence $D(y_0\gamma) = h(\gamma)$. Thus a concrete representation for q is

$$q(D)(\gamma) := D(y_0\gamma). \tag{3.3}$$

We have the important result

Proposition 3.1. $\mathscr{D}$ is a locally trivial, principal G bundle over $\mathscr{R}$ with projection map q.

Proof: (a) q is continuous since $q^{-1}(N_{\gamma,O}) = \{D \in \mathscr{D} \mid D(y_0\gamma) \subset O\}$, which is open in the compact open topology. The action of G^Σ on $\mathscr{D}$ is continuous and free.

(b) The main step is to show that $\mathscr{D}$ is locally trivial. As an analytic set $\mathrm{Hom}(\pi_1(\Sigma), G)$ is triangulable and hence each element possesses an open neighborhood that is contractible over itself. Furthermore, since $\mathrm{Hom}(\pi_1(\Sigma), G)$ is a closed subspace of the compact space $X_{i=1}^n G$ it is itself compact and can be covered by a finite number of such sets. We will prove local triviality by constructing a continuous local section σ_V over each contractible set $V \subset \mathscr{R}$.

Let V be such a set with a map $F: V \times I \to V$ with $F(h,0) = h$ and $F(h,1) = \hat{h}$ for some $\hat{h}$. This induces a deformation retraction of $\Sigma \times V$ onto $\Sigma \times \{\hat{h}\}$. Define a $\pi_1(\Sigma)$ action on $\hat{\Sigma} \times V \times G$ by

$$\gamma:(y,h,g) \rightsquigarrow (y\gamma,h,h(\gamma^{-1})g) \tag{3.4}$$

and let $(\hat{\Sigma} \times V) \times_T G$ denote the set of equivalence classes under this action. This is a principal G bundle over $\Sigma \times V$ with the projection map $\eta[y,h,g] = (r(y),h)$ and right G action $[y,h,g]g' := [y,h,gg']$. Now the map $f: \Sigma \times V \to \Sigma \times V$, $(x,h) \rightsquigarrow (x,\hat{h})$ induces a pullback from $(\hat{\Sigma} \times V) \times_T G$ to give a bundle that is isomorphic to the trivial bundle $(\hat{\Sigma} \times_{\hat{h}} G) \times V$. Moreover f is homotopic to the identity and hence $(\hat{\Sigma} \times V) \times_T G$ is also a trivial principal G bundle and therefore admits global cross sections. These are of the form

$$\varphi(x,h) = [y,h,D^{-1}(y,h)], \tag{3.5}$$

where $D(y\gamma,h) = D(y,h)h(\gamma)$.

Thus $\sigma_V(h)(y) := D(y,h)$ is the desired local section over V.

(c) Finally we show that $\mathscr{R}$ is homeomorphic to $\mathscr{D}/G^\Sigma$. The projection map $p: \mathscr{D} \to \mathscr{D}/G^\Sigma$ is continuous and open when $\mathscr{D}/G^\Sigma$ carries the usual quotient topology. The natural map $j: \mathscr{D}/G^\Sigma \to \mathscr{R}, j[D] := q(D)$ is continuous and bijective with an inverse $i: \mathscr{R} \to \mathscr{D}/G^\Sigma, i(h) := [D_h]$ where D_h is any element in $\mathscr{D}$ with $q(D) = h$. Over a contractible open set V, $i_{|V} = p \cdot \sigma_V$ is continuous and since $\mathscr{R}$ is covered by such sets, i is continuous on $\mathscr{R}$. Hence $\mathscr{R}$ is homeomorphic to $\mathscr{D}/G^\Sigma$. Q.E.D.

Although the bundle is locally trivial it will not in general be trivial and global cross sections will not exist. Thus we have a type of Gribov phenomenon,[12] although it should be emphasized that our bundle of flat connections is quite different from the bundle of irreducible connections considered in Refs. 13 and 14.

4. THE HOMOTOPY EXACT SEQUENCE FOR $\pi_0(\mathscr{D})$

A. The exact sequence

Considerable information on $\pi_0(\mathscr{D})$ may be extracted from the homotopy exact sequence of the fiber bundle in Proposition (3.1). $\mathscr{D}$ decomposes into disjoint pieces over the components of $\mathscr{R}$ and in each one we choose a base point D_h lying over the base point h in the corresponding $\mathscr{R}$ component C_h. The group G^Σ is injected into the fiber over h by the map $i: \Omega \rightsquigarrow \Omega D_h$. The homotopy exact sequence of the bundle is

$$\pi_1(\mathscr{R},h) \xrightarrow{\partial} [\Sigma,G] \xrightarrow{i_*} \pi_0(\mathscr{D},D_h) \xrightarrow{q_*} \pi_0(\mathscr{R},h) \to \{*\}, \tag{4.1}$$

where the last two entries are sets (not groups) whose base points are the components containing D_h and h, respectively. Note that we have used $\pi_0(G^\Sigma) = [\Sigma,G]$ and that the surjection of q_* follows from that of q.

The general theory[15] of this exact sequence of sets and groups provides the crucial result

$$\begin{aligned} q_*^{-1}(C_h) &= [\Sigma,G]/\ker i_* \\ &\approx [\Sigma,G]/\partial(\pi_1(\mathscr{R},h)). \end{aligned} \tag{4.2}$$

Thus in principle we have achieved a complete specification of $\pi_0(\mathscr{D})$ in terms of the components $\{C_h\}$ of $\mathscr{R}$ and the abelian group $[\Sigma,G]/\partial(\pi_1(\mathscr{R},h))$. Note that, if $\pi_1(\mathscr{R},h) = 0$ for all h, the vacuum-state classification is simply

$$\pi_0(\mathscr{D}) = [\Sigma,G] \times \pi_0(\mathscr{R}). \tag{4.3}$$

In particular, if $\Sigma = S^3$ then $\pi_1(\Sigma) = 0$ and $\mathscr{R} = \{*\}$ and the usual classification by $[S^3,G] = \pi_3(G)$ is reproduced. Another example where Eq. (4.3) is applicable is $\Sigma = \mathbb{R}P^3$ and $G = \mathrm{SU2}$. Then $\pi_1(\Sigma) = Z_2 \approx (e,a)$ and $\mathrm{Hom}(\pi_1(\Sigma),G)$ possesses just two elements given by $h(a) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $h(a) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, which induce connections with trivial- and Z_2-holonomy groups, respectively. Since all SU2 bundles over a three-manifold are trivial[5] it follows that $\mathrm{Hom}(\pi_1(\Sigma),\mathrm{SU2}) = \mathscr{R}$ and $\pi_0(\mathscr{D})$ is the disjoint union of two copies of $[\Sigma,G] = [\mathbb{R}P^3,\mathrm{SU2}] = Z$.

B. The quantum states

We will denote the states corresponding to C_h and $l \in [\Sigma,G]/\partial(\pi_1(\mathscr{R},h))$ by $|C_h,l\rangle$ and assume that, as usual, there is a gauge transformation operator T_Ω acting on these states such that

$$T_\Omega|C_h,l\rangle = |C_h,[\Omega]l\rangle. \tag{4.4}$$

Thus the states are not only invariant under "small gauge" transformations (i.e., those that are homotopic to a constant) but also under those whose homotopy classes belong to the little group $\partial(\pi_1(\mathscr{R},h))$. We may diagonalize the gauge action in analogy with the usual flat space theory and construct the "θ-vacua"

$$|C_h,\theta\rangle = \sum_l \theta(l)|C_h,l\rangle, \tag{4.5}$$

where θ is a character of the abelian group $[\Sigma,G]/\partial(\pi_1(\mathscr{R},h))$. These new states are gauge invariant up to phase factors and have the inner products

$$\langle C_h,\theta|C_h,\theta'\rangle = \delta_{\theta\theta'} \sum_l \langle C_h,0|C_h,l\rangle\theta(l), \tag{4.6}$$

where the sum is over $[\Sigma,G]/\partial(\pi_1(\mathscr{R},h))$. They also diagonalize the tunneling amplitudes within a fixed holonomy sector (i.e., the "in" and "out" states have the same C_h label) but there is no reason to expect that transitions between different elements of $\pi_0(\mathscr{R})$ can be easily diagonalized.

Tunneling is expected between different holonomy sectors on a compact Σ because the affine sum of any two flat, nonsingular connections describes a configuration of zero energy. In other words, the time-dependent configuration $\omega_\mu(\mathbf{x},t)$ defined by

$$\begin{aligned} \omega_i(\mathbf{x},t) &:= \lambda(t)\omega_i^{(1)}(\mathbf{x}) + (1-\lambda(t))\omega_i^{(2)}(\mathbf{x}), \\ \omega_0(\mathbf{x},t) &:= 0 \end{aligned} \tag{4.7}$$

interpolates between the two flat, static connections $\omega_i^{(1)}$ and $\omega_i^{(2)}$ with finite Euclidean action $L[\lambda]$. This configuration can contribute to the transition amplitude within the functional integral approach to quantum field theory, even though ω_μ is not a solution to the classical field equations. It is interesting to note that the function $\tilde{\lambda}(t) := \lambda(t) - \frac{1}{2}$ which minimizes $L[\lambda]$ satisfies

$$\left(\frac{d\tilde{\lambda}}{dt}\right)^2 - \frac{K_1}{K_2}\left(\tilde{\lambda}^2 - \frac{1}{4}\right) = 0, \tag{4.8}$$

where

$$K_1 := \int_\Sigma d^3x\,([\omega^{(1)} - \omega^{(2)}] \wedge [\omega^{(1)} - \omega^{(2)}])^2,$$

$$K_2 := \int_\Sigma d^3x\,(\omega^{(1)} - \omega^{(2)})^2.$$

Equation (4.8) can be recognized from standard ϕ^4 theory to have a kink solution.[16] (Note that the above argument will be inappropriate if Σ is originally noncompact and boundary conditions are imposed that lead to non-square-integrable fields).

It is clear that the main mathematical problem is to represent the boundary map ∂ in as concrete a form as possible. In terms of the bundle $\mathscr{D}$, this map is obtained by taking a loop ρ in $\mathscr{R}$ such that $\rho(0) = h$, and considering its lift to a curve $\lambda \rightsquigarrow D_\lambda$ in $\mathscr{D}$ with $D_1 = D_h$. Then the homotopy class of $D_0 D_h^{-1}$ in G^Σ is independent of the choice of lift and $\partial[\rho] = [D_0 D_h^{-1}]$. Thus ∂ defines the characteristic class[9] of the bundle $\mathscr{D}$ restricted to the component C_h of $\mathscr{R}$ and may be viewed as an element of

$$\mathrm{Hom}(\pi_1(\mathscr{R},h),[\Sigma,G]) = H^1(C_h;[\Sigma,G]). \tag{4.9}$$

C. Prime three-manifolds

It should be noted that a special role is played by prime three-manifolds. Such a manifold cannot be decomposed into a topological sum $\Sigma_1 * \Sigma_2$ unless either Σ_1 or Σ_2 is S^3 and conversely every compact three-manifold can be decomposed into a connected sum of a finite set of prime manifolds. Now $\pi_1(\Sigma_1 * \Sigma_2)$ is the free product[17] $\pi_1(\Sigma_1) * \pi_1(\Sigma_2)$ and hence

$$\mathrm{Hom}(\pi_1(\Sigma_1 * \Sigma_2),G) = \mathrm{Hom}(\pi_1(\Sigma_1),G) \times \mathrm{Hom}(\pi_1(\Sigma_2),G). \tag{4.10}$$

Thus the computation of $\mathrm{Hom}(\pi_1(\Sigma),G)$ may be reduced to that for prime manifolds. As an example consider the prime space $S^1 \times S^2$ with $\pi_1(S^1 \times S^2)$ the free group Z. Then $\mathscr{R} \approx \mathrm{Hom}(\pi_1(\Sigma),G) \approx G$ and (4.1) becomes

$$\to \pi_1(G) \overset{\partial}{\to} [\Sigma,G] \overset{i_*}{\to} \pi_0(\mathscr{D}) \mapsto \{*\}, \tag{4.11}$$

where for simplicity it has been assumed that G is connected so that $\pi_0(G) = \{*\}$. It follows at once that if $\pi_1(G) = 0$ then $\pi_0(\mathscr{D}) = [\Sigma,G] = [S^1 \times S^2,G] \approx Z$ and so the vacuum states are labeled by a single integer. On the other hand if $\Sigma = (S^1 \times S^2) * (S^1 \times S^2)$ then $\mathrm{Hom}(\pi_1(\Sigma),G) = G \times G$ and (4.11) now contains $\pi_1(G) \times \pi_1(G)$. Once again $\pi_0(\mathscr{D}) = Z$ if $\pi_1(G) = 0$.

In general it is very difficult to compute $\mathrm{Hom}(\pi_1(\Sigma),G)$ directly and information on ∂ needs to be obtained by other means.

5. SOME RESULTS ON THE BOUNDARY MAP ∂

A. $\pi_1(\Sigma)$ is infinite

We will now discuss the computation of ∂ for the component of $\mathscr{R}$ which contains the trivial homomorphism $h_0(\gamma) = \mathbb{1}$. Let the loop space $\Omega\,\mathrm{Hom}(\pi_1(\Sigma),G)$ carry the compact open topology.[18] Since the loops pass through the base point h_0 we have $\Omega\,\mathrm{Hom}(\pi_1(\Sigma),G) = \Omega\mathscr{R}$ and hence $\pi_1(\mathscr{R},h_0) = \pi_0(\Omega\,\mathrm{Hom}(\pi_1(\Sigma),G))$. Our first result is

Proposition 5.1. $\pi_1(\mathscr{R},h_0) = 0$ unless $\pi_1(\Sigma)$ contains elements of infinite order.

Proof: Let $\gamma = \pi_1(\Sigma)$ be such that $\gamma^p = 1$ for some p. Then $(h(\gamma))^p = \mathbb{1}$ for all $h \in \mathrm{Hom}(\pi_1(\Sigma),G)$. The exponential map is a local diffeomorphism from the Lie algebra of G onto G and hence there is some open neighborhood U of $\mathbb{1} \in G$ such that $g^p = \mathbb{1}$ if and only if $g = \mathbb{1}$. Define $V_1 := \{h \,|\, h(\gamma) = \mathbb{1}\}$ and $V_2 := \{h \,|\, h(\gamma) \notin U\}$. These are closed subsets of $\mathrm{Hom}(\pi_1(\Sigma),G)$ with $V_1 \cap V_2 = \emptyset$ and $V_1 \cup V_2 = \mathrm{Hom}(\pi_1(\Sigma),G)$.

Now consider $\rho \in \Omega\,\mathrm{Hom}(\pi_1(\Sigma),G)$. Then $w_1 := \rho^{-1}(V_1) = \{\lambda \in S^1 \,|\, \rho(\lambda)(\gamma) = \mathbb{1}\}$ and $w_2 := \rho^{-1}(V_2) = \{\lambda \in S^1 \,|\, \rho(\lambda)(\gamma) \notin U\}$ are closed subsets of S^1 with $S^1 = w_1 \cup w_2$ and $w_1 \cap w_2 = \emptyset$. Thus $w_2 = w_1^c$ and hence w_2 is open. But S^1 is connected and therefore $w_2 = \emptyset$ or S^1. However, $\rho(0)(\gamma) = \mathbb{1}$ and so $w_1 \neq \emptyset$ which implies $w_2 = \emptyset$ and $w_1 = S^1$. Thus $\rho(\lambda)(\gamma) = \mathbb{1}$ and so ρ is trivial if all elements of $\pi_1(\Sigma)$ are of finite order.

In particular, if $\pi_1(\Sigma)$ is finite, $q_*^{-1}(C_{h_0}) = [\Sigma,G]$ and the quantum vacua in the C_{h_0} holonomy sector are labeled by the winding numbers in $[\Sigma,G]$.

B. Reduction to a lifting problem

Assume now that $\pi_1(\Sigma)$ is infinite and that ρ is a nontrivial loop in $\Omega\mathscr{R}$. We seek a curve C_λ in $\mathscr{D}$ such that $C_1(y) = 1$ and $C_\lambda(y\gamma) = C_\lambda(y)\rho(\lambda)(\gamma)$ as then $\partial[\rho]$ is equal to the homotopy class of $C_0 \in G^\Sigma$. The computation of C_0 is equivalent to finding the pairs of gauge functions that can be linked by paths in $\mathscr{D}$ in the sense that Ω_1 can be joined to Ω_2 if and only if $\Omega_1\Omega_2^{-1}$ can be joined to the identity. The homotopy class of C_0 lies in $[\Sigma,G]$ and can be classified by the primary and secondary winding numbers belonging to $H^1(\Sigma;\pi_1(G))$ and $H^3(\Sigma;\pi_3(G))$, respectively. Ideally one would relate the group elements to the homotopical properties of ρ and the rest of this section is devoted to deriving results of this type.

Let E_λ be a curve in $\mathscr{D}$ satisfying $E_0(y) = E_1(y) = \mathbb{1}$ and $E_\lambda(y\gamma) = E_\lambda(y)\rho(\lambda)(\gamma)$. If such a curve exists then $C_\lambda(y\gamma)E_\lambda^{-1}(y\gamma) = C_\lambda(y)E_\lambda^{-1}(y)$ and hence $C_\lambda(y)E_\lambda^{-1}(y)$ may be viewed as a function Ω_λ from Σ into G. However, $\Omega_1 = 1$ and $\Omega_0 = C_0$, so that $\lambda \rightsquigarrow \Omega_\lambda$ is a homotopy from C_0 to $\mathbb{1}$. Thus $\partial[\rho]$ is trivial if and only if an E_λ exists. We will start an investigation by studying the obstructions to the construction of E_λ.

Let ΩG denote the loop space of G equiped with the compact open topology. This function space is a topological group with the product law

$$(\omega_1 \cdot \omega_2)(\lambda) := \omega_1(\lambda)\omega_2(\lambda). \tag{5.1}$$

Of course ΩG is also an H space under the loop composition

$$\begin{aligned}(\omega_1 \vee \omega_2)(\lambda) &= \omega_1(2\lambda), \quad 0\leqslant\lambda\leqslant\tfrac{1}{2}\\ &= \omega_2(2\lambda-1), \quad \tfrac{1}{2}\leqslant\lambda\leqslant 1 \end{aligned} \tag{5.2}$$

and these two H structures can be shown to be homotopic,[19] i.e., $\omega_1\cdot\omega_2 \sim \omega_1 \vee \omega_2$ for all pairs ω_1 and ω_2. There is the natural homeomorphism

$$\begin{aligned}&\Omega\,\mathrm{Hom}(\pi_1(\Sigma),G)\rightarrow\mathrm{Hom}(\pi_1(\Sigma),\Omega G),\\ &\qquad\qquad \rho \rightsquigarrow \chi_\rho,\end{aligned} \tag{5.3}$$

where $\chi_\rho(\gamma)(\lambda) := \rho(\lambda)(\gamma)$ and $\mathrm{Hom}(\pi_1(\Sigma),\Omega G)$ has the point open topology.

Let $(E\pi,\eta_\pi,B\pi)$ be a universal $\pi_1(\Sigma)$ bundle with a map $u:\Sigma\rightarrow B\pi$ inducing the universal covering space $\hat{\Sigma}$ and a corresponding bundle map $\bar{u}:\hat{\Sigma}\rightarrow E\pi$. By analogy with the discussion in Sec. 2 we may regard $E\pi$ as a principal (flat) ΩG bundle via the homomorphism

$$\chi_\rho:\pi_1(\Sigma)\rightarrow\Omega G. \tag{5.4}$$

Thus we form the quotient $E\pi\times_{\chi_\rho}\Omega G$ of $E\pi\times\Omega G$ by the $\pi_1(\Sigma)$-action $\gamma(e,\omega) = (e\gamma,\chi_\rho(\gamma)^{-1}\omega)$ and define a projection $\eta:E\pi\times_{\chi_\rho}\Omega G\rightarrow B\pi$, $\eta[e,\omega] := \eta_\pi(e)$ and a right ΩG-action $[e,\omega]\omega' := [e,\omega\omega']$. This principal ΩG bundle is induced by a map $B\chi_\rho$ into the base space of a universal ΩG bundle. We show in the Appendix that there exists such a bundle whose base space is G and we have the picture

$$\begin{tikzcd}
\Omega G \arrow[r] & E\pi\times_{\chi_\rho}\Omega G \arrow[d] & \\
\Sigma \arrow[ur, dashed, "\phi"] \arrow[r, "u"'] & B\pi \arrow[r, "B\chi_\rho"'] & G.
\end{tikzcd}$$

(5.5)

Now a lift ϕ of u is necessarily of the form $\phi(x) = [\bar{u}(y),E^{-1}(y)]$, with $E(y\gamma)(\lambda) = E(y)(\lambda)\rho(\lambda)(\gamma)$ and $E(y)(0) = E(y)(1) = \mathbb{1}$. Hence the definition $E_\lambda(y) := E(y)(\lambda)$ provides the one-parameter family of functions that we seek. Conversely, any such E_λ gives rise to a lift ϕ. But η is a principal fibration with classifying map $B\chi_\rho$ and hence u lifts if and only if $B\chi_\rho\cdot u$ is homotopically trivial.[20,21] Since $B\chi_\rho\cdot u$ is a map from Σ into G it is classified[7] by elements of the cohomology groups $H^1(\Sigma;\pi_1(G))$ and $H^3(\Sigma;\pi_3(G))$, and an E_λ exists if and only if $B\chi_\rho\cdot u\sim *$, which will be true if and only if both group elements vanish ($*$ denotes the constant map from Σ into G).

C. Obstructions to constructing a C_λ

Before discussing the cohomological relations between $B\chi_\rho$ and ρ let us return to the problem of C_λ. By analogy with the above, we construct the principal ΩG bundle $E\pi\times_\chi PG$ over $B\pi\times G$ with the projection map $t[e,p] = (\eta_\pi(e),p(0))$ and ΩG action $[e,p]\omega = [e,p\omega]$. Consider the map $(u,A):\Sigma\rightarrow B\pi\times G$, $(u,A)(x) := (u(x),A(x)^{-1})$, where $A\in G^\Sigma$, and contemplate the lifting problem

$$\begin{tikzcd}
\Omega G \arrow[r] & E\pi\hat{\times}_\chi PG \arrow[d] & \\
\Sigma \arrow[ur, dashed, "\phi"] \arrow[r, "(u{,}A)"'] & B\pi\times G \arrow[r, "l"'] & G.
\end{tikzcd}$$

(5.6)

Lifts of (u,A) are in one-to-one correspondence with curves C_λ in $\mathscr{D}$ such that $C_\lambda(y,\gamma) = C_\lambda(y)\rho(\lambda)(\gamma)$, $C_1(y) = \mathbb{1}$, and $C_0(y) = A(r(y))$, where r is the projection from $\hat{\Sigma}$ into Σ. Thus the main problem is to find the homotopy class of A in terms of ρ and the first step is to compute the classifying map l.

Proposition 5.3. The classifying map l is given by the chain of maps

$$B\pi\times G \overset{B\chi\times 1}{\rightarrow} G\times G \overset{\mu}{\rightarrow} G \quad\text{and}\quad \mu(g_1,g_2) := g_1g_2.$$

Proof: (a) There is an ΩG-equivariant map $B\bar{\chi}:E\pi\times_\chi\Omega G\rightarrow PG$ of the form $B\bar{\chi}[e,\omega] = L(e)\omega$, where $L(e\gamma) = L(e)\chi(\gamma)$ for all $\gamma\in\pi_1(\Sigma)$. Define $PG\times_T PG$ by the equivalence relation $(p_1,p_2)\equiv(p_1\omega,\omega^{-1}p_2)\forall\omega\in\Omega G$. Then $PG\times_T PG$ is a principal ΩG bundle under the action $[p_1,p_2]\omega = [p_1,p_2\omega]$ and there is a bundle-map pair $L\times 1:E\pi\times_\chi\Omega G\rightarrow PG\times_T PG$; $B\chi\times 1:B\pi\times G\rightarrow G\times G$ with the projection $PG\times_T PG\overset{\delta}{\rightarrow} G\times G$, $\delta(p_1,p_2) = (p_1(0),p_2(0))$.

(b) There is another bundle-map pair $PG\times_T PG\rightarrow PG$, $[p_1,p_2]\rightsquigarrow p_1p_2$; $G\times G\rightarrow G$, $(g_1,g_2)\rightsquigarrow g_1g_2$ and hence $E\pi\times_\chi PG$ is isomorphic to the pull back of the universal bundle PG by tne composition of these two maps. Q.E.D.

Lifts of (u,A) (and hence C_λ) exist if and only if $l\cdot(u,A)\sim *$, which in turn is true if and only if the corresponding members of $H^1(\Sigma;\pi_1(G))$ and $H^3(\Sigma;\pi_3(G))$ vanish. The element in $H^1(\Sigma;\pi_1(G))$ is $l\cdot(u,A)^*\iota_1$ where ι_1 is the characteristic element[22] in $H^1(G;\pi_1(G)) = \mathrm{Hom}(\pi_1(G),\pi_1(G))$; [it is the identity homomorphism from $\pi_1(G)$ onto $\pi_1(G)$] and we have

Lemma 5.4: If $h\in H^1(G;\pi_1(G))$ then $l\cdot(u,A)^*h = ((B\chi\cdot u)^* - A^*)h$.

Proof: Any $h\in H^1(G;\pi_1(G))$ is primitive and hence $\mu^*(h) = h\times 1 + 1\times h$ in $H^1(G\times G;\pi_1(G))$. Thus $l^*(h) = B\chi^*(h)\times 1 + 1\times h$ in $H^1(B\pi\times G;\pi_1(G))$. Now $(\mu,A):\Sigma\rightarrow B\pi\times G$ may be factored as $\Sigma\overset{\Delta}{\rightarrow}\Sigma\times\Sigma\overset{u\times A}{\rightarrow}\mathrm{B}\pi\times\mathrm{G}\overset{1\times\nu}{\rightarrow}B\pi\times G$, where Δ is the diagonal map $x\rightsquigarrow(x,x)$ and $\nu(g) := g^{-1}$. We have the exact sequence

$$\begin{aligned}&0\rightarrow H^1(\Sigma;\pi_1(G))\oplus H^1(\Sigma;\pi_1(G))\rightarrow H^1(\Sigma\times\Sigma;\pi_1(G))\\ &\quad\rightarrow H^2(\Sigma;\pi_1(G))*\pi_1(G)\oplus\pi_1(G)*H^2(\Sigma;\pi_1(G))\rightarrow 0\end{aligned}$$

and, on the subgroup $H^1(\Sigma;\pi_1(G))\oplus H^1(\Sigma;\pi_1(G))$, the map Δ^* is $\Delta^*(a,b) = a+b$ in $H^1(\Sigma;\pi_1(G))$. However, $l^*(h)$ has no components in the torsion-product parts of $H^1(B\pi\times G;\pi_1(G))$ and so $\Delta^*\cdot(u\times A)^*\cdot(1\times\nu)^*\cdot l^*(h) = (B\chi\cdot u)^*(h) - A^*(h)$, which proves the result. Q.E.D.

Similar considerations apply to the pullback of the generator ι_3 of $H^3(G;\pi_3(G))$ which is isomorphic to Z if G is simple and nonabelian so that $\pi_3(G) = Z$. [Of course there are no secondary numbers if $\pi_3(G) = 0$]. If $\pi_1(G) = 0$ then ι_3 is primitive and, as in Lemma 5.4,

$$(l\cdot(u,A))^*\iota = ((B\chi\cdot u)^* - A^*)\iota \tag{5.7}$$

for all ι in $H^3(G;\pi_3(G))$. If $\pi_1(G)\neq 0$ then ι may not be primitive and $\mu^*(\iota) = \iota\times 1 + 1\times\iota$ plus terms in $H^2(G;Z)\otimes H^1(G;Z)$ and $H^2(G;Z)*H^2(G;Z)$. Now $\pi_1(G)$ is abelian and, if finite, it is a sum of cyclic groups and hence $H^1(G;Z) = \mathrm{Hom}(\pi_1(G),Z) = 0$. On the other hand if $\pi_1(G) = Z$, then the Serre exact cohomology sequence[22,23] applied to the fibration $\hat{G}\rightarrow G\rightarrow B\pi_1(G)$ ($\hat{G}$ is the universal covering group of G) shows that $H^2(G;Z) = H^2(B\pi_1(G);Z) = H^2(BZ;Z) = H^2(S^1;Z) = 0$. Thus there are no $H^2(G;Z)\otimes H^1(G;Z)$ terms for any $\pi_1(G)$. In addition, $H^2(G;Z)*H^2(G;Z)$ contains only elements of finite order and these can be neglected since their pullback to $H^3(\Sigma;Z)$ must vanish. This is because $H^3(\Sigma;Z)\approx Z$ for Σ compact and orientable (which we are assuming). The net effect is that ι acts as if it were primitive and we have shown

Lemma 5.5: If $\iota \in H^3(G;\pi_3(G))$ then $l\cdot(u,A)^*\iota = ((B\chi\cdot u)^* - A^*)\iota$.

We can now derive the important result

Proposition 5.6: The map $\partial:\pi_1(\mathscr{R},h_0)\to[\Sigma,G]$ is $\partial[\rho] = [B\chi_\rho\cdot u]$.

Proof: A curve C_λ in $\mathscr{D}$ linking A to $\mathbb{1}$ exists if $l\cdot(u,A)\sim *$, which is true if $l\cdot(u,A)^*\iota_1 = 0 = l\cdot(u,A)^*\iota_3$, which by the results above, is equivalent to $(B\chi_\rho\cdot u)^*\iota_1 = A^*\iota_1$ and $(B\chi_\rho\cdot u)^*\iota_3 = A^*\iota_3$. However, the homotopy class of a map is uniquely determined by its pullback of ι_1 and ι_3 and consequently $l\cdot(u,A)\sim *$ iff $[B\chi_\rho\cdot u] = [A]$. But $\partial[\rho] = [C_0] = [A]$.

Q.E.D.

Thus, in the component of $\mathscr{R}$ containing h_0, we have in principle solved the problem of computing the vital map ∂ in terms of the homotopy properties of ρ. In practical terms it is most advantageous to represent $[B\chi_\rho\cdot u]$ cohomologically and the next step is an attempt to compute the primary and secondary classes associated with this map thus splitting the problem into two parts.

D. A representation of the primary class

We want to relate the properties of $\chi:\pi_1(\Sigma)\to\Omega G$ with $B\chi^*:H^1(G;\pi_1(G))\to H^1(B\pi;\pi_1(G))$. Now $H^1(G;\pi_1(G)) = \mathrm{Hom}(\pi_1(G),\pi_1(G))$ and $H^1(B\pi;\pi_1(G)) \approx \mathrm{Hom}(\pi_1(\Sigma),\pi_1(G))$ and it suffices to study $B\chi^*$ on the identity homomorphism ι_1. In this language the primary cohomology class of $B\chi$ is

Proposition 5.7: $(B\chi^*\iota_1)(\gamma) = [\chi(\gamma)]$ the homotopy class of $\chi(\gamma)$ in ΩG.

Proof: If $h\in\mathrm{Hom}(\pi_1(G),\pi_1(G))$ then $(B\chi^*(h))(\gamma) = h(B\chi_*(\gamma))$ for all γ in $\pi_1(\Sigma)$, where $B\chi_*$ maps $\pi_1(\Sigma)$ into $\pi_1(G)$. The homtopy exact sequence of the fibration [cf. (5.5)] $E\pi\times_\chi\Omega G\to B\pi \xrightarrow{B\chi} G$ is

$$\to\pi_1(E\pi\times_\chi\Omega G)\to\pi_1(\Sigma)\to\pi_1(G)$$
$$\to\pi_0(E\pi\times_\chi\Omega G)\to\pi_0(B\pi)\to *,$$

where,since$\pi_1(\Sigma)$isdiscrete,$B\pi_1(\Sigma) = K(\pi_1(\Sigma),1)$andhence $\pi_1(B\pi) = \pi_1(\Sigma)$. On the other hand, by mapping S^0 into the sequence[20] $\to\pi_1(\Sigma)\xrightarrow{\chi}\Omega G\to E\pi\times_\chi\Omega G\to B\pi$ we get

$$\to\pi_1(E\pi\times_\chi\Omega G)\to\pi_1(\Sigma)\xrightarrow{\chi_*}\pi_0(\Omega G)\to\pi_0(E\pi\times_\chi\Omega G)\to\pi_0(B\pi)$$

where χ_* is the composite map $\pi_1(\Sigma)\xrightarrow{\chi}\Omega G\xrightarrow{\epsilon}\pi_0(\Omega G)\approx\pi_1(G)$ with $\epsilon(\omega) = [\omega]$. Thus $B\chi_*(\gamma) = \epsilon\cdot\chi(\gamma) = [\chi(\gamma)]$.

Q.E.D.

We note that $u^*:H^1(B\pi;\pi_1(G))\to H^1(\Sigma;\pi_1(G))$ is the identity map from $\mathrm{Hom}(\pi_1(\Sigma),\pi_1(G))$ onto itself and hence

Corollary 5.8. The primary class in $H^1(\Sigma;\pi_1(G))$ representing $B\chi_\rho\cdot u$ is just the homomorphism $\gamma\rightsquigarrow[\chi(\gamma)]$ from $\pi_1(\Sigma)$ into $\pi_1(G)$. In particular the primary class of the endpoint C_0 of the curve C_λ must be $\gamma\rightsquigarrow[\chi(\gamma)]$.

As an illustration of Corollary 5.8 we can consider the case $G = \mathrm{U}1$, where there is no secondary class ($\pi_3(\mathrm{U}1) = 0$). Let A be any function from Σ into U1. Then A_* is a map from $\pi_1(\Sigma)$ into $\pi_1(\mathrm{U}1) = Z$ and we define $\rho\in\Omega\,\mathrm{Hom}(\pi_1(\Sigma),\mathrm{U}1)$ by $\rho(\lambda)(\gamma) = e^{i\lambda A_*(\gamma)} = \chi_\rho(\gamma)(\lambda)$. Then $A_*(\gamma) = [\chi(\gamma)]$ for all $\gamma\in\pi_1(\Sigma)$ and so for any $A\in\mathrm{U}1^\Sigma$ there exists a ρ such that $A\sim B\chi_\rho\cdot u$. Hence the boundary map ∂ is onto. Furthermore, $\pi_0(\mathscr{R}) = \{*\}$ and hence $\pi_0(\mathscr{D}) = \{*\}$ i.e., there is just a single vacuum sector in this canonical quantization scheme. Of course this is already obvious from the observation that the affine sum of two zero-energy U1 connections is itself zero energy.

More generally let A map Σ into any compact Lie group G and consider the exact sequence

$$\mathbb{1}\to(\Omega G)_0\to\Omega G\xrightarrow{\epsilon}\pi_1(G)\to 0, \tag{5.8}$$

where $(\Omega G)_0$ is the base component of ΩG. If A_* is the induced homomorphism from $\pi_1(\Sigma)$ into $\pi_1(G)$, we seek a homomorphism χ from $\pi_1(\Sigma)$ to ΩG such that $\in\cdot\chi = A_*$. Now $\pi_1(G)$ is abelian and hence A_* factors through $\pi_1(\Sigma)/[\pi_1(\Sigma),\pi_1(\Sigma)]\approx H_1(\Sigma;Z)$, which is a sum of b_1 (the first Betti number) copies of the integers Z plus a set of cyclic groups. It is clearly sufficient to find a homomorphism ϕ from $H_1(\Sigma;Z)$ into ΩG such that $\in\cdot\phi = \beta$, where $A_* = \beta\cdot\alpha$ and α is the canonical projection from $\pi_1(\Sigma)$ onto $H_1(\Sigma;Z)$. Now suppose for example that $\pi_1(G) = Z_q$ or Z with generator μ and choose $\omega\in\Omega G$ such that $\epsilon(\omega)\equiv[\omega] = \mu$. The map ϕ must vanish on the cyclic groups (proposition 5.1) and we define it on $\oplus_{i=1}^{b_1} Z$ by $\phi(n_1,\dots,n_{b_1}) := \omega_1^{\Sigma_{i=1}^{b_1} m_i n_i}$, where the integers m_i are defined by $A_*(0,0,\dots,1,0,\dots,0)$ (1 in the ith place) $= \mu^{m_i}$. Then $\epsilon\cdot\phi(n_1,\dots,n_{b_1}) = \mu^{\Sigma m_i n_i} = A_*(n_1,\dots,n_{b_1})$.

Hence, given any $A:\Sigma\to G$ and provided that $b_1 > 0$, we can find a χ such that the primary cohomology classes of A and $B\chi\cdot u$ are equal. Thus ∂ maps *onto* the primary classes in $[\Sigma,G]$ or, more precisely, there is a surjective map $\delta:[\Sigma,G]\to H^1(\Sigma;\pi_1(G))$ and $\delta\cdot\partial$ is surjective. For most groups (SO 3 is one exception) δ splits and $[\Sigma,G]$ can be expressed as a direct sum,[7] in which case ∂ maps onto the $H^1(\Sigma;\pi_1(G))$ subgroup. Thus in the holonomy sector C_{h_o} the primary classes drop out of the vacuum state classification. We do not know if this is true for the other sectors. Note that if $b_1 = 0$ then the primary classes may reappear in the classification.

We have been unable to find a general expression for the secondary winding number comparable to that in Proposition (5.7) although, since $\chi_1\sim\chi_2$ implies $B\chi_1\sim B\chi_2$, only the homotopy class of χ can be relevant. However, in a few special cases it is possible to show that the secondary class vanishes and this might always be so. For example, if $\pi_1(\Sigma)$ is a free group with generators $\gamma_1,\gamma_2,\dots,\gamma_n$ and χ is a homomorphism into ΩG, let $\omega_1,\dots,\omega_n$ be the loops in G such that $\chi(\gamma_i) = \omega_i$, $i = 1,\dots,n$. Now suppose that $\pi_1(G) = 0$. Then there are paths $\omega_i^{(t)}$ in ΩG such that $\omega_i^{(1)} = \omega_i$ and $\omega_i^{(0)} = \mathbb{1}$ (the trivial loop). Define $\chi^{(t)}(\gamma_{i_1}\gamma_{i_2}\cdots) = \omega_{i_1}^{(t)}\omega_{i_2}^{(t)}\cdots$ for any word $\gamma_{i_1}\gamma_{i_2}\cdots$. Then $t\rightsquigarrow\chi^{(t)}$ is a homotopy of χ with the trivial homomorphism and hence $B\chi$ is homotopically trivial. Thus the secondary winding number of $B\chi\cdot u$ vanishes identically.

A more interesting example is afforded by the case when $\pi_1(\Sigma)$ is abelian.

E. Abelian $\pi_1(\Sigma)$

Only prime manifolds can have an abelian $\pi_1(\Sigma)$ and the only possibilities,[17,24] with $\pi_1(\Sigma)$ infinite, are $\pi_1(\Sigma) = Z$ (eg., $S^1\times S^2$) or $Z\oplus Z\oplus Z$ (eg., $S^1\times S^1\times S^1$). Any homormorphism of Z into G necessarily factors through a U1 subgroup which in turn is mapped into G via a maximal torus. Thus if ρ is a loop in Hom$(\pi_1(\Sigma),G)$, ρ_λ factors as

$$Z \xrightarrow{a_\lambda} \mathrm{U1} \xrightarrow{e} T \xrightarrow{i_\lambda} G;$$

$$Z\oplus Z\oplus Z \xrightarrow{c_\lambda} \mathrm{U1}\times\mathrm{U1}\times\mathrm{U1} \xrightarrow{f_\lambda} T \xrightarrow{i_\lambda} G. \tag{5.9}$$

Now any two maximal tori are conjugate and hence χ_ρ factors as

$$Z \xrightarrow{a} \Omega\,\mathrm{U1} \xrightarrow{e} \Omega T \xrightarrow{\Omega i} \Omega G \xrightarrow{A\mathrm{d}p} \Omega G,$$

$$Z\oplus Z\oplus Z \xrightarrow{c} \Omega\,(\mathrm{U1}\times\mathrm{U1}\times\mathrm{U1}) \xrightarrow{f} \Omega T \xrightarrow{\Omega i} \Omega G \xrightarrow{\mathrm{Ad}p} \Omega G, \tag{5.10}$$

where $a(n)(\lambda) := a_\lambda(n)$, etc. and $(\mathrm{Ad}p)(\lambda) := \mathrm{Ad}(p_\lambda)$ for some curve p_λ in G. This latter term has no effect on $B\chi_\rho$ and we finally obtain

$$B\chi_\rho: BZ \xrightarrow{Ba} \mathrm{U1} \xrightarrow{Be} T \xrightarrow{i} G,$$

$$B(Z\oplus Z\oplus Z) \xrightarrow{Bc} \mathrm{U1}\times\mathrm{U1}\times\mathrm{U1} \xrightarrow{Bf} T \xrightarrow{i} G, \tag{5.11}$$

and we recall that $BZ\sim S^1$ and $B(Z\oplus Z\oplus Z)\sim S^1\times S^1\times S^1$.

Now suppose that G has rank 1 or 2, such as $G=$ U1, U2, SU2, SU3, SO3, SO4, SO5, etc.,; then $H^3(T;Z)=0$ and hence $B\chi_\rho^*$ must vanish on $H^3(\Sigma;\pi_3(G))$—there is no secondary number. On the other hand since $\dim\Sigma = 3$, simple obstruction theory[15,22] shows that

$$\begin{aligned}
[\Sigma,\mathrm{U}(n)] &= [\Sigma,\mathrm{U2}] \quad n\geqslant 3,\\
[\Sigma,\mathrm{SU}(n)] &= [\Sigma,\mathrm{SU2}] \quad n\geqslant 3,\\
[\Sigma,\mathrm{SO}(n)] &= [\Sigma,\mathrm{SO5}] \quad n\geqslant 6,
\end{aligned} \tag{5.12}$$

and so the secondary numbers of $B\chi_\rho$ vanish in these cases too.

F. Nonhomotopic C_λ paths

Let us conclude this study of the map ∂ and the paths C_λ by observing that in Eq. (5.6) homotopically *inequivalent* lifts may occur. The corresponding curves C_λ still link $\mathbb{1}$ with $A(x)$ but they cannot be deformed into each other. The classes of inequivalent lifts are labeled by $[\Sigma,\Omega G] = [S\Sigma,G]$ where $S\Sigma$ is the (reduced) suspension of Σ.[20,21] These homotopy classes may be expressed cohomologically using the type of Postnikov technique described in Ref. (7). Some sample results are

$$\begin{aligned}
[S\Sigma,\mathrm{SO5}] &= [S\Sigma,\mathrm{SU2}] = H^3(\Sigma;Z_2)\oplus H^2(\Sigma;Z)\\
&= Z_2\oplus H_1(\Sigma;Z),
\end{aligned} \tag{5.13}$$

$$[S\Sigma,\mathrm{SU}(n)] = H_1(\Sigma;Z), \quad n\geqslant 3, \tag{5.14}$$

$$[S\Sigma,\mathrm{SO}(n)] = H_1(\Sigma;Z), \quad n\geqslant 6, \tag{5.15}$$

and it is worth noting that even in the conventional S^3 case we get such an effect since $[S(S^3),\mathrm{SU2}] = \pi_4(\mathrm{SU2}) = Z_2$ and so there are two classes of curves linking $\mathbb{1}$ with A. On the other hand, $\pi_4(\mathrm{SU}(n)) = 0$ if $n\geqslant 3$ and in this case the phenomenon is absent. Thus there is a new type of "topological charge" associated with the SU2 vacuum states but its physical significance is unclear to us.

6. CONCLUSIONS

We have seen that the canonically-quantized Yang–Mills field on a general three-space Σ has a vacuum structure that differs significantly from the familiar one where $\Sigma = S^3$. New phenomena arise from the nonvanishing of $\pi_1(\Sigma)$ and the corresponding possibility of a nonvanishing discrete holonomy group giving zero energy solutions that are not pure gauge. These can either increase or decrease the naive $[\Sigma,G]$ classification by respectively increasing the holonomy sector or by permitting new zero-energy paths which enlarge the class of gauge functions that are physically equivalent.

We have identified the space of zero-energy solutions with the function space $\mathscr{D}^{(r)}$ $(r\geqslant 2)$ and classified the quantum vacua by $\pi_0(\mathscr{D})$. A crucial result is the exact homotopy sequence (4.1) and the ensuing enumeration by $q_*^{-1}(C_h) = [\Sigma,G]/\partial\pi_1(\mathscr{R},h)$ of the states associated with the holonomy sector C_h. States and transition amplitudes can be diagonalized using the characters of this group and in Sec. 5 we have presented a number of results on $\partial\pi_1(\mathscr{R},h)$. The major problems that remain to be solved are:

(1) The calculation of the secondary winding number of $[B\chi_\rho\cdot u]$ in terms of the homomorphism ρ for a general $\pi_1(\Sigma)$ (i.e., other than abelian).

(2) The derivation of analogous results in holonomy sectors other than C_{h_0}.

(3) If (1) is impossible in general it would be useful to solve it for at least a selection of prime manifolds such as those having nilpotent or solvable fundamental groups.[17,25,26]

(4) An exhibition of a definite example in which the secondary winding number of $[B\chi_\rho\cdot u]$ is nonvanishing. This would drastically change the appearance of the θ vacua. Alternatively, one would like a proof that this number vanishes for all $\pi_1(\Sigma)$.

(5) The entire construction should be repeated for the case where the G bundle over Σ is nontrivial.

It is possible to write down analogs of the lifting problem (5.5) which apply to an arbitrary holonomy sector, but it is difficult to recast the information into a usable cohomological form. A useful tool in this respect is the "dual" version of (5.5) or its extensions. We have the dual relation $\Omega G^{\Sigma} = G^{S\Sigma}$ and, for example, Eq. (5.5) dualizes as

$$\begin{tikzcd}
 & (E\pi \wedge \mathscr{R})_T \times G \arrow[d] \\
S\wedge\Sigma \equiv S\Sigma \arrow[ur, dashed, "\phi"] \arrow[r, "\rho\wedge u"] & B\pi \wedge \mathscr{R},
\end{tikzcd} \tag{6.1}$$

where $\wedge$ denotes the smash product of pointed spaces and the equivalence relation on $(E\pi\wedge\mathscr{R})\times G$ is $(e,h,g)\equiv(e\gamma,h,h(\gamma)^{-1}g)$.

(cf. $(\hat{\Sigma}\times V)\times_T G$ in Sec. 3.B). A lift ϕ is necessarily of the form $\phi[\lambda,x] = [\bar{u}(y),\rho(\lambda),E(\lambda,y)]$ and thus we recover the $E(\lambda,y)$ function.

We hope to return to these problems in a later paper as well as to the question of the analogous effects in the canonical quantization of the gravitational field.

APPENDIX

Proposition A.1: $\pi_0(\mathscr{D}^{(r)}) = \pi_0(\mathscr{D})$ for all r where $\mathscr{D}\equiv\mathscr{D}^{(0)}$.

Proof: There is a natural map i_* from $\pi_0(\mathscr{D}^{(r)})$ into $\pi_0(\mathscr{D})$ which we wish to show is a bijection.

(a) To prove that i_* is surjective it suffices to show that any element of $\mathscr{D}$ is homotopic to some element of $\mathscr{D}^{(r)}$. We know, however, that there is a one-to-one correspondence between elements of $\mathscr{D}$ induced by a given homomorphism $h\in\mathscr{R}$ and cross sections ϕ of the C^∞ fiber bundle $\hat{\Sigma}\times_h G\to\Sigma$ with $\phi(x) = [y,D(y)^{-1}]$. Every continuous cross section of a C^∞ fiber bundle over a compact base space is homotopic to a C^r section[27,28] and hence for every element of $\mathscr{D}$ there exists a suitable element of $\mathscr{D}^{(r)}$.

(b) To show that i_* is injective, we must show that any two elements of D_0 and D_1 of $\mathscr{D}^{(r)}$ which can be joined by a path of D^λ in $\mathscr{D}$ can also be joined by a path in $\mathscr{D}^{(r)}$. First we triangulate the real analytic set $\mathscr{R}$ with h_0 and h_1 (the homorphisms corresponding to D_0 and D_1) as two of the vertices. We deform the projection of $\lambda\rightsquigarrow D_\lambda$ in $\mathscr{R}$ into a path $\lambda\rightsquigarrow h_\lambda$ which lies entirely in the one-skeleton of $\mathscr{R}$. Since $\mathscr{D}$ is a fiber bundle (Proposition 3.10) the homotopy lifting property[20,22] guarantees that D_λ can be simultaneously deformed into a new path $\tilde{D}_\lambda$ in $\mathscr{D}$ such that $\tilde{D}_0 = D_0$ and $\tilde{D}_1$ is homotopic to D_1. Thus $\tilde{D}_1(y) = \Omega(x)D_1(y)$ with $\Omega(x)$ homotopic to $\mathbb{1}$ through a one-parameter family $\lambda\rightsquigarrow\Omega_\lambda(x)$ of C^r functions. Then $\lambda\rightsquigarrow\Omega_{1-\lambda}^{-1}(x)\tilde{D}_\lambda(y)$ is a path of continuous functions whose initial and final points equal D_0 and D_1. Thus without loss of generality we can assume that $\lambda\rightsquigarrow D_\lambda$ covers the deformed path $\lambda\rightsquigarrow h_\lambda$ in the base space $\mathscr{R}$.

The curve h_λ intersects the singular set in $\mathscr{R}$ at a finite number of vertices which we will label with the corresponding values of λ so that $0<\lambda_1<\lambda_2<\cdots<\lambda_n<1$ (h_0 and/or h_1 could also be singular points). Consider the interval $I_1 = [0,\lambda_1]$ and construct the trivial principal G bundle $(\hat{\Sigma}\times I_1)_T\times G$ over $\Sigma\times I_1$ with the equivalence relation $(y,\lambda,g)\equiv(y\gamma,\lambda,h_\lambda^{-1}(\gamma)g)$ and projection $t:[y,\lambda,g]\rightsquigarrow(r(y),\lambda)$. A continuous cross section is of the form $\psi(x,\lambda) = [y,\lambda,D_\lambda^{-1}(y)]$ and, because $\Sigma\times I_1$ is a C^∞ manifold (with boundary), there exists a C^r cross section[29] $\psi^{(r)}(x,\lambda)$ such that $d(D_\lambda(y),D_\lambda^{(r)}(y))<\epsilon$ for all $\lambda\in[0,\lambda_1]$ and with ϵ sufficiently small that $D_\lambda\sim D_\lambda^{(r)}$. Here $d(\ ,\ )$ is a metric on G.

Now $D_0^{(r)}(y) = \Phi(x)D_0(y)$ for some C^r function $\Phi(x)$ which is homotopic to $\mathbb{1}$ through a family $t\rightsquigarrow\Phi_t$ of C^r functions. We can replace $D_\lambda^{(r)}(y)$ by the family of C^r functions $\tilde{D}_\lambda^{(r)}(y) = \Phi_{(1-\lambda)/\lambda_1}^{-1}D_\lambda^{(r)}(y)$ which satisfies $\tilde{D}_0^{(r)} = D_0$ and of course $d(D_1(y),\tilde{D}_1^{(r)}(y))<\epsilon$. We can repeat this procedure for the finite set of intervals $\{[\lambda_i,\lambda_{i+1}]\}$ and end with a continuous path $\lambda\rightsquigarrow\tilde{D}_\lambda(r)$ of C^r functions with the property that, by choosing ϵ small enough, $\tilde{D}_1^{(r)}\sim D_1$. Once again one uses a family Ω_t of C^r functions in order to set $\tilde{D}_1^{(r)} = D_1$ while leaving alone the condition $\tilde{D}_0^{(r)} = D_0$. This final curve in $\mathscr{D}^{(r)}$ is the one we seek.

Q.E.D.

Proposition A.2. There is a universal ΩG bundle whose base space is G.

Proof: (a) Let PG denote the space of paths in G, whose endpoint is $\mathbb{1}$, equipped with the compact open topology. Then PG is a topological group and ΩG is a closed normal subgroup. Form the canonical ΩG bundle

$\Omega G\to PG\xrightarrow{\alpha} PG/\Omega G$, where $PG/\Omega G$ is given the usual quotient topology (which is compatible with its group structure). The projection map α is a continuous, open, homomorphism with kernel ΩG.

We define $\pi:PG\to G$ by $\pi(p) = p(0)$ and $t:PG/\Omega G\to G$ by $t[p] = p(0)$. Then π and t are continuous and t is a bijection. However, π is a homomorphism from the metrizable group PG to the compact group G and hence π is open [Ref. (30) p. 98] which in turn implies t is open. Thus t is a homeomorphism and we obtain the principal ΩG bundle

$\Omega G\to PG\xrightarrow{\pi} G.$

(b) To prove local triviality, let U be a neighborhood of $\mathbb{1}\in G$ such that the exponential map is a diffeomorphism onto U. Define a map $\sigma:U\to\pi^{-1}(U)$ by $\sigma(g) = \{g_t\}$ where $t\rightsquigarrow g_t$ is the unique geodesic in U with $g_0 = g$ and $g_1 = \mathbb{1}$. Then σ is a continuous local section and local sections can be constructed everywhere by pulling σ around $G\approx PG/\Omega G$ with the PG action.

(c) PG is a contractible space and hence the locally trivial bundle $\Omega G\to PG\to G$ is a univeral ΩG bundle. Thus G is a model for $B(\Omega G)$.

Q.E.D.

[1] *Proceedings of the 1980 Cambridge Workshop in Supergravity* (Cambridge U. P., Cambridge, England, 1981).

[2] R. Jackiw and C. Rebbi, Phys. Rev. Lett. **37**, 172 (1976); C. G. Callen, R. F. Dashen, and D. T. Gross, Phys. Lett. B **63**, 334 (1976).

[3] N. Steenrod, *The Topology of Fiber Bundles* (Princeton U. P., Princeton, N.J., 1951).

[4] S. J. Avis and C. J. Isham, in *Recent Developments in Gravitation—Cargese 1978*, edited by M. Levy and S. Deser (Plenum, New York, 1979).

[5] The effects of holonomy on gauge theories has been discussed in a different context: M. Asorey, J. Math. Phys. **22**, 179 (1981).

[6] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, Vol. I (Interscience, New York, 1963).

[7] C. J. Isham, in *Essays in honour of Wolfgang Yourgrau*, edited by A. van

der Merwe (Plenum, New York, to appear in 1981).

[8] C. J. Isham and G. Kuntstatter, "Yang–Mills canonical vacuum structure in a general three-space," Phys. Lett. B **102**, 417 (1981).

[9] J. W. Milnor, Commun. Math. Helv. **32**, 215 (1957).

[10] F. Kamber and P. Tondeur, *Lecture Notes in Mathematics*, Vol. 67 (Springer, New York, 1968).

[11] A. Weil, Ann. Math. **72**, 369 (1960).

[12] V. N. Gribov, Nucl. Phys. B **139**, 1 (1978).

[13] I. Singer, Commun. Math. Phys. **60**, 7 (1978).

[14] P. K. Mitter and C. M. Viallet, "On the bundle of connections and the gauge orbit manifold in Yang–Mills theory," CNRS preprint (1979).

[15] H. J. Baues, *Lecture Notes in Mathematics*, Vol. 28 (Springer, New York, 1977).

[16] See for example R. Rajaraman, Phys. Rep. **21**, 227 (1975).

[17] J. Hempel, *3-manifolds* (Princeton University, New Jersey, 1976).

[18] J. Dugundje, *Topology* (Allyn and Bacon, Boston, 1966).

[19] F. Kamber and P. Tondeur, Am. J. Math. **89**, 857 (1967).

[20] R. M. Switzer, *Algebraic Topology-Homotopy and Homology* (Springer, New York, 1975).

[21] B. Grey, *Homotopy Theory* (Academic, New York, 1975).

[22] E. H. Spanier, *Algebraic Topology* (McGraw-Hill, New York, 1967).

[23] A. Borel, *Lecture Notes in Mathematics*, Vol. 36 (Springer, New York, 1967).

[24] D. Epstein, Quart. Math. Oxford **12**, 205 (1961).

[25] C. Thomas, Proc. Cambridge Philos. Soc. **64**, 303 (1968).

[26] B. Evans and L. Moser, Trans. Am. Math. Soc. **168**, 189 (1972).

[27] J. T. Schwarz, *Differential Geometry and Topology* (Gordon and Breach, New York, 1968).

[28] J. R. Munkres, *Elementary Differential Topology* (Princeton U. P., Princeton, N. J., 1966).

[29] R. Palais, *Foundations of Global Nonlinear Analysis* (Benjamin, New York, 1968).

[30] T. Husain, *Introduction to Topological Groups* (Saunders, Philadelphia, 1966).

Boson operator representation of Brownian motion

Ulrich R. Steiger
Department of Chemistry, University of California, Davis, California 95616
Ronald F. Fox
School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 15 January 1982; accepted for publication 12 March 1982)

In the framework of the classical theory of Brownian motion the time evolution of the distribution function in the full phase space of a particle immersed in a fluid is governed by a Fokker–Planck equation. The reduced distribution function in coordinate space fulfills the Smoluchowski equation in first approximation. This work improves previous derivations by including higher order corrections and by using an expansion which permits the discussion of the size of the error made by truncating the infinite series. The derivation is based on the adaption of a powerful mathematical tool used in quantum field theory: The Fokker–Planck equation is written in terms of boson operators. Conditional equilibrium averages of operators are defined which play the role of vacuum expectation values. The time-ordered cumulant expansion is used to calculate the formal diffusion operator in terms of conditional equilibrium averages of powers of the "Liouville operator in the interaction picture." It is shown that all these averages can be obtained from a *Gaussian* generating functional which is explicitly calculated using the time-ordered version of Glauber's theorem. The resulting diffusion equation, a fourth order partial differential equation in the position space, is obtained by calculating the cumulant expansion up to sixth order. Conditions on the potential are established which guarantee that these equations are dissipative and it is shown that all solutions approach the Boltzmann distribution as $t \rightarrow \infty$. Curvilinear, non-Euclidean coordinates are introduced in order to interpret these diffusion equations. Nonlinear diffusion equations and their application regarding the self-avoiding random walk are discussed.

PACS numbers: 05.40. + j

1. INTRODUCTION

The motion of a particle immersed in a fluid is governed by the Smoluchowski equation. This is the time evolution equation for the reduced probability density $P(t,q)$ as a function of the time t and the position q. Under the influence of an external potential U and the fluid the reduced probability density changes with time according to

$$\frac{\partial}{\partial t} P(t,q) = \frac{\partial}{\partial q} \cdot D \left(\frac{\partial}{\partial q} + \frac{1}{kT} \frac{\partial U}{\partial q} \right) P(t,q). \tag{1}$$

D denotes the diffusion constant, T is the temperature, and k is the Boltzmann constant.

A simple example shows that the Smoluchowski equation cannot be completely correct in general. The Smoluchowski equation is an approximation in contrast to Einstein's result, which holds in absence of external forces and turns out to be exact[1]—compare Sec. 4.

For instance, consider the harmonic oscillator under the influence of random forces described by the Kramers–Liouville equation—a completely solved problem.[2] The first moment of the reduced probability density is[3]

$$\langle q(t) \rangle = \langle q(0) \rangle e^{-(c/2m)t} \left(\cosh \omega_1 t - \frac{c}{2m\omega_1} \sinh \omega_1 t \right)$$

with the friction coefficient c, the mass m, the frequency ω, and $\omega_1 = (c^2/4m^2 - \omega^2)^{1/2}$. The time dependence of $\langle q(t) \rangle$ is in good agreement with the Smoluchowski equation for the reduced probability density $P(t,q)$.

$$\frac{\partial}{\partial t} P(t,q) = \frac{\partial}{\partial q} \Lambda \left(\frac{\partial}{\partial q} + q \frac{m\omega^2}{kT} \right) P(t,q) \tag{2}$$

for large times, $t \gg \omega_1^{-1}$, if Einstein's diffusion coefficient $\Lambda_0 = kT/c$ is replaced by the ω-dependent expression

$$\Lambda = \Lambda(\epsilon) = \Lambda_0 \frac{1 - (1 - 4\epsilon)^{1/2}}{2\epsilon}, \tag{3}$$

with $\epsilon = (m\omega/c)^2$. In the limit as $\epsilon \rightarrow 0$, $\Lambda(\epsilon)$ approaches Λ_0. As an infinite series

$$\Lambda(\epsilon) = \Lambda_0(1 + \epsilon + 2\epsilon^2 + \cdots). \tag{4}$$

This example shows that, in general, the diffusion equation contains higher order corrections which depend on the potential.

A very good survey of previous derivations of the Smoluchowski equation (and corrections) is contained in Ref. 4. The present work improves these derivations in three ways. First, the boson operator representation and the introduction of appropriate conditional equilibrium averages reduces the calculation to a purely algebraic problem. Secondly, we are able to show that all necessary averages in the momentum space can be obtained from a Gaussian generating functional. This important result makes it possible to calculate the time-ordered cumulant expansion up to sixth order, which leads to new corrections of previously published diffusion equations. Thirdly, the expansion used here is physically motivated though all calculations are exact and do not contain any further approximations. It is necessary to keep the physical picture in mind because some mathematical manipulations are only formally correct. In our picture the correlation of the momenta can be viewed as a particle which decays exponentially.

The starting point of our discussion is the Fokker–

0022-2488/82/091678-10$02.50

Planck equation for translational Brownian motion, which is called the Kramers–Liouville equation in the following. We consider the Brownian motion only in the Markovian limit.[5]

2. CUMULANT EXPANSION OF THE OPERATOR G

The time-ordered cumulants give an explicit expansion of the formal diffusion operator G describing the time evolution of the reduced probability density $P(t,q)$,

$$\frac{\partial}{\partial t}P(t,q) = G(t,q)P(t,q),$$

$$G(t,q) = \sum_{n=1}^{\infty} G^{(n)}(t,q). \tag{5}$$

The description of the motion of a particle moving in a fluid leads to the Kramers–Liouville equation.[2] This serves as a starting point for the description of a particle influenced by an arbitrary external potential $U(q)$ and a "Brownian fluid." The fluid is considered to be composed of particles which exert a fluctuating force with vanishing mean on the particle immersed in the fluid and it is assumed that this force has a white spectrum. The only constants entering the description are the Boltzmann constant k, the temperature T with $\beta \equiv (kT)^{-1}$, the mass m, and a friction coefficient c which depends on the size of the particle and the viscosity of the fluid.

The Kramers–Liouville equation is a first order partial differential equation describing the time evolution of the probability density $f(t,q,p)$ which depends on the time t, the position q, and the canonical conjugate momenta p. q and p are vectors in $\mathbb{R}$. the Kramers–Liouville equation is

$$\frac{\partial}{\partial t}f(t,q,p) = (L+K)f(t,q,p), \tag{6}$$

$$L = -m^{-1}p\cdot\frac{\partial}{\partial q} + \frac{\partial U}{\partial q}\cdot\frac{\partial}{\partial p}, \tag{7}$$

$$K = c\frac{\partial}{\partial p}\cdot\left(m^{-1}p + \beta^{-1}\frac{\partial}{\partial p}\right). \tag{8}$$

L is Liouville's operator and K denotes Kramers' operator, which describes the effect of the random force acting on the Brownian particle. Equation (6) describes how the initial distribution, given by a function $f_0 = f(0,q,p)$, changes in time. Any initial distribution will approach the Maxwell–Boltzmann distribution g_{MB} in the limit as $t\to\infty$. With the partition function Z,

$$g_{\mathrm{MB}}(q,p) = Z^{-1}e^{-\beta[p^2/2m + U(q)]}. \tag{9}$$

For a large particle moving in a dense fluid, the main contribution in Eq. (6) is due to the Kramers operator. The Kramers operator forces the relaxation of the momentum distribution to the Maxwell distribution g_{M},

$$g_{\mathrm{M}}(p) = (2\pi/\beta)^{-3/2}e^{-\beta(p^2/2m)}. \tag{10}$$

In order to calculate that diffusion operator G of the time evolution equation

$$\frac{\partial}{\partial t}P = GP \tag{11}$$

for the reduced probability distribution $P(t,q)$,

$$P(t,q) = \int_{\mathbf{R}^3} d^3p\, f(t,q,p), \tag{12}$$

an approach very often used in quantum mechanics turns out to be useful here, too. In analogy to the interacting picture used, for instance, for the quantum mechanical treatment of radiation, the time dependent transformation

$$f(t) \equiv e^{tK}\tilde{f}(t) \tag{13}$$

leads to the Kramers–Liouville equation in the "interaction picture",

$$\frac{\partial}{\partial t}\tilde{f} = e^{-tK}Le^{tK}\tilde{f} \equiv \tilde{L}(t)\tilde{f}. \tag{14}$$

In order to calculate this new operator $\tilde{L}(t)$, and also for the further analysis, it is useful to introduce the following operators:

$$a \equiv -\left(\frac{m}{\beta}\right)^{1/2}\frac{\partial}{\partial p} - \left(\frac{\beta}{m}\right)^{1/2}p, \tag{15}$$

$$a^{\dagger} \equiv \left(\frac{m}{\beta}\right)^{1/2}\frac{\partial}{\partial p}, \tag{16}$$

$$b \equiv -\left(\frac{1}{m\beta}\right)^{1/2}\frac{\partial}{\partial q} - \left(\frac{\beta}{m}\right)^{1/2}\frac{\partial U}{\partial q}, \tag{17}$$

$$b^{+} \equiv \left(\frac{1}{m\beta}\right)^{1/2}\frac{\partial}{\partial q}. \tag{18}$$

These operators consist of three components, e.g., $a = (a_1,a_2,a_3)$. The operators a and $a^{\dagger}$ are dimensionless, but the operators b and b^{+} have the dimension of an inverse time. The Liouville operator L and the Kramers operator K are

$$L = b^{+}\cdot a - b\cdot a^{\dagger}, \tag{19}$$

$$K = -c/ma^{\dagger}\cdot a. \tag{20}$$

The dot "·" denotes the usual scalar product of two vectors with three components. The commutator algebra of the components of the operators a, $a^{\dagger}$, b and b^{+} is

$$\begin{aligned}
&[a_i,a_j] = [a_i^{\dagger},a_j^{\dagger}] = 0,\\
&[a_i,a_j^{\dagger}] = \delta_{ij},\\
&[a_i,b_j] = [a_i,b_j^{\dagger}] = [a_i^{\dagger},b_j] = [a_i^{\dagger},b_j^{\dagger}] = 0,\\
&[b_i,b_j] = [b_i^{\dagger},b_j^{\dagger}] = 0,\\
&[b_i,b_j^{\dagger}] = m^{-1}\frac{\partial^2 U}{\partial q_i\partial q_j}.
\end{aligned} \tag{21}$$

The Liouville operator in the interaction picture is given by the infinite series

$$\tilde{L}(t) = e^{-tK}Le^{tK} = \sum_{n=0}^{\infty}\frac{t^n}{n!}A^{(n)}. \tag{22}$$

The operators $A^{(n)}$ are defined by recursion, $A^{(0)} = L$ and $A^{(n+1)} = [A^{(n)},K]$. This identity is proved in Ref. 6. The commutator algebra of the components of the operators a and $a^{\dagger}$—a three dimensional Weyl algebra—and the fact that all a-operators commute with all b-operators, Eq. (21), leads immediately to

$$A^{(n)} = \left(\frac{-c}{m}\right)^n b^{+}\cdot a - \left(\frac{c}{n}\right)^n b\cdot a^{\dagger}. \tag{23}$$

Substituting these expressions into the sum on the right-hand side of Eq. (22) gives

$$\begin{aligned}&\tilde{L}(t) = b^{\dagger}(t)\cdot a - b(t)\cdot a^{\dagger},\\ &b^{\dagger}(t) \equiv b^{\dagger} e^{-(c/m)t},\\ &b(t) \equiv b e^{(c/m)t}.\end{aligned} \tag{24}$$

In order to take full advantage of the properties of the operators introduced by Eqs. (15)–(18) it is useful to consider the scalar products

$$\langle u,v\rangle \equiv \int_{\mathbf{R}^3} d^3p\, g_{\mathrm{M}}^{-1} u^*(p) v(p), \tag{25}$$

$$\langle\langle f,h\rangle\rangle \equiv \int_{\mathbf{R}^3} d^3q \int_{\mathbf{R}^3} d^3p\;\; g_{\mathrm{MB}}^{-1} f^*(q,p) h(q,p). \tag{26}$$

The inverse of the Maxwell distribution, respectively, the Maxwell–Boltzmann distribution, serves as weight function. Consequently

$$\langle au,v\rangle = \langle u,a^{\dagger}v\rangle, \tag{27}$$

$$\langle\langle af,h\rangle\rangle = \langle\langle f,a^{\dagger}h\rangle\rangle, \tag{28}$$

$$\langle\langle bf,h\rangle\rangle = \langle\langle f,b^{\dagger}h\rangle\rangle. \tag{29}$$

These identities hold for all functions which vanish sufficiently rapidly at infinity; they are proved by integrating by parts. The operators $a^{\dagger}$ and $b^{\dagger}$ are the formal adjoint operators of a, respectively b.

The definition of the scalar product $\langle\cdot,\cdot\rangle$, Eq. (25), can be extended to functions which depend also on q.

$$\begin{aligned}\langle f,h\rangle &\equiv \int_{\mathbf{R}^3} d^3p\, g_{\mathrm{M}}^{-1} f^*(q,p) h(q,p)\\ &= \langle f,h\rangle(q).\end{aligned} \tag{30}$$

We will see that the diffusion operator G can be expressed in terms of a sum of products of operators of the form

$$\int_{\mathbf{R}^3} d^3p \tilde{L}(t_1)\cdots\tilde{L}(t_n) g_{\mathrm{M}}. \tag{31}$$

Keeping the definition (30) in mind, this expression can be interpreted as the *conditional equilibrium average* of the nth order product $\tilde{L}(t_1)\cdots\tilde{L}(t_n)$,

$$\begin{aligned}\langle\tilde{L}(t_1)\cdots\tilde{L}(t_n)\rangle &\equiv \langle g_{\mathrm{M}},\tilde{L}(t_1)\cdots\tilde{L}(t_n) g_{\mathrm{M}}\rangle\\ &= \int_{\mathbf{R}^3} d^3p \tilde{L}(t_1)\cdots\tilde{L}(t_n) g_M.\end{aligned} \tag{32}$$

The cumulant expansion, which will be used to calculate the diffusion operator G, is based on the following theorem:

Theorem 1 (Theorem on time-ordered exponentials): For an arbitrary operator $M(t)$, the nth cumulant average $\langle M(t_1)\cdots M(t_n)\rangle_c$ for $t_1 > t_2 > \cdots > t_n$ is defined by

$$\langle M(t_1)\cdots M(t_n)\rangle_c \equiv \sum_{\text{partitions of } n} (-1)^{k-1} \sum_{p} \prod_{s=1}^{k} \left\langle \prod_{l=1}^{n_{p(s)}} M(t_{i_{p(s),l}}) \right\rangle. \tag{33}$$

The sum runs over all ordered partitions of the first n integers in k subsets $(i_{11},\dots,i_{1n_1})\cdots(i_{k1},\dots,i_{kn_k})$, with $i_{lr} < i_{ls}$ for $r < s$, all l, and $i_{11} = 1$. p is a permutation of the integers $2,\dots,k$ and $p(1) \equiv 1$. Then the following identity holds:

$$\begin{aligned}&\left\langle \underset{\leftarrow}{T} \exp \int_0^t ds\, M(s) \right\rangle\\ &\quad = \underset{\leftarrow}{T} \exp \int_0^t ds \left\{ \left\langle \underset{\leftarrow}{T} \exp \int_0^s ds' M(s') \right\rangle_c - \mathbb{1} \right\}.\end{aligned} \tag{34}$$

The time-ordered exponential is viewed as a formal power series,

$$\underset{\leftarrow}{T} \exp \int_0^t ds\, M(s) = \mathbb{1} + \sum_{n=1}^{\infty} \int_0^t ds_1 \cdots \int_0^{s_{n-1}} ds_n M(s_1)\cdots M(s_n). \tag{35}$$

Equation (34) is provide in Ref. 7.

The connection between the operator cumulants $G^{(l)}$ and the cumulant average (33) is

$$G^{(n)}(t_1) = \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n \langle M(t_1)\cdots M(t_n)\rangle_c. \tag{36}$$

The complete cumulant expansion $G(t) = \sum_{n=1}^{\infty} G^{(n)}(t)$ can be written as the cumulant average of the time-ordered exponential of the operator $M(t)$,

$$G(t) = \left\langle \underset{\leftarrow}{T} \exp \int_0^t ds\, M(s) \right\rangle_c - \mathbb{1}.$$

Again, the time-ordered exponential is viewed as a formal power series in the operator $M(t)$. The cumulant average of powers of the operator $M(t)$ of the form $\langle M(t_1)\cdots M(t_n)\rangle_c$ is given by Eq. (33).

The Theorem on time-ordered exponentials shows that the cumulant expansion leads indeed to the time evolution equation for the reduced probability density $P(t,q)$ if the initial momentum distribution is Maxwellian $f(0,q,p) \equiv g_{\mathrm{M}} P(0,q)$. [This is only an apparent restriction since, as it turns out, the diffusion operator $G(t)$ is independent of the initial momentum distribution in the limit as $t \to \infty$]. We have

$$\begin{aligned}P(t,q) &= \langle g_{\mathrm{M}}, f\rangle = \langle g_{\mathrm{M}}, e^{tK}\tilde{f}\rangle = \left\langle e^{(-tc/m)a^{\dagger}\cdot a} g_{\mathrm{M}}, \tilde{f}\right\rangle\\ &= \langle g_{\mathrm{M}}, \tilde{f}\rangle = \left\langle \underset{\leftarrow}{T} \exp \int_0^t ds\, \tilde{L}(s) \right\rangle (0,q)\\ &= \underset{\leftarrow}{T} \exp \int_0^t ds\, G(s) P(0,q).\end{aligned} \tag{37}$$

Hence, the reduced probability density $P(t,q)$ fulfills the evolution equation $(\partial/\partial t)P = GP$.

Sometimes it is useful to have a recursive definition of the cumulant averages (for instance, for the proof of the cluster property for cumulants). The nth cumulant average $\langle M(t_1)\cdots M(t_n)\rangle_c$ can be expressed in terms of the average $\langle M(t_1)\cdots M(t_m)\rangle_c$ with $m < n$.

Theorem 2 (Inversion formula[8,9]):

$$\begin{aligned}&\langle M(t_1)\cdots M(t_n)\rangle\\ &= \sum_{\text{partitions of } n} \langle M(t_{i_{11}})\cdots M(t_{i_{1n_1}})\rangle_c \cdots \langle M(t_{i_{k1}})\cdots M(t_{i_{kn_k}})\rangle_c.\end{aligned} \tag{38}$$

The sum runs over all partitions of the first n integers in k subsets $(i_{11},\dots,i_{1n_1})\cdots(i_{k1},\dots,i_{kn_k})$. The subsets are ordered $i_{ls} < i_{lr}$ for $s < r$. The order of the cumulant average is determined by the condition $i_{l1} < i_{l'1}$ for $l < l'$.

Proof: Using the Theorem on time-ordered exponentials, one obtains a functional equation. For all functions

$$\lambda(s):\mathbb{R}^+\rightarrow C,$$

$$\langle \underset{\leftarrow}{T} \exp \int_0^t ds\, \lambda(s) M(s) \rangle$$
$$= \underset{\leftarrow}{T} \exp \int_0^t ds \Big\{ \langle \underset{\leftarrow}{T} \exp \int_0^s ds'\, \lambda(s') M(s') \rangle_c - \mathbf{1} \Big\}. \tag{39}$$

Taking the functional derivatives $\delta^n / \prod_{i=1}^n \delta\lambda(t_i)$ and setting $\lambda(t_i) \equiv 0$ gives the desired result (38). This completes the proof of Theorem 2.

3. DIFFUSION EQUATIONS

In this section the cumulant expansion up to sixth order will be calculated. The first nonvanishing term, the second cumulant is exactly Smoluchowski's result.

Equations (33) and (36) give an expansion of the diffusion operator G in terms of the averages,

$$\langle \tilde{L}(t_1) \cdots \tilde{L}(t_n) \rangle. \tag{40}$$

The first step consists of calculating all necessary averages of this form. Though this procedure seems simple enough, the higher order corrections, the fourth, sixth, etc., cumulants, have never been calculated before because this calculation becomes very cumbersome. A breakthrough can be achieved by realizing the simple structure of the Liouville operator in the interaction picture: $\tilde{L}(t)$ is linear in the creation and destruction operators $a^\dagger$ and a. Their properties are summed up by

$$[a_i, a_j^\dagger] = \delta_{ij}, \tag{41}$$

$$\langle a_i u, v \rangle = \langle u, a_i^\dagger v \rangle, \tag{42}$$

$$a_i g_{\rm M} = 0. \tag{43}$$

However, all this means that for each i the operators a_i and $a_i^\dagger$ are a pair of destruction and creation operators formally identical with the boson operators used in quantum field theories.[10] The properties of these operators are much simpler than the properties of the original differential operator $\partial/\partial p$ and the multiplication operator p. They also allow the construction of an orthonormal basis for the momentum space,[11,12]

$$\psi_{n_1 n_2 n_3} \equiv \frac{(a_1^\dagger)^{n_1} (a_2^\dagger)^{n_2} (a_3^\dagger)^{n_3}}{(n_1)!^{1/2} (n_2)!^{1/2} (n_3)!^{1/2}} g_{\rm M}. \tag{44}$$

The Maxwell distribution $g_{\rm M}$ corresponds to the vacuum in quantum field theory.

The first cumulant $G^{(1)}$, Eq. (36), is the average of $\tilde{L}(t)$. This operator contains only first powers of a and $a^\dagger$. However, the expectation value of these operators vanishes, $\langle a \rangle = 0$ and $\langle a^\dagger \rangle = 0$. Hence $G^{(1)}(t) = \langle \tilde{L}(t) \rangle = 0$. The second term of the cumulant expansion (5) is according to Eq. (36) equal to $\int_0^t ds \langle \tilde{L}(t) \tilde{L}(s) \rangle$. Using the orthogonality relation $\langle \psi_n, \psi_{n'} \rangle = \delta_{n_1 n_1'} \delta_{n_2 n_2'} \delta_{n_3 n_3'}$ for the base functions (44) leads to $\langle \tilde{L}(t) \tilde{L}(s) \rangle = - b^\dagger(t) \cdot b(s)$.

$$G^{(1)}(t) = 0,$$
$$G^{(2)}(t) = - \int_0^t ds\, b^\dagger(t) \cdot b(s). \tag{45}$$

The second cumulant gives the first diffusion equation for translational motion. Evaluating the time integral and replacing the operators b and $b^\dagger$ by the original definitions (17), (18), and (24) leads to

$$\frac{\partial}{\partial t} P(t,q) \cong G^{(2)}(t) P(t,q)$$
$$= \frac{\partial}{\partial q} \cdot \Lambda(t) \Big(\beta \frac{\partial U}{\partial q} + \frac{\partial}{\partial q} \Big) P(t,q) \tag{46}$$
$$= \frac{\partial}{\partial q} \cdot \Lambda(t) e^{-\beta U} \frac{\partial}{\partial q} e^{\beta U} P(t,q), \tag{47}$$
$$\Lambda(t) = \frac{1 - e^{-(c/m)t}}{\beta c}.$$

The reduced distribution $P(t,q)$ fulfills the Smoluchowski equation (46). The diffusion coefficient $\Lambda(t)$ is time dependent and vanishes for $t \rightarrow 0$ because initially the momentum distribution is Maxwellian. At the beginning, there is no correlated motion because the average momentum $\langle p \rangle$ vanishes for $t = 0$. The diffusion coefficient becomes stationary after a short time (long, however,compared with m/c).

The calculation of the higher order corrections is based on a formula for the averages $\langle \tilde{L}(t_1) \cdots \tilde{L}(t_n) \rangle$, which shows that the operator $\tilde{L}(t)$, and a Gaussian random variable $\tilde{x}$ with mean zero, have an important property in common: All moments of $\tilde{x}$ can be expressed in terms of the second moment $\langle \tilde{x}^2 \rangle \equiv \sigma^2$. The odd moments vanish and the even moments are $\langle \tilde{x}^{2m} \rangle = 1 \cdot 3 \cdot \cdots (2m-1) \sigma^{2m}$. The following theorem represents a generalization of this Gaussian property. It is an important results of this work.

Theorem 3 (Gaussian Property)[13]:

For $t_1 > t_2 > \cdots > t_{2m}$,

$$\langle \tilde{L}(t_1) \tilde{L}(t_2) \cdots \tilde{L}(t_{2m} \rangle$$
$$= \sum_{\text{partitions of } 2m} \underset{\leftarrow}{T} \Big\{ \prod_{j=1}^m \langle \tilde{L}(t_{i_{2j-1}}) | \cdot | \tilde{L}(t_{i_{2j}}) \rangle \Big\}, \tag{48}$$

with $\langle \tilde{L}(t)| \equiv b^\dagger(t)$ and $|\tilde{L}(t)\rangle \equiv - b(t)$. The sum runs over all partitions of the first $2m$ integers in subsets $(i_1 i_2) \cdots (i_{2m-1} i_{2m})$ with $i_{2j-1} < i_{2j}$. The operators on both sides of Eq. (48) are time ordered.

Proof: The moments of the operator $\tilde{L}(t)$ can be calculated using the generating functional $\phi[\lambda]$,

$$\phi[\lambda] \equiv \Big\langle \underset{\leftarrow}{\dot{T}} \exp \int_0^\infty dt \int_{\mathbf{R}^3} d^3 q \int_{\mathbf{R}^3} d^3 q' \{\cdots\} \Big\rangle, \tag{49}$$
$$\{\cdots\} = [b^\dagger(t,q,q') \cdot a - b(t,q,q') \cdot a^\dagger] \lambda(t,q,q').$$

The distributions $b^\dagger(t,q,q')$ and $b(t,q,q')$ are obtained by applying the operator $b^\dagger(t)$, respectively, $b(t)$ on the delta function $\delta(q-q')$.

$$b^\dagger(t,q,q') \equiv b^\dagger(t) \delta(q-q'), \tag{50}$$
$$b(t,q,q') \equiv b(t) \delta(q-q').$$

The delta function $\delta(q-q')$ is considered as a function of q. If the support of λ is bounded, the integral in Eq. (49) is finite and hence $\phi[\lambda] < \infty$.

The moments of the operator $L(t)$ are generated by taking the functional derivatives of $\phi[\lambda]$ and setting λ to zero.

$$\langle\tilde L(t_1)\cdots\tilde L(t_n)\rangle f(q_0),$$

$$\int_{\mathbb{R}^3} d^3q_1\cdots\int_{\mathbb{R}^3} d^2q_n \frac{\delta^n}{\prod_{i=1}^{n}\delta\lambda(t_i,q_{i-1},q_i)}\phi[\lambda]\bigg|_{\lambda\equiv 0}^{f(q_n)}. \tag{51}$$

$\phi[\lambda]$ is a real valued functional, but taking the functional derivatives leads to a generalized function, a distribution in the variables $t_1,\ldots,t_n$, $q_0,\ldots,q_n$. This distribution multiplied with $f(q_n)$ gives, after integrating over $\mathbb{R}^{3n}$, the function $\langle\tilde L(t_1)\cdots\tilde L(t_n)\rangle f(q_0)$.

Since the argument of the exponential in the definition of $\phi[\lambda]$, Eq. (49), is linear in a and $a^\dagger$, the functional $\phi[\lambda]$ can be calculated using the time-ordered version of Glauber's theorem.[14] Along the lines of Ref. 15 one obtains

$$\begin{aligned}\phi[\lambda] &= \exp\left[-\int_0^\infty ds_1\int_0^{s_1} ds_2(\cdots)\right],\\ (\cdots) &= \int_{\mathbb{R}^3} d^3q\int_{\mathbb{R}^3} d^3q'\int_{\mathbb{R}^3} d^3q''\int_{\mathbb{R}^3} d^3q'''\\ &\quad\times\sum_{l=1}^{3}\{b_l^\dagger(s_1,q,q')\lambda(s_1,q,q')\\ &\quad\times b_l(s_2,q'',q''')\lambda(s_2,q'',q''')\}.\end{aligned} \tag{52}$$

The sum $\sum_{l=1}^{3}$ in Eq. (52) can also be written as the scalar product $b^\dagger(s_2,q,q')\cdot b(s_2,q'',q''')$. The argument of the exponential is a quadratic form in the test function λ. Taking the functional derivatives according to Eq. (51) gives, therefore, the sum of partitions in ordered subsets of two elements. For instance,

$$\begin{aligned}&\frac{\delta^4}{\prod_{i=1}^{4}\delta\lambda(t_i,q_{i-1},q_i)}\phi[\lambda]\bigg|_{\lambda=0}\\ &\quad= b^\dagger(t_1,q_0,q_1)\cdot b(t_2,q_1,q_2)b^\dagger(t_3,q_2,q_3)\cdot b(t_4,q_3,q_4)\\ &\qquad+ b^\dagger(t_1,q_0,q_1)\cdot b(t_3,q_2,q_3)b^\dagger(t_2,q_1,q_2)\cdot b(t_4,q_3,q_4)\\ &\qquad+ b^\dagger(t_1,q_0,q_1)\cdot b(t_4,q_3,q_4)b^\dagger(t_2,q_1,q_2)\cdot b(t_3,q_2,q_3).\end{aligned} \tag{53}$$

Integrating over the variables q_1 to q_4 gives

$$\begin{aligned}\langle L(t_1)\cdots L(t_4)\rangle &= \underbrace{b(t_1)^\dagger b(t_2)}\underbrace{b^\dagger(t_3)b(t_4)}\\ &\quad+ b^\dagger(t_1)b^\dagger(t_2)b(t_3)b(t_4) + \underbrace{b^\dagger(t_1)\underbrace{b^\dagger(t_2)b(t_3)}b(t_4)}.\end{aligned} \tag{54}$$

In this equation the lines indicate how the scalar products have to be calculated, e.g., the last term is equal to $\sum_{i,j} b_i^\dagger(t_1)b_j^\dagger(t_2)b_j(t_3)b_i(t_4)$. This illustration concludes the proof of the Theorem 3.

Equations (38) and (48) provide all tools needed for the following calculation of the fourth and sixth cumulant. Only the results are listed here; for more details see Ref. 16. The calculation is cumbersome but patience leads to the goal. The results are, adopting Einstein's convention of summation,

$$\begin{aligned}G^{(4)}(t) &= \left(1-\frac{2c}{m}te^{-(c/m)t}-e^{-(2c/m)t}\right)\\ &\quad\times\frac{\partial}{\partial q_i}\frac{m}{c^3}\frac{\partial^2U}{\partial q_i\partial q_j}\left(\beta^{-1}\frac{\partial}{\partial q_j}+\frac{\partial U}{\partial q_j}\right),\end{aligned} \tag{55}$$

$$\begin{aligned}&G^{(6)}(t)\\ &= \left(\frac{1}{2}-\frac{3c}{m}te^{-(c/m)t}-\frac{3c}{m}te^{-(2c/m)t}-\frac{1}{2}e^{-(3c/m)t}\right.\\ &\quad\left.-\frac{9}{2}e^{-(2c/m)t}+\frac{9}{2}e^{-(c/m)t}\right)\frac{m^2}{c^5}\{\cdots\}\\ &\quad+\left[\frac{1}{2}-\left(\frac{ct}{m}\right)^2e^{-ct/m}+\frac{2c}{m}te^{-ct/m}+\frac{c}{m}te^{-(2c/m)t}\right.\\ &\quad\left.-\frac{4c}{m}e^{-(c/m)t}+e^{-2(c/m)t}+\frac{5}{2}e^{-(2c/m)t}\right]\frac{m^2}{c^5}(\cdots),\\ \{\cdots\} &= \frac{\partial^2}{\partial q_i\partial q_j}U_{il}\\ &\quad\times\left(\beta^{-1}\frac{\partial}{\partial q_j}+\frac{\partial U}{\partial q_j}\right)\left(\beta^{-1}\frac{\partial}{\partial q_l}+\frac{\partial U}{\partial q_l}\right)\\ &\quad+\frac{\partial}{\partial q_i}U_{ij}U_{jl}\left(\beta^{-1}\frac{\partial}{\partial q_l}+\frac{\partial U}{\partial q_l}\right)\\ &\quad-\frac{\partial}{\partial q_i}U_{ij}\left(\beta^{-1}\frac{\partial}{\partial q_j}+\frac{\partial U}{\partial q_j}\right)\\ &\quad\times\frac{\partial}{\partial q_l}\left(\beta^{-1}\frac{\partial}{\partial q_l}+\frac{\partial U}{\partial q_l}\right)\\ (\cdots) &= \frac{\partial}{\partial q_i}\left(\beta^{-1}\frac{\partial}{\partial q_i}+\frac{\partial U}{\partial q_i}\right)\frac{\partial}{\partial q_j}U_{jl}\left(\beta^{-1}\frac{\partial}{\partial q_l}+\frac{\partial U}{\partial q_l}\right)\\ &\quad-\frac{\partial}{\partial q_i}U_{ij}\left(\beta^{-1}\frac{\partial}{\partial q_j}+\frac{\partial U}{\partial q_j}\right)\\ &\quad\times\frac{\partial}{\partial q_l}\left(\beta^{-1}\frac{\partial}{\partial q_l}+\frac{\partial U}{\partial q_l}\right)\\ &\quad+2\frac{\partial}{\partial q_i}U_{ij}U_{jl}\left(\beta^{-1}\frac{\partial}{\partial q_l}+\frac{\partial U}{\partial q_l}\right).\end{aligned} \tag{56}$$

The matrix U_{ij} stands for the second derivative $\partial^2U/\partial q_i\partial q_j$. In the limit as t goes to ∞, one obtains

$$G^4(\infty) = \frac{m}{c^3}\frac{\partial}{\partial q_i}U_{ij}\left(\beta^{-1}\frac{\partial}{\partial q_j}+\frac{\partial U}{\partial q_{ji}},\right) \tag{57}$$

and for the sixth cumulant

$$\begin{aligned}G^{(6)}(\infty) &= \frac{m^2}{c^5}\frac{\partial}{\partial q_i}\left[2U_{ij}U_{jl}+U_{ilj}U_j+\frac{1}{2\beta}U_{iljj}\right]\\ &\quad\times\left(\beta^{-1}\frac{\partial}{\partial q_l}+U_l\right)\\ &\quad+\frac{3m^2}{2c^5\beta}\frac{\partial}{\partial q_i}\left(U_{ilj}\frac{\partial}{\partial q_j}\right)\left(\beta^{-1}\frac{\partial}{\partial q_l}+U_l\right).\end{aligned} \tag{58}$$

Discussion

The higher order corrections $G^{(4)}$, $G^{(6)}$, etc., vanish in the absence of external forces. This follows from the fact that the cumulant averages $\langle\tilde L(t_1)\cdots\tilde L(t_n)\rangle_c$ vanish for $n>2$ if the operators b and $b^\dagger$ commute. In general, the second cumulant is exact for a multiplicative stochastic Gaussian process if the stochastic operators commute.[17] Einstein's result,

$$\frac{\partial}{\partial t}P(t,q) = \Lambda(t)\nabla^2P(t,q), \tag{59}$$

is therefore exact. The stationary value of the diffusion coefficient is obtained in the limit as $t\to\infty$, $D=\lim_{t\to\infty}\Lambda(t)=1/\beta c$.

In the presence of external forces the second cumulant leads in the limit as $t \rightarrow \infty$ precisely to the Smoluchowski equation. This equation can be interpreted as the conservation law of the current,

$$j \equiv JP, \tag{60}$$

$$J \equiv \frac{-1}{\beta c} \frac{\partial}{\partial q} - \frac{1}{c} \frac{\partial U}{\partial q}, \tag{61}$$

$$\frac{\partial}{\partial t} P \cong \frac{-\partial}{\partial q} \cdot j. \tag{62}$$

Adding the corrections given by the fourth cumulant, one obtains a diffusion equation with position dependent diffusion coefficients.

$$\frac{\partial}{\partial t} P \cong \frac{-\partial}{\partial q_k} D_{kl} j_l, \tag{63}$$

$$D_{kl} = \delta_{kl} + (m/c^2) U_{kl}. \tag{64}$$

This equation agrees with the results given in Ref. 4.

It can be shown that all solutions of the Smoluchowski equation (62) tend toward the equilibrium distribution, the Boltzmann distribution g_{B} as $t \rightarrow \infty$. This is not always true for Eq. (63) because the diffusion coefficients $D_{kl}/c\beta$ become negative if U_{kl} has a negative eigenvalue smaller than c^{-2}/m. Obviously this is nonsense. The diffusion equation (63) can only be applied if the potential is sufficiently smooth. If the curvature of the potential U is very small, Eq. (63) reduces to the Smoluchowski equation. If the curvature is very large the new equation leads to significant changes of the solutions. On the other hand, if U_{ij} is very large the higher order corrections may become more and more important. In this case it is also questionable if the cumulant expansion converges. Equation (63) can be tested directly for simple cases, for example, the harmonic oscillator. Equation (63) gives the first correction of the diffusion coefficient given by Eq. (4). This series converges only if $\epsilon < 1/4$. This means that the correction given by Eq. (63) and (64) can lead to a quantitative change of the diffusion coefficient in the order of 0–25%.

Equation (63) is further improved if the sixth cumulant, Eq. (58), is included. This leads to

$$\begin{aligned} \frac{\partial}{\partial t} P = & - \frac{\partial}{\partial q_k} \Big[\delta_{kl} + \frac{m}{c^2} U_{kl} \\ & + \frac{2m^2}{c^4} U_{ki} U_{il} + \frac{m^2}{c^4} U_{kli} U_i + \frac{m^2}{2\beta c^4} U_{klii} \Big] j_l \\ & - \frac{\partial}{\partial q_k} \left(\frac{3m^2}{2c^4 \beta} U_{klj} \frac{\partial}{\partial q_j} \right) j_l. \end{aligned} \tag{65}$$

It seems that for the harmonic oscillator the higher order corrections lead always to a second order partial differential equation of the same form as the Smoluchowski equation but with a modified diffusion coefficient.

In the next section we will show that all solutions of Eq. (65) reach the Boltzmann distribution as $t \rightarrow \infty$ if the potential is sufficiently smooth.

In general, for a diffusion equation

$$\frac{\partial}{\partial t} P = AP \tag{66}$$

one requires that the diffusion operator A is dissipative;[18] this means

$$\mathrm{Re}\langle\langle \psi, A\psi \rangle\rangle \leqslant 0 \quad \text{for all } \psi \in D(A). \tag{67}$$

The scalar product used in the following discussion is defined

$$\langle\langle \psi,\phi \rangle\rangle \equiv \int_{\mathbf{R}^3} d^3q \, g_B^{-1} \psi^* \phi. \tag{68}$$

We denote by H the set of all functions ϕ on $\mathbb{R}^3$ which are measureable on $\mathbb{R}^3$ such that the expression $\langle\langle \phi,\phi \rangle\rangle$ defined by Eq. (68) exists, $\langle\langle \phi,\phi \rangle\rangle < \infty$. Functions which are equal almost everywhere are identified. The norm on H is defined by

$$\|\phi\| \equiv \langle\langle \phi,\phi \rangle\rangle^{1/2} \quad \text{for } \phi \in H. \tag{69}$$

The diffusion operator given by Eq. (62) is dissipative because for $A^{(2)} = -(\partial/\partial q) \cdot J$,

$$\begin{aligned} \mathrm{Re}\langle\langle \psi, A^{(2)}\psi \rangle\rangle &= \mathrm{Re}\langle\langle \psi, \frac{-\partial}{\partial q} J\psi \rangle\rangle \\ &= -\sum_{i=1}^{3} \mathrm{Re}\langle\langle J_i\psi, J_i\psi \rangle\rangle (c\beta)^{-1} \leqslant 0. \end{aligned} \tag{70}$$

Similarly one shows that the operator $A^{(4)} = (-\partial/\partial q_k) D_{kl} J_l$ is dissipative if all eigenvalues of the symmetric matrix D_{kl} are non-negative.

These concepts are useful because one can now prove that the solution of these diffusion equations approach the Boltzmann distribution in the limit as $t \rightarrow \infty$. First, the norm $\|P(t)\|$ of a solution of the diffusion equation (66) decreases monotonically in time,

$$\begin{aligned} \frac{d}{dt} \|P(t)\|^2 &= \frac{d}{dt} \mathrm{Re}\langle\langle P(t), P(t) \rangle\rangle \\ &= 2 \,\mathrm{Re}\, \langle\langle P(t), AP(t) \rangle\rangle \\ &\leqslant 0. \end{aligned} \tag{71}$$

The norm $\|P(t)\|$ must reach its minimum as $t \rightarrow \infty$ because $\|P(t)\|$ is bounded from below, $\|P(t)\| \geqslant 0$. Hence,

$$\frac{d}{dt} \|P(t)\|^2 \Big|_{t=\infty} = 0. \tag{72}$$

This means that $\mathrm{Re}\langle\langle P, AP \rangle\rangle = 0$ at $t = \infty$. But the equation $\mathrm{Re}\langle\langle P, AP \rangle\rangle = 0$ implies that $P = g_{\mathrm{B}}$ in the case of the diffusion equations (62), $A = A^{(2)}$, and (63), $A = A^{(4)}$. For instance,

$$\begin{aligned} & \mathrm{Re}\langle\langle P, \frac{\partial}{\partial q} \cdot JP \rangle\rangle = 0, \\ & \Rightarrow \langle\langle J_l P, J_l P \rangle\rangle = 0 \qquad \text{for } l = 1,2,3, \\ & \Rightarrow J_l P = 0 \qquad \text{for } l = 1,2,3, \\ & \Rightarrow P = g_{\mathrm{B}}. \end{aligned}$$

Therefore, by requiring that the diffusion operator A is dissipative and also that g_{B} is the only function which fulfills the equation $\mathrm{Re}\langle\langle g_{\mathrm{B}}, Ag_{\mathrm{B}} \rangle\rangle = 0$, one makes sure that all solutions reach finally the Boltzmann distribution g_{B}.

Calculating higher order cumulants leads very fast to extremely complicated expressions for the diffusion operators. The sixth cumulant leads to a third partial differential operator; compare Eq. (58). From the investigation of $A^{(4)}$ one suspects that, again, the diffusion operator $A^{(6)} = G^{(2)} + G^{(4)} + G^{(6)}$ is only meaningful if the potential U is sufficiently smooth. This is actually the case. Without going into all details the result may be stated as follows:

$$\text{(i)}\mathrm{Re}\langle\langle\psi, A^{(6)}\psi\rangle\rangle = 0 \Rightarrow \psi = g_{\mathrm{B}}, \tag{73}$$

(ii)$A^{(6)}$ is dissipative if

$$1 - \lambda_2 - \tfrac{5}{2}\lambda_4 \geqslant 0, \tag{74}$$

with

$$\lambda_l = \sup|D^{\nu_i}U(q)|\cdots|D^{\nu_k}U(q)|\left(\frac{3}{c}\right)^l m^{l/2}\beta^{k-l/2},$$
$$q \in \mathbb{R}^3, |\nu_i| \geqslant 1,$$
$$l = |\nu_i| + \cdots + |\nu_k|. \tag{75}$$

The derivatives D^{ν} are defined $D^{\nu} = (\partial/\partial q_i)^{\sigma}(\partial/\partial q_2)^{\beta}(\partial/\partial q_3)^{\gamma}$, $\nu = (\alpha, \beta, \gamma)$ and $|\nu| = \alpha + \beta + \gamma$.

λ_l is dimensionless constant which is typical for the cumulants $G^{(k)}$ for $k \geqslant l + 2$. For more details see again Ref. 16.

From Eqs. (73) and (74) and from the previous remarks it follows that all solutions of the diffusion equation $(\partial/\partial t)P = A^{(6)}P, P(t = 0) \in H$, approach the Boltzmann distribution as $t \to \infty$ (for sufficiently smooth potentials). This is an important result because the form of the operator $A^{(6)}$ differs very much from a typical diffusion operator, for example, such as $A^{(4)}$.

4. INTERPRETATION, APPLICATIONS, AND REMARKS ON THE CONVERGENCE OF THE CUMULANT EXPANSION

Position dependent diffusion coefficients. First, we consider the diffusion equation (63). It was already pointed out that one must require

$$\frac{\partial^2 U}{\partial q_i \partial q_j} < \frac{c^2}{m} \quad \text{for all } i, j. \tag{76}$$

The diffusion equation (63) can be reduced to a diffusion equation with constant diffusion coefficients. This is achieved by introducing new variables $Q = Q(q)$, a coordinate transformation which has an interesting physical interpretation, illustrating the interplay of the Newtonian dynamics and the pure Brownian motion.

Consider the symmetric matrix $D_{kl}(q)$ defined by Eq. (64). It is possible to find a square root $g_{\alpha\beta}(q)$ such that

$$g_{\alpha\beta}(q)g_{\beta\gamma}(q) = D_{\alpha\gamma}(q), \tag{77}$$
$$g_{\alpha\beta}(q) = g_{\beta\alpha}(q).$$

We introduce the new variables Q_1, Q_2, Q_3 by

$$dQ_\mu = g^{-1}{}_{\mu\nu}dq_\nu. \tag{78}$$

The new coordinates Q_1, Q_2, Q_3 are the coordinates in a nonorthogonal coordinate frame. These coordinates belong to a differentiable manifold with the metric

$$\frac{dQ_\mu}{dq_\alpha}\frac{dQ_\nu}{dq_\alpha} = D_{\mu\nu}^{-1}(q). \tag{79}$$

The metric is the inverse of the diffusion matrix $[D_{\mu\nu}(q)]$.

The new coordinates $Q_\mu(q_1, q_2, q_3)$ are well defined by Eq. (78) if dQ_μ is an exact differential. This condition can be written

$$\frac{\partial g_{\alpha\mu}^{-1}}{\partial q_\nu} - \frac{\partial g_{\alpha\nu}^{-1}}{\partial q_\mu} = 0. \tag{80}$$

The new density $P'(t,Q)$ is obtained from the density $P(t,q)$ by the transformation

$$P'(t,Q) = gP(t,q(Q)), \tag{81}$$

with

$$g \equiv \det(g_{\alpha\beta}). \tag{82}$$

In order to derive the diffusion equation for the new distribution function $P'(t,Q)$, Eqs. (78) and (81) are substituted into Eq. (63). This leads to

$$\frac{\partial}{\partial t}P'(t,Q) = g\frac{1}{c\beta}g_{\mu\alpha}^{-1}\frac{\partial}{\partial Q_\alpha}D_{\mu\nu}\,e^{-\beta U}g_{\nu\beta}^{-1}\frac{\partial}{\partial Q_\beta} \times e^{\beta U}g^{-1}P'(t,Q). \tag{83}$$

It is convenient to write current operator J in the form $J = (-kT/c)e^{-U\beta}(\partial/\partial q)e^{U\beta}$. The potential U is now considered as a function of Q;

$$U = U(q(Q)) = U(Q). \tag{84}$$

The derivative $(\partial/\partial Q_\alpha)gg_{\mu\alpha}^{-1}$ vanishes as a consequence of the integrability condition (80). Hence Eq. (83) can be written

$$\frac{\partial}{\partial t}P'(t,Q) = \frac{1}{c\beta}\frac{\partial}{\partial Q_\alpha}e^{-U\beta+\ln g}\frac{\partial}{\partial Q_\alpha}e^{U\beta-\ln g}P'(t,Q). \tag{85}$$

This is the diffusion equation in the new coordinates Q. In these curvilinear coordinates, the diffusion equation looks like a diffusion equation in an Euclidean space with constant diffusion coefficients $\delta_{\alpha\beta}(1/c\beta)$ and potential $U - \beta^{-1}\ln g$.[19]

The transformation $Q = Q(q)$ has a physical meaning. Recall that the inverse of the square root of the matrix D_{ij} is in first order in m/c^2 given by

$$g_{\mu\nu}^{-1} \cong \delta_{\mu\nu} - \frac{m}{2c^2}\frac{\partial^2 U}{\partial q_\mu \partial q_\nu}. \tag{86}$$

The approximation is consistent because the matrix D_{ij} is accurate only up to first order in m/c^2. The higher order corrections would be specified by the higher order cumulants. The expression (86) satisfies the integrability condition (80). The new variables Q are

$$Q = q + \frac{t_k{}^2 a(q)}{2}. \tag{87}$$

t_k is the correlation time of the momenta,

$$t_k = m/c. \tag{88}$$

$a(q)$ is the acceleration due to the force $-\partial U/\partial q$,

$$a_\mu(q) = -m^{-1}\frac{\partial U}{\partial q_\mu}. \tag{89}$$

It is easy to verify the Eq. (87) is a solution of Eq. (78) and (87).

Equation (85) shows that the "effective potential" in the new variables Q is equal to $U - \beta^{-1}\ln g$. We use again the approximation (86). Hence,

$$U' \equiv U - \beta^{-1}\ln g \cong U - \frac{m}{2c^2\beta}\det\left(\frac{\partial^2 U}{\partial q_i \partial q_j}\right). \tag{90}$$

The diffusion equation (62) can now be written

$$\frac{\partial}{\partial t} P'(t,Q) = - \frac{\partial}{\partial Q} \cdot J_Q P'(t,Q), \tag{91}$$

with

$$J_Q = - \frac{1}{c\beta} \left(\frac{\partial}{\partial Q} + \beta \frac{\partial U'}{\partial Q} \right). \tag{92}$$

Equation (92) has the same *form* as the Smoluchowski equation (46) obtained for translational diffusion in first nonvanishing order. But Eq. (92) describes a different process. The new interpretation is given by the coordinate transformation (87) and the new potential U' defined by Eq. (90).

Comparison of Eqs. (87) and (91) shows a competition between the Newtonian dynamics and a Brownian motion. Equation (87) is the orbit of a Newtonian particle under the influence of a force which is constant in time: $q(t) = q_0 + v_0 t + at^2/2$. In Eq. (87) the term containing the initial velocity v_0 vanishes. the Newtonian dynamics applies only for a very short time the correlation time of the momenta t_k. Applying the transformation (87) shows that everything else is a pure Brownian motion with external potential U' governed by the Smoluchowski equation (92).

Nonlinear diffusion equations. In a manner similar to the preceding presentation the diffusion equation can be derived for N interacting particles.[4,16] This is leads to a partial differential equation in the $3N$ dimensional position space. There are only few applications for the N particle diffusion equation because it cannot be solved in general for large N. On the other hand, one may ask the question if the reduced density

$$\rho(t,q) = \int_{\mathbf{R}^3} d^3 q_2 \cdots \int_{\mathbf{R}^3} d^3 q_n \int_{\mathbf{R}^3} d^3 p_1 \cdots \int_{\mathbf{R}^3} d^3 p_n \times f(t,q,q_2,\ldots,q_N,p_1,\ldots,p_N) \tag{93}$$

fulfills a simple diffusion equation if N is large and the interaction between the particles is weak. [In Eq. (93) $f(t,q,q_2,\ldots,q_N,p_1,\ldots,p_N)$ is the probability density in the $6N$ dimensional phase space.] If the correlations between the particles can be neglected, the N body problem reduces to the one particle dynamics. For this case we propose the following diffusion equation: In one dimension,

$$\frac{\partial}{\partial t} \rho(t,q) = D \frac{\partial}{\partial q} \left(1 + \frac{m}{c^2} \frac{\partial^2 U[\rho]}{\partial^2 q} \right) \left(\frac{\partial}{\partial q} + \beta \frac{\partial U[\rho]}{\partial q} \right) \rho(t,q), \tag{94}$$

and similarly in three dimensions. The potential U is now a functional of the density ρ. For a specific two particle interaction potential $V(q)$ one can write (only the one dimmensional case is considered in the following)

$$U[\rho](q) = \int_{\mathbf{R}^3} dq' V(q - q') \rho(q'). \tag{95}$$

The diffusion equation (94) can be used to describe excluded volume effects if one chooses the potential

$$V(q) = \alpha \delta(q). \tag{96}$$

α monitors the strength of the excluded volume force. Combining Eqs. (94)–(96) leads to

$$\rho_t = A \rho_{qq} + B [\rho \rho_{qq} + (\rho_q)^2] + C [(\rho_{qq})^2 + \rho \rho_{qqq}] + \cdots. \tag{97}$$

ρ_t and ρ_q denote the partial derivatives $(\partial / \partial t) \rho$ and $(\partial / \partial q) \rho$, respectively. This equation serves only as a simple illustration and therefore we did not write out the terms proportional to α^2. The constants are $A = D$, $B = \alpha D / kt$, and $C = \alpha m D / c^2$.

Equation (97) represents a description of the self-avoiding random walk.[20] For classical diffusion (or random walk) the variance σ^2 of the distribution $\rho(t,q)$ increases proportional to time. One expects that the distribution $\rho(t,q)$ spreads out faster in the case of the self-avoiding random walk.

Equations for the first and second moment of the distribution $\rho(t,q)$ can easily be obtained from Eq. (97).

$$\frac{d}{dt} \langle q \rangle = 0, \tag{98}$$

$$\frac{d}{dt} \langle q^2 \rangle = 2A + B \int_{-\infty}^{+\infty} dq\, \rho^2(t,q) + C \int_{-\infty}^{+\infty} dq\, \rho^2{}_q(t,q) + \cdots. \tag{99}$$

The first term in Eq. (99) alone would lead to the old result $\sigma^2 \sim t$. The second and third term are always positive; the variance σ^2 increases now faster in time. In the sense of a first approximation one can assume that the distribution $\rho(t,q)$ is Gaussian for all times if the initial distribution was Gaussian. This approximation leads to a simple differential equation for the variance σ^2:

$$\frac{d}{dt} \sigma^2 = 2A + \frac{\beta}{2\pi^{1/2}} \sigma^{-1} + \frac{C}{2(2\pi)^{1/2}} \sigma^{-3} + \cdots. \tag{100}$$

For large values of σ^2 the higher order corrections can be neglected. In this case one obtains again $\sigma \sim t^{1/2}$. On the other hand, if σ is small the contributions of the higher order terms become more and more important. If the right-hand side of the differential equation (100) is dominated by the third term one obtains the relation $\sigma \sim t^{1/5}$. In this case the higher order corrections of the diffusion equation do not only lead to a quantitative modification of the solution but they lead to a qualitatively different solution.

Remarks on the convergence of the cumulant expansion. The following remarks are not a substitute for a rigorous proof of the convergence of the cumulant expansion. From a practical point of view it is more important to understand the physsical implications and to establish criteria which allow identification of the physical situations (type of potentials, possible values of the parameters m, c, β) which are best described by the diffusion equations derived in the previous section and to be able to distinguish these situations from those where these diffusion equations cannot be applied. In this case one would have to go back to the Kramer–Liouville equation.

Consider the cumulant expansion

$$G = \sum_{n=1}^{\infty} G^{(n)}. \tag{101}$$

There is a formal similarity between the cumulant expansion and Mayer's expansion of the grand canonical partition function for the real gas.[21] Furthermore, the cumulants have a cluster property analogous to the cluster property of the Ursell functions. The cluster property for the cumulants is a consequence of the following factorization property. We say that the moments $\langle M(t_1)...M(t_n)\rangle$ satisfy the factorization property if, for all n and $1\leqslant i\leqslant n-1$,

$$\langle M(t_1+\tau)\cdots M(t_i+\tau)M(t_{i+1})\cdots M(t_n)\rangle$$
$$-\langle M(t_1+\tau)\cdots M(t_i+\tau)\rangle\langle M(t_{i+1})\cdots M(t_n)\rangle \underset{\tau\to\infty}{\longrightarrow} 0. \quad (102)$$

One can show[22] that, if this factorization property holds, the cumulant averages go to zero if the "distance" τ between two "time clusters" becomes large. We have for all n and $1\leqslant i\leqslant n-1$

$$\langle M(t_1+\tau)\cdots M(t_i+\tau)M(t_{i+1})\cdots M(t_n)\rangle_c \underset{\tau\to\infty}{\longrightarrow} 0. \quad (103)$$

The cumulants are appreciably bigger than zero only if the time moments $t_1,...,t_n$ are clustered together, for instance, if all variables $t_1,...,t_n$ are contained in a short interval of time.

One can easily verify that the factorization property (102) holds also for the Liouville operator in the interaction picture, $\tilde{L}(t)$ [compare also Eq. (48)].

$$\langle \tilde{L}(t_1+\tau)\cdots\tilde{L}(t_i+\tau)\tilde{L}(t_{i+1})\cdots\tilde{L}(t_n)\rangle$$
$$-\langle \tilde{L}(t_1+\tau)\cdots\tilde{L}(t_i+\tau)\rangle\langle\tilde{L}(t_{i+1})\cdots\tilde{L}(t_n)\rangle \sim e^{-(c/m)\tau}. \quad (104)$$

This means in physical terms: The momenta $p(t)$ and $p(t+\tau)$ become independent random variables if $\tau\gg t_k=m/c$, since the momentum $p(t+\tau)$ has been changed by many random collisions with the Brownian fluid particles. Therefore, the operator $\tilde{L}(t)$ can also be considered as an independent random operator acting on the spatial distribution of the time difference τ is large compared with the correlation time of the momenta t_k. The autocorrelation function of the momenta and also the expression (104) decrease exponentially in τ, the lapse of time.[23]

Equation (96) represents a strong version of the factorization property (94), which leads to a strong cluster property

$$\langle \tilde{L}(t_1+\tau)\cdots\tilde{L}(t_i+\tau)\tilde{L}(t_{i+1})\cdots\tilde{L}(t_n)\rangle_c \sim e^{-(c/m)\tau}. \quad (105)$$

This equation and the invariance of the cumulant averages under time translations,

$$\langle\tilde{L}(t_1+\tau)\cdots\tilde{L}(t_n+\tau)\rangle_c=\langle\tilde{L}(t_1)\cdots\tilde{L}(t_n)\rangle_c, \quad (106)$$

can be used to show that the limit of $G^{(n)}$ as $t\to\infty$ exists for all n. Moreover, the cluster property (105) shows that the higher order cumulants are very small if the correlation time of the momenta, $t_k=c/m$, is small measured on a macroscopic time scale with units $t_{\rm mac}$. The nth cumulant is proportional to $(t_k/t_{\rm mac})^{n-1}$,

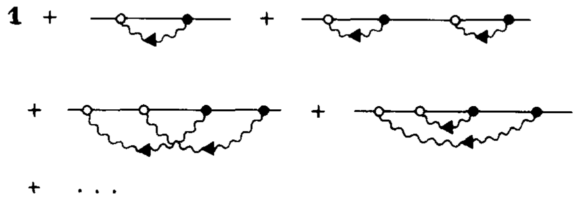
$$G^{(n)}\sim(t_k/t_{\rm mac})^{n-1}. \quad (107)$$

In the following it is useful to represent the exact solution, obtained by solving the Kramers–Liouville equation $P(t)$, graphically. We assume that the time-ordered exponential

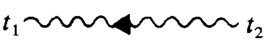
$$P(t)=\langle \underset{\leftarrow}{T}\exp\int_0^t ds\,\tilde{L}(s)\rangle P_0 \quad (108)$$

converges for some macroscopic time t, which can be small but still large compared with $t_k=m/c$. The time-ordered exponential can be represented graphically.

$$\langle \underset{\leftarrow}{T}\exp\int_0^t ds\,\tilde{L}(s)\rangle = \quad (109)$$



The straight lines denotes the t axis. The circles represent the operators b, respectively, $b^\dagger$ and wiggled lines stand for a factor $e^{-c/m(t_1-t_2)}$.

———	:	t axis
○	:	$-b_i^\dagger$
●	:	b_j
	:	$e^{-c/m(t_1-t_2)}\delta_{ij}$ (110)

It is understood that an integration over all possible ordered sequences of time $t>t_1>\cdots>t_n>0$ is performed in all diagrams.

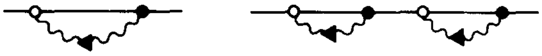
The lines can be interpreted as the propagator of a virtual particle describing the correlation of the momenta.

Equation (109) is an expression of the Gaussian property. The first diagram is the second moment of $\tilde{L}(t)$, $\langle\tilde{L}(t_1)\tilde{L}(t_2)\rangle$. The next three diagrams add up to the fourth moment, etc.

The graphical representation (109) is very convenient to compare the exact solution $P(t)$ gained from the Kramers–Liouville equation with the approximate solutions, for instance the first approximation $P^{(2)}(t)$ which is the solution of the Smoluchowski equation $(\partial/\partial t)P^{(2)}(t)=G^{(2)}P^{(2)}(t)$. The function $P^{(2)}(t)$,

$$P^{(2)}(t)=\underset{\leftarrow}{T}\exp\int_0^t ds\,G^{(2)}(s)P_0$$
$$=\underset{\leftarrow}{T}\exp\Big[-\int_0^t ds\int_0^s ds' b^\dagger(s)\cdot b(s')\Big]P_0, \quad (111)$$

could also be represented graphically but the operators $b^\dagger(t)$ and $b(t)$ are no longer time ordered in the expansion (111)! All nonoverlapping graphs appearing in (111) agree with the corresponding graphs in (109). These graphs are

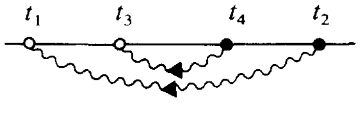


etc. (112)

On the other hand, the expansion (111) also contains terms of the form

$$b^\dagger(t_1)\cdot b(t_2)b^\dagger(t_3)\cdot b(t_4), \quad (113)$$

with $t_1>t_3>t_4>t_2$! The corresponding term in the expansion (109) is the time-ordered product of (113) given by



(114)

The two expressions (113) and (114) are not equal if the operators b_l and $b_m^\dagger$ do not commute. The "errors" which arise from the incorrect ordering of these operators in the expression (111) are corrected if the fourth cumulant is included. The fourth cumulant is therefore proportional to the commutator $[b_i,b_j^\dagger]$. Multiple "overlaps" are corrected by higher cumulants. The higher cumulants generate nested commutators, which lead to expressions in terms of derivatives of the potential U. For instance,

$$[b_i,b_j^\dagger] = m^{-1}\frac{\partial^2 U}{\partial q_i \partial q_j},$$

$$[b_i,[b_j,b_l^\dagger]] = -m^{-3/2}\beta^{-1/2}\frac{\partial^3 U}{\partial q_i \partial q_j \partial q_l}, \qquad (115)$$

etc.

If the operators b_i and $b_j^\dagger$ commute, only the second cumulant survives. Rotational diffusion has also been studied.[24,25] In this case the higher order cumulants do not vanish even in the absence of an external potential, quite contrary to the translations diffusion without external forces. These corrections, which have been calculated up to fourth order,[24] are due to the noncommutativity of the operators M_1, M_2, M_3, the infinitesimal generators of SO(3), which correspond to the operators

$$\frac{1}{i}\frac{\partial}{\partial q}, \quad \frac{1}{i}\frac{\partial}{\partial q_2}, \quad \frac{1}{i}\frac{\partial}{\partial q_3},$$

and the infinitesimal generators of the translations in $\mathbb{R}^3$.

The main reason why the cumulant expansion gives such as excellent description of the diffusion process is given by the fact that not only the first few diagrams in (109) but also the higher order diagrams are represented correctly to a large extent if the correlation time t_k is small.

5. CONCLUDING REMARKS

It has been shown that the diffusion equation for translational Brownian motion can be calculated using a boson representation of the Kramers–Liouville equation and the time-ordered cumulants. The first six terms of this expansion have been calculated. They have led to a fourth order partial differential operator. It has been shown that Einstein's result is exact in the limit as $t \to \infty$. In general, the higher order cumulants are small because the cluster property holds.

The boson representation can be extended to coupled translational and rotational diffusion of molecules or arbitrary shape.[16] The Gaussian property [Eq. (48)] does *not* apply for rotational diffusion but rules have been derived for calculating the moments $\langle \tilde{L}(t_i)\cdots\tilde{L}(t_n)\rangle$ in the most general case.[16] The boson representation allows one to calculate or estimate the higher order corrections in a straightforward manner. This work generalizes earlier results on rotational diffusion[24] and coupled translational and rotational diffusion.[25,26]

It can be shown that the effects on diffusion due to the so-called long time tails,[27] the nonexponentially decreasing tails of the velocity autocorrelation, are very small in three dimensions. These non-Markovian aspects of Brownian motion will be discussed in an upcoming publication.

Note added in proof: It has been brought to the author's attention that a one dimensional version of Eq. (65) and the proof of the conjecture on the harmonic oscillator below Eq. (65) have been published previously.[28]

ACKNOWLEDGMENTS

This work was supported in part by NSF grant PHY 7921541 to one of us (R.F.F.) and in part by NSF grant CHE 80-09816 (U.R.S.).

[1] Einstein's diffusion equation follows from the Kramers–Liouville equation without further approximations in the limit as $t \to \infty$.

[2] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943) [reprinted in N. Wax, *Selected Papers, on Noise and Stochastic Processes* (Dover, New York, 1954)].

[3] This expression holds if $\omega < c/2m$ and the first moment of the momentum distribution vanishes initially, $\langle p(0)\rangle = 0$.

[4] G. Wilemski, J. Stat. Phys. **14**, 153 (1975).

[5] In particular, the non-Markovian aspects ("long-time-tails" of the velocity autocorrelation function!) are not discussed in this paper, but it can be shown that the effects on diffusion are very small in three dimensions.

[6] R. F. Fox, Phys. Rep. **48**, 179 (1978), p. 250.

[7] See Ref. 6, p. 240.

[8] This is a generalization of the $U - W$ relations which connect the Boltzmann factors and the Ursell functions used in the context of Mayer's theory of the imperfect gas (see Ref. 9).

[9] D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969), Eq. (4.5), p. 87.

[10] J. J. Sakurai, *Advanced Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1967), Chaps. 2 and 4.

[11] These functions are

$$\psi_{n_1 n_2 n_3} = \psi_{n_1}\psi_{n_2}\psi_{n_3};$$

$$\psi_n(p) = (\beta m/n!2^n 2\pi)^{1/2} e^{-\beta(p^2/2m)} H_n((\beta/2m)^{1/2} p).$$

[12] P. Reśibois and M. De Leener, *Classical Kinetic Theory of Fluids* (Wiley, New York, 1977), Appendix A.

[13] Consider the function $\tilde{\mathscr{L}}(\mathbf{x})$ formally defined by

$$\tilde{\mathscr{L}}(x) \equiv \tilde{\mathscr{L}}(\mathbf{t},q,q') \equiv \tilde{L}(t)\delta(q-q'); x = (t,q,q') \in \mathbb{R}^7.$$

The function $\mathscr{L}(\tilde{x})$ depends implicitly on the momenta, but it is remarkable that $\tilde{\mathscr{L}}(x)$ can be considered as a Gaussian random variable on $\mathbb{R}^3 \times \mathbb{R}^3$. For $\tilde{\mathscr{L}}(x)$ Theorem 3 reads

$$\langle \tilde{\mathscr{L}}(x_1)\cdots\tilde{\mathscr{L}}(x_n)\rangle = \sum_{\substack{\text{all pairs}\\(ij)}} \pi\langle \tilde{\mathscr{L}}(x_i)\tilde{\mathscr{L}}(x_j)\rangle.$$

[14] A. Messiah, *Quantum Mechanics* (Wiley, New York, 1961), Vol. 1, p. 442.

[15] see Ref. 6, p. 278 and 279.

[16] U. R. Steiger, "*Analysis of Coupled Translational and Rotational Diffusion Using Operator Calculus*," Ph.D. Thesis, Georgia Institute of Technology (University Microfilms, 1981).

[17] R. F. Fox, J. Math. Phys. **17**, 1148 (1976).

[18] K. Yosida, *Functional Analysis* (Springer, Berlin, 1974), Chap. 8.

[19] For comparison, see Eq. (46).

[20] S.-K. Ma, *Modern Theory of Critical Phenomena* (Addison-Wesley, Reading, Mass. 1976), Chap. X.5.

[21] R. F. Fox, J. Math. Phys. **16**, 289 (1975).

[22] See Ref. 17, compare also Ref. 9, Chap. 4.4.

[23] This is true for "classical Brownian motion" only; if the non-Markovian aspects ("long-time-tails") are considered too, the exponential $e^{-(c/m)\tau}$ in Eqs. (104) and (105) goes over into a power law $t^{-3/2}$ (in three dimensions) for very large values of τ.

[24] G. W. Ford, J. T. Lewis, and J. McConnell, Phys. Rev. A **19**, 907 (1979).

[25] U. R. Steiger and R. F. Fox, J. Math. Phys. **23**, 296 (1982).

[26] D. W. Condiff and J. S. Dahler, J. Chem. Phys. **44**, 3988 (1966).

[27] See, for instance, Ref. 12, Chap. 7.

[28] U. M. Titulaer, Physica A **91**, 321 (1978).

Lattice dynamics, random walks, and nonintegral effective dimensionality [a)]

Barry D. Hughes
Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, Minnesota 55455

Michael F. Shlesinger
Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742 and La Jolla Institute, La Jolla, California 92038

(Received 21 July 1981; accepted for publication 7 January 1982)

Definitions of the nonintegral effective dimensionality of recursively defined lattices (fractal lattices) may be based on scaling properties of the lattices, or on the qualitative behavior of cooperative phenomena supported by the lattices. We examine analogs of these definitions for regular (i.e., periodic) lattices supporting long-range interactions. In particular, we show how to calculate a harmonic oscillator effective dimension, a scaling dimension, and a random walk effective dimension for simple cubic lattices with a class of long-range interactions. We examine the relationship between these three dimensions for regular lattices, and conjecture a constraint on the analogs of these dimensions for fractal lattices.

PACS numbers: 05.50. + q

I. INTRODUCTION

Lattices of effectively nonintegral dimensionality have been studied recently by a number of authors[1–4] with a view to obtaining information about the effect of dimensionality, degree of symmetry, and topological structure on critical phenomena. Such lattices are frequently defined recursively (and consequently are hard to visualize), but have the advantage that renormalization group techniques are easily applied and analytic results are sometimes available.

There is no *a priori* method for assigning a numerical value to the dimension of a nonstandard lattice, and several definitions currently in use fail to asssign the same dimension to particular lattices (although they do coincide, as they should, with the usual value of the dimension of any Bravais lattice). Nelson and Fisher[1] and Gefen, Mandelbrot, and Aharony[2] employ a definition based on scaling properties, while Dhar[3] uses a dimension based on lattice dynamics.[5] Neither of these definitions gives the value 1 for the lower critical dimension of the Ising model.[6] To gain insight into the relationship between the inequivalent definitions, we consider here some model systems in which the effective dimensionality of a simple cubic lattice is changed when the physical processes (interactions) it supports are long ranged. We are able to calculate a harmonic lattice dimension h (in the manner of Dhar[3]), a scaling dimension f (analogous to that of Nelson and Fisher[1] and Gefen *et al.*[2]), and a random walk effective dimension r (proposed by Hughes, Shlesinger, and Montroll[7,8]).

II. EQUATIONS OF MOTION

We begin by writing down a few results from the theory of lattice dynamics[9,10] and the theory of random walks,[11,12] and noting certain mathematical similarities between them. Consider an s-dimensional simple cubic lattice of coupled harmonic oscillators, each oscillator possessing only one degree of freedom (for simplicity). If $x(\mathbf{l},t)$ denotes the displacement of the oscillator at site $\mathbf{l}$ and time t, and each oscillator has mass m, then the equations of motion of the lattice may be written in the form

$$m\frac{\partial^2}{\partial t^2}x(\mathbf{l},t)=\sum_{\mathbf{l}'}\gamma(\mathbf{l}')\{x(\mathbf{l}+\mathbf{l}',t)-x(\mathbf{l},t)\}; \qquad (1)$$

here the interaction constants are assumed to satisfy the relation $\gamma(\mathbf{l})=\gamma(-\mathbf{l})\geqslant 0$ to ensure stability of the lattice. The trial solution

$$x(\mathbf{l},t)\propto\exp(i\boldsymbol{\theta}\cdot\mathbf{l}-i\omega t) \qquad (2)$$

leads to the dispersion relation

$$\omega^2=W(\boldsymbol{\theta})=(1/m)\sum_{\mathbf{l}}\gamma(\mathbf{l})[1-\exp(i\boldsymbol{\theta}\cdot\mathbf{l})]. \qquad (3)$$

For a random walk on the same lattice commencing at the origin, the probability $P_n(\mathbf{l})$ of the walker being at the site $\mathbf{l}$ after n steps may be easily calculated in terms of the single step jump distribution $p(\mathbf{l})$, since

$$P_{n+1}(\mathbf{l})=\sum_{\mathbf{l}'}p(\mathbf{l}-\mathbf{l}')P_n(\mathbf{l}'). \qquad (4)$$

Here we assume that the walk is symmetric, i.e., $p(\mathbf{l})=p(-\mathbf{l})$, in addition to the usual requirements that $\Sigma_{\mathbf{l}}\, p(\mathbf{l})=1$ and $p(\mathbf{l})\geqslant 0$. It follows immediately from Eq. (4) that

$$P_{n+1}(\mathbf{l})-P_n(\mathbf{l})=\sum_{\mathbf{l}'}p(\mathbf{l}')\{P_n(\mathbf{l}+\mathbf{l}')-P_n(\mathbf{l})\}, \qquad (5)$$

and the similarity between the right hand sides of (1) and (5) implies a mathematical relationship between the solutions of the lattice dynamics and random walk problems, as first pointed out by Montroll.[13] In terms of the structure function

[a)] This research was supported in part by a postdoctoral fellowship from the Commonwealth Scientific and Industrial Research Organization of Australia, held by B. D. H., and by DARPA.

$$\lambda(\boldsymbol{\theta}) = \sum_{\mathbf{l}} p(\mathbf{l}) e^{i\mathbf{l}\cdot\boldsymbol{\theta}}, \tag{6}$$

the random walk generating function

$$P(\mathbf{l},z) \equiv \sum_{n=0}^{\infty} P_n(\mathbf{l}) z^n \tag{7}$$

is given by

$$P(\mathbf{l},z) = \frac{1}{(2\pi)^s} \int_B \frac{\exp(-i\mathbf{l}\cdot\boldsymbol{\theta}) d^s\boldsymbol{\theta}}{1 - z\lambda(\boldsymbol{\theta})}, \tag{8}$$

the volume integral being taken over the first Brillouin zone, i.e., the hypercube $B = \{\boldsymbol{\theta} = (\theta_1,\theta_2,\ldots\theta_s) | -\pi \leqslant \theta_j \leqslant \pi\}$. Writing

$$\xi = \sum_{\mathbf{l}} \gamma(\mathbf{l}) \quad \text{and} \quad p(\mathbf{l}) = \xi^{-1}\gamma(\mathbf{l}), \tag{9}$$

we set up the explicit mathematical correspondence between the lattice dynamics and random walk problems, so that the dispersion relation (3) becomes

$$\omega^2 = W(\boldsymbol{\theta}) = (\xi/m)\{1 - \lambda(\boldsymbol{\theta})\}. \tag{10}$$

In the terminology of probability theory, $\lambda(\boldsymbol{\theta})$ is the characteristic function of the discrete probability distribution $p(\mathbf{l})$. Although it is clear that $|\lambda(\boldsymbol{\theta})| \leqslant 1$, with equality holding at the points $(2\pi m_1, 2\pi m_2, \ldots, 2\pi m_s)$, m_j integral, i.e., the centers of the Brillouin zones, it is not obvious at which other points $\lambda(\boldsymbol{\theta})$ may attain the value 1. [This is an important question, because such points influence the asymptotic properties of the walk, as can be seen from Eq. (8).] A necessary and sufficient condition that no such additional points exist has been given by Spitzer[14] (the walk must be "aperiodic" in a certain sense). In the Appendix we give a short proof that in fewer than four dimensions no such additional points may exist, provided that $p(\mathbf{l})$ is nonzero for certain nearest-neighbor transitions.

III. HARMONIC LATTICE DIMENSION

Let $g(\omega)$ denote the distribution of frequencies[9,10] and $G(\omega^2)$ the distribution of squared frequencies for a lattice of harmonic oscillators. For a simple cubic lattice of spatial dimension s, with nearest-neighbor coupling only, it can be shown that as $\omega \to 0$, $g(\omega) \propto \omega^{s-1}$, or equivalently

$$\int_0^{\omega} g(\omega')d\omega' = \int_0^{\omega^2} G(\omega'^2)d(\omega'^2) \sim A\omega^s, \tag{11}$$

with A constant. For any infinite lattice of oscillators,[15] Dhar defines the dimension of the lattice as h if

$$H(\omega^2) \equiv \int_0^{\omega^2} G(\omega'^2)d(\omega'^2) \sim A(\omega)\omega^h \quad \text{as } \omega \to 0, \tag{12}$$

with $A(\omega)$ bounded, but not vanishing, in the neighborhood of $\omega = 0$. [The need to allow for nonconstant A in generalizing Eq. (11) is seen from Dhar's analysis[3] of the truncated n-simplex lattice.] For nonperiodic lattices, the determination of $H(\omega^2)$ is a matter of some difficulty.[16] In contrast, for the translationally invariant lattices considered in the present paper, $H(\omega^2)$ is easily found from the dispersion relation of the lattice at small wavenumbers. For the remainder of this paper, let C denote a constant, the value of which is not necessarily the same from line to line. Since

$$G(\omega^2) = C \int_B \delta(\omega^2 - W(\theta)) d^s\boldsymbol{\theta}, \tag{13}$$

it follows that

$$H(\omega^2) = C \int_B \Theta(\omega^2 - W(\boldsymbol{\theta})) d^s\boldsymbol{\theta}, \tag{14}$$

with Θ denoting the Heaviside unit step function. For sufficiently small values of ω only the local behavior of $W(\boldsymbol{\theta})$ in the neighborhood of its zeros contributes to the asymptotic form of $H(\omega^2)$, and from known properties of lattice walk structure functions we will usually only have zeros of $W(\boldsymbol{\theta})$ at the centers of the Brillouin zones (see the Appendix).

If the coupling constants $\gamma(\mathbf{l})$ of the lattice satisfy the condition

$$\sum_{\mathbf{l}} \mathbf{l}^2 \gamma(\mathbf{l}) < \infty \tag{15}$$

or, equivalently (in the random walk terminology), if the mean-squared displacement per step,

$$\langle \mathbf{l}^2 \rangle = \sum_{j=1}^{s} \langle l_j^2 \rangle = \sum_{\mathbf{l}} \mathbf{l}^2 p(\mathbf{l}), \tag{16}$$

is finite, then it is not difficult to establish that

$$1 - \lambda(\boldsymbol{\theta}) \sim \frac{1}{2} \sum_{j=1}^{s} \langle l_j^2 \rangle \theta_j^2. \tag{17}$$

Consequently, for small ω, it follows from Eq. (14) that $H(\omega^2)$ is given by the volume of a hyperellipsoid with semiaxes proportional to ω, i.e.,

$$H(\omega^2) \sim C\omega^s. \tag{18}$$

The harmonic lattice dimension h is thus the same as the usual dimension s of the space lattice unless the coupling constants are so long ranged that Eq. (15) is violated (or, equivalently, the associated random walk has infinite mean-squared displacement per step). We consider the canonical example of an interaction for which $h > s$:

$$1 - \lambda(\boldsymbol{\theta}) \sim C|\theta|^{\mu}, \tag{19}$$

where $\mu < 2$. Gillis and Weiss[17] have shown that Eq. (19) arises if

$$p(\mathbf{l}) \sim C|\mathbf{l}|^{-s-\mu} \quad \text{as} \quad |\mathbf{l}| \to \infty. \tag{20}$$

It is easily shown that for the lattice system governed by Eq. (19)

$$H(\omega^2) \sim C\omega^{2s/\mu}, \tag{21}$$

so that

$$h = 2s/\mu. \tag{22}$$

Not only does h exceed s but also, by choosing μ sufficiently small, h may be made arbitrarily large. In Secs. IV and V we study random walks with structure function similar to Eq. (19).

IV. SCALING DIMENSION

For self-similar lattices (as considered in Refs. 1–6), it is possible to define a scaling dimension or fractal dimension[18] as follows. Suppose that the lattice is generated by breaking a finite portion of it into N identical parts, each similar to the original, but scaled down by a linear factor L. Then

$$f = \ln N / \ln L \tag{23}$$

is the fractal or scaling dimension. An analog of this for random walks or lattice dynamics on cubic lattices may be considered[7,8] by building a self-similarity property into $p(\mathbf{l})$. In one dimension the canonical example is the "Weierstrass random walk"[7]

$$p(l) = \frac{a-1}{2a} \sum_{n=0}^{\infty} a^{-n} \{\delta_{l,b^n} + \delta_{l,-b^n}\} \tag{24}$$

$(a,b > 1$; b integral), for which

$$\lambda(\theta) = \frac{a-1}{a} \sum_{n=0}^{\infty} a^{-n} \cos(b^n \theta)$$
$$= a^{-1} \lambda(b\theta) + \frac{a-1}{a} \cos(\theta). \tag{25}$$

The mean-squared displacement per step is infinite when $b^2 \geqslant a$, and the long-time behavior of the walk is not Gaussian if the inequality holds strictly (which we now assume).

The walker makes on the average about a jumps of length 1, forming a cluster of sites visited, before making a jump of length b to begin a new cluster; about a such new clusters are formed before a jump of length b^2 occurs, and so on. While this self-similar cluster formation is obvious for a walk of limited duration, it is necessary that the random walk be transient, i.e., not every point is certain to be reached, for the clusters to persist for a walk of infinite number of steps. In one spatial dimension, transience of the walk[7] requires the inequality $b > a$. The set of points visited in a walk of infinite duration may be assigned the fractal dimension

$$f = \ln a / \ln b, \tag{26}$$

provided that the walk is transient, i.e.,

$$0 < f < 1. \tag{27}$$

The quantity f arises as a scaling dimension not only from this probabilistic argument, but also directly from the functional equation [Eq. (25)]. It can be shown that if $f < 2$, the small θ behavior of the structure function is given by

$$\lambda(\theta) - 1 \sim |\theta|^f Q(\theta), \tag{28}$$

where Q is a continuous but highly oscillatory function [being periodic in $\ln|\theta|$ with period $\ln b$, i.e., $Q(\theta) = Q(b\theta)$]. As $|l| \to \infty$,

$$p(l) = O(|l|^{-1-f}); \tag{29}$$

there is an effective power-law decay (of exponent $-1-f$) with some superimposed noise.

A well-defined scaling dimension does not exist for general one-dimensional long-tailed jump distributions for which $p(l) = O(|l|^{-1-\mu})$ with $0 < \mu < 2$, such as the example of Gillis and Weiss,[17]

$$\lambda(\theta) = \zeta(1+\mu)^{-1} \sum_{n=1}^{\infty} n^{-1-\mu} \cos(n\theta), \tag{30}$$

where ζ is the Riemann zeta function. However, by analogy with the Weierstrass walk, we assign to it the fractal dimension $f = \mu$. For the case $p(l) \sim C|l|^{-1-\mu}$ as $|l| \to \infty$, we have scaling in the *tail* of $p(l)$, but not for small values of l. With $\mu = \ln N / \ln L$, the scale change $l \to Ll'$ transforms $p(l)dl$ from $C|l|^{-1-\mu}dl$ to $N^{-1}C|l'|^{-1-\mu}dl' = N^{-1}p(l')dl'$. Analogs of the Weierstrass random walk may be constructed in higher dimensions,[8] but are most naturally discussed in terms of continuous space and we do not enter into these questions here.

V. RANDOM WALK EFFECTIVE DIMENSION

Pólya[19] showed that the probability u for a symmetric random walker, taking nearest neighbor steps on a simple cubic lattice, to return to the starting point is unity in one and two dimensions, and less than unity in three or more dimensions. These results are most easily obtained[11,20] from the generating function $P(\mathbf{l},z)$ [Eq. (8)], where $u = 1 - R^{-1}$, and in s dimensions,

$$R = \lim_{z \to 1-} P(\mathbf{0},z) = \lim_{z \to 1-} \frac{1}{(2\pi)^s} \int_B \frac{d^s\boldsymbol{\theta}}{1 - z\lambda(\boldsymbol{\theta})}. \tag{31}$$

For Pólya's walk, and for any symmetric walk with finite mean-squared displacement per jump, $1 - \lambda(\boldsymbol{\theta}) \sim$ constant$\cdot|\boldsymbol{\theta}|^2$ as $|\boldsymbol{\theta}| \to 0$. Pólya's conclusions must be modified when $1 - \lambda(\boldsymbol{\theta}) = O(|\boldsymbol{\theta}|^{\mu})$, as in Eq. (28), and the following argument shows that in this case the random walk may be considered to take place in a space of effective dimension greater than the Euclidean dimension.

We restrict our attention to the case when the only zero of $1 - \lambda(\boldsymbol{\theta})$ in B is at $\boldsymbol{\theta} = \mathbf{0}$. The convergence or divergence of the integral is determined by the behavior of the integral in a small hypersphere (of radius ρ) centered on the origin. Introducing polar coordinates, we see that if the mean-squared displacement per jump is finite then

$$\int_{|\boldsymbol{\theta}| \leqslant \rho} \frac{d^s\boldsymbol{\theta}}{1 - \lambda(\boldsymbol{\theta})} \sim C \int_o^{\rho} \frac{|\boldsymbol{\theta}|^{s-1}}{|\boldsymbol{\theta}|^2} d|\boldsymbol{\theta}|, \tag{32}$$

while if

$$1 - \lambda(\theta) \sim C|\theta|^{\mu} \quad (0 < \mu < 2), \tag{33}$$

then

$$\int_{|\theta| \leqslant \rho} \frac{d^s\boldsymbol{\theta}}{1 - \lambda(\boldsymbol{\theta})} \sim C \int_0^{\rho} \frac{|\boldsymbol{\theta}|^{s-1} d|\boldsymbol{\theta}|}{|\boldsymbol{\theta}|^{\mu}}$$
$$= C \int_0^{\rho} \frac{|\boldsymbol{\theta}|^{(s+2-\mu)-1}}{|\boldsymbol{\theta}|^2} d|\boldsymbol{\theta}|. \tag{34}$$

We infer from Eqs. (32) and (34) that for a walk on an s-dimensional space lattice, if Eq. (33) holds then in analogy with Eq. (32), the effective dimensionality of the random walk[7,8] is

$$r = s + 2 - \mu. \tag{35}$$

For example, even in one dimension the random walker need not return to the starting point if $\mu < 1$, and thus the random walker exhibits the qualitative behavior of a walker in a higher dimensional space. We note that as $0 < \mu \leqslant 2$, we have

$$s \leqslant r < s + 2. \tag{36}$$

The above definition of effective dimension is of course not the only one which may be based on random walk properties. Other random walk statistics (such as the mean number of distinct sites visited in n steps[21]) may also be used to define a random walk dimension.

VI. DISCUSSION

To cast light on the problem of assigning useful dimensions to nonstandard lattices, we have considered the problem of nonstandard interactions on an s-dimensional simple cubic lattice. We have shown that it is possible to define a harmonic lattice dimension h, a scaling dimension f, and a random walk dimension r. The examples considered yield the following inequalities and relations among the dimensions:

$$h \geqslant s, \quad 0 < f \leqslant 2, \quad s \leqslant r < s+2,$$
$$h = 2s/f, \quad r = s + 2 - f.$$

If we eliminate s from the preceding equations we find the relation between f, h, and r:

$$fh = 2[r + f - 2]. \tag{37}$$

Since $r \geqslant 1$, it follows that

$$fh \geqslant 2[f - 1]. \tag{38}$$

Although this inequality has been derived for cubic lattices supporting fractal interactions (necessarily restricted to the case $f < 2$) we conjecture that it holds *for a class of fractal lattices* with f the usual fractal dimension and h the harmonic lattice dimension, calculated with nearest-neighbor couplings only. It is easily verified that the truncated n-simplex lattice[3] [$f = \ln n/\ln 2$, $h = 2 \ln n/\ln(n+2)$] and the modified rectangular lattice[3] [$f = 2$, $h = 1.5$] are consistent with (38), and indeed the inequality holds strictly. Counterexamples to (38) might be able to be constructed by forming suitable direct products[22] of fractal lattices, but the authors have encountered none to date. A systematic investigation of possible relations between h and the key topological parameters of a fractal lattice (such as f, lacunarity, ramification,[2] etc.) would be of considerable interest.

APPENDIX

It is well known[23] that for a one-dimensional lattice walk, $\lambda(\theta) = 1$ if and only if $\theta = 2\pi m/a$ (a = lattice spacing, $m = 0, \pm 1, \pm 2, \ldots$). It is not possible to generalize this in the obvious manner to an s-dimensional cubic lattice [$\lambda(\theta) = 1$ only at the centers of the Brillouin zones] without some further restrictions on the walk. We prove (for a lattice of unit spacing) that if $s = 2$ or $s = 3$, the structure function $\lambda(\boldsymbol{\theta})$ attains the value 1 at no point other than $\boldsymbol{\theta} = \mathbf{0}$ inside or on the boundary of the first Brillouin zone $B = \{\boldsymbol{\theta} | -\pi \leqslant \theta_j \leqslant \pi\}$, provided that certain nearest-neighbor transitions have nonzero probability. [Equivalently, in the lattice dynamics terminology, $W(\boldsymbol{\theta})$ vanishes in B only at $\boldsymbol{\theta} = \mathbf{0}$, if certain nearest-neighbor couplings are nonzero.]

We assume that for $s \geqslant 2$ and unit lattice spacing,

$$\lambda(\boldsymbol{\theta}) = \sum_{\mathbf{l}} \exp(i\mathbf{l}\cdot\boldsymbol{\theta}) p(\mathbf{l}) \tag{A1}$$

attains the value 1 at some point $\boldsymbol{\phi}$ in B, with $\boldsymbol{\phi} \neq \mathbf{0}$. It follows from Eq. (A1) that

$$\begin{aligned} 0 &= \sum_{\mathbf{l}} [\exp(i\mathbf{l}\cdot\boldsymbol{\phi}) - 1] p(\mathbf{l}) \\ &= \sum_{\mathbf{l}} [\cos(\mathbf{l}\cdot\boldsymbol{\phi}) - 1] p(\mathbf{l}) + i \sum_{\mathbf{l}} \sin(\mathbf{l}\cdot\boldsymbol{\theta}) p(\mathbf{l}). \end{aligned} \tag{A2}$$

As the summand in the real part of the right-hand side of Eq. (A2) is never positive, we deduce that $p(\mathbf{l})$ can be nonzero at most at points $\mathbf{l}$ lying on the family of hyperplanes,

$$\mathbf{l}\cdot\boldsymbol{\phi} = 2\pi m \quad (m = 0, \pm 1, \pm 2, \ldots). \tag{A3}$$

If we define $d = 2\pi/|\boldsymbol{\phi}|$ and $\hat{\mathbf{n}} = \boldsymbol{\phi}/|\boldsymbol{\phi}|$, we may rewrite the family of hyperplanes as

$$\mathbf{l}\cdot\hat{\mathbf{n}} = md, \tag{A4}$$

with d the distance from the origin ($\mathbf{l} = \mathbf{0}$) of the two hyperplanes closest to the origin which do not pass through the origin. Since $\boldsymbol{\phi}$ lies in B, $|\boldsymbol{\phi}|^2 \leqslant s\pi^2$, with equality if and only if $\boldsymbol{\phi}$ lies at a corner of B, and so

$$d \geqslant 2s^{-1/2}. \tag{A5}$$

In particular, if $s = 2$ or $s = 3$, then $d > 1$, and the nearest-neighbor sites for which $p(\mathbf{l}) > 0$ can only lie on the line or plane through the origin ($\mathbf{l}\cdot\boldsymbol{\phi} = 0$). Hence for $s = 2$ or $s = 3$, if $p(\mathbf{l}) > 0$ for s orthogonal nearest-neighbor transitions, $\lambda(\boldsymbol{\theta}) = 1$ in B if and only if $\boldsymbol{\theta} = \mathbf{0}$. When $s = 4$, (A5) shows only that $d \geqslant 1$, with equality if and only if $\boldsymbol{\phi}$ is one of the 16 vertices of the Brillouin zone B, so that if $p(\mathbf{l}) > 0$ for four orthogonal nearest-neighbor transitions, $\lambda(\boldsymbol{\theta}) = 1$ on B at $\boldsymbol{\theta} = \mathbf{0}$ and at most also at the 16 vertices of B. For $s > 4$, the present argument gives no information.

The hypothesis that certain nearest-neighbor transitions have nonzero probability cannot be removed. To see this, we note that a nearest-neighbor random walk on a body-centered cubic lattice (in three dimensions) can be described in terms of a random walk or a simple cubic lattice, with nearest-neighbor transitions forbidden. It has been pointed out by Joyce[24] that for such a walk, singular points other than the origin must be considered in Eq. (8).

The breakdown of the simple argument used here to examine where $\lambda(\boldsymbol{\theta}) = 1$ for $s > 4$ appears to be another curious example of the strong qualitative difference between walks in four or fewer dimensions and walks in more than four dimensions.[25,26]

[1] D. R. Nelson and M. E. Fisher, Ann. Phys. (NY) **91**, 226 (1975).

[2] Y. Gefen, B. B. Mandelbrot, and A. Aharony, Phys. Rev. Lett. **45**, 855 (1980).

[3] D. Dhar, J. Math. Phys. **18**, 577 (1977); *ibid.* **19**, 5 (1978).

[4] M. Kaufman and R. B. Griffiths, Phys. Rev. B **24**, 496 (1981).

[5] See also H. J. Stapleton, J. P. Allen, C. P. Flynn, D. G. Stinson, and S. R. Kurtz, Phys. Rev. Lett. **45**, 1456 (1980).

[6] D. Dhar, Pramāna **15**, 545 (1980).

[7] B. D. Hughes, M. F. Shlesinger, and E. W. Montroll, Proc. Natl. Acad. Sci. USA **78**, 3287 (1981).

[8] B. D. Hughes, E. W. Montroll, and M. F. Shlesinger, J. Stat. Phys. **28**, 111 (1982).

[9] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic, New York, 1963).

[10] A. Isihara, *Statistical Physics* (Academic, New York, 1971), Chap. 8.

[11] E. W. Montroll, Proc. Symp. Appl. Math. **16**, 193 (1964).

[12] E. W. Montroll and G. H. Weiss, J. Math. Phys. **6**, 167 (1965).

[13] E. W. Montroll, in *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability* (U. Cal. Berkeley, 1955), pp. 209–240.

[14] F. Spitzer, *Principles of Random Walk* (Springer, New York, 1976) 2nd ed., pp. 67–70.

[15] The lattice must be infinite to yield a continuous frequency distribution.

[16] See, for example, the elegant discussion by R. J. Rubin and R. Zwanzig, J. Math. Phys. **2**, 861 (1961), of the frequency spectrum of a Cayley tree (Bethe lattice).

[17] J. E. Gillis and G. H. Weiss, J. Math. Phys. **11**, 1307 (1970).

[18] B. B. Mandelbrot, *Fractals: Form, Chance and Dimension* (Freeman, San Francisco, 1977).

[19] G. Pólya, Math. Ann. **84**, 149 (1921).

[20] The generating function formalism holds for an arbitrary single-step probability distribution $p(\mathbf{l})$, and allows u to be calculated for face-centered and body-centered cubic lattices also.

[21] If $P(\mathbf{0}, 1 - n^{-1}) \sim n^{\beta} L(n)$, with L "slowly varying" as $n \rightarrow \infty$, then the expected number of distinct sites visited after n steps is (Ref. 12) $\langle S_n \rangle \sim n / P(\mathbf{0}, 1 - n^{-1})$ as $n \rightarrow \infty$. For the long-ranged jump distributions considered in the present paper it can be shown that $\langle S_n \rangle \sim C n^{h/2}$ if $h < 2$. It has been brought to our attention by the referee that if $h < 2$ this relation also holds for fractal lattices.

[22] D. Dhar, J. Phys. A **14** L185 (1981).

[23] E. Lukacs, *Characteristic Functions* (Griffin, London, 1970), 2nd ed.

[24] G. S. Joyce, J. Math. Phys. **12**, 1390 (1971).

[25] K. Lindenberg, V. Seshadri, K. E. Shuler, and G. H. Weiss, J. Stat. Phys. **23**, 11 (1980).

[26] P. Erdös and S. J. Taylor, Acta Math. Acad. Sci. Hung. **11**, 231 (1960).

Lorentz invariance of the extended object

J. L. Jacquot
Centre Universitaire de Sétif, Sétif, Algérie,[a] *and Centre de Recherches Nucléaires, Université Louis Pasteur, 67037 Strasbourg Cedex, France*

M. Umezawa[b]
Centre de Recherches Nucléaires, Université Louis Pasteur, 67037 Strasbourg Cedex, France

(Received 11 December 1981; accepted for publication 9 April 1982)

It has been known for a few years that the Heisenberg field of the extended object can be obtained in expanded form as a power series of quantum operators and creation and annihilation operators by solving the Yang–Feldman equation. Such an expression is called a dynamical map (mapping of the infield into the Heisenberg field). We will show that the Heisenberg field thus expressed (the dynamical map) is Lorentz covariant if it satisfies the equal time canonical commutation relation. In this paper we limit ourselves to the invariance of the first order term. Also, our Heisenberg field is $(1 + 1)$ dimensional and is of the tree approximation. In the course of the calculation, we find that the quantum mechanical operator and the quantized field may be mixed by the Lorentz transformation if the space derivative of the classical field soliton solution is assumed to decrease not faster than $1/x^2$. It indicates that the Hilbert space may not be the direct product of two sub-Hilbert spaces, even though it is a product of two subspaces, one of which is a Fock space of the quantized field and the other is of the quantum mechanical operator.

PACS numbers: 11.10. – z, 11.30.Cp

I. INTRODUCTION

In the last few years, through the study of extended objects in quantum field theory,[1] great progress has been made in our understanding of physical systems which present both quantum mechanical and quantum field properties. One of the main approaches to the subject uses the so-called boson transformation method,[2] which is applied to systems described by the scalar field theory.

In this method, first using the Yang–Feldman strategy, the scalar Heisenberg field is given by a functional of the asymptotical free field. The trick is to introduce a C-number function which obeys the same homogeneous equation of motion as the free field. Thus the Heisenberg field is now, by means of the boson transformation method, a functional of the sum of the free field and the C-number function. This full Heisenberg field, called the dynamical map, describes the extended object, and due to the boson transformation theorem,[2] obeys the same Heisenberg equation of motion as the free field. Now, if we consider only static extended objects, i.e., where the C-number function depends only on the space variable, the requirement that the dynamical map obey the equal time canonical commutation relation implies the introduction of classical quantum coordinates. As has been shown in the tree approximation, these quantum coordinates reflect the invariance of the dynamical map under space translations.[3] Moreover, in $(1 + 1)$ dimensions, if Lorentz invariance is assumed for the description of the extended object in the one particle approximation, then the dynamical map depends only on two generalized coordinates[4] X and T. Therefore we can assume that the dynamical map can be written as $\Psi(X,T)$ at least in the tree approximation.

Our purpose in this paper is to show in the tree approximation that $\Psi(X,T)$ describes completely the extended object in the second order in quantum coordinates and field operators. Thus if the equal time commutation relations are assumed, then in this approximation, by calculating the energy stress tensor, we obtain the infinitesimal Lorentz transformation of the dynamical map $\Psi(X,T)$ to first order in the quantum coordinates and field operators. Therefore as expected, by the Lorentz transformation, quantum coordinates and field operators mix together.

The paper is organized into five sections. In Sec. II we recall briefly the main results which concern the dynamical map in the tree approximation. In Sec. III, in this approximation, assuming that the extended object in $(1 + 1)$ dimensions is described by the dynamical map $\Psi(X,T)$ and that this latter function satisfies the equal time commutation relations, we obtain the equal time commutation relations of the quantum coordinates and field operators up to third order. Section IV is devoted to the calculation of the second order field stress tensor and to the first order infinitesimal Lorentz transform of the quantum coordinates and field operators. In the final section we draw some conclusions.

II. USEFUL PROPERTIES OF THE DYNAMICAL MAP IN THE TREE APPROXIMATION

If one begins with the Heisenberg equation of motion for the full scalar field $\psi(x,t)$,

$$(\Box - m^2)\psi = \frac{\delta V}{\delta \psi}, \tag{2.1}$$

and performs a static boson transformation over the free field φ_0, i.e.,

$$\varphi_0(x,t) \rightarrow \varphi_0(x,t) + f(x), \tag{2.2}$$

which together with the C-number static function $f(x)$ satisfy the homogeneous part of Eq. (2.1), then by virtue of the

[a] Present address.

[b] Postal address: Centre de Recherches Nucléaires, Physique Théorique des Hautes Energies, 67037 Strasbourg Cedex, France.

Yang–Feldman method and by the boson transformation theorem,[2] the dynamical map $\Psi(x,t)$, which describes the extended object, is given in the tree approximation by the Taylor expansion

$$\Psi(x,t) = \frac{1}{n!} \sum_{n=0}^{\infty} \psi^{(n)}(x,t), \tag{2.3}$$

where $\psi^{(0)}$ stands for the vacuum expectation value $\phi(x)$ of the dynamical map Ψ and $\psi^{(n)}$ is an n-order field operator in the free field φ_0. In addition the field operator $\psi^{(n)}$ is determined by a recursion formula,[3] which gives the different equations of motion up to third order:

$$(\Box - m^2)\phi = \frac{\delta V}{\delta\phi}, \tag{2.4}$$

$$(\Box - m^2)\psi^{(1)} = \frac{\delta^2 V}{\delta^2\phi}\,\psi^{(1)}, \tag{2.5}$$

$$\left(\Box - m^2 - \frac{\delta^2 V}{\delta^2\phi}\right)\psi^{(2)} = \frac{\delta^3 V}{\delta^3\phi}\,\psi^{(1)^2}. \tag{2.6}$$

From the fact that the equal time commutation relation must have the canonical form, it follows that the field $\psi^{(1)}$ is a sum of a field χ, which describes both the scattered and bound states, and a quantum mechanical piece (the quantum coordinates) so that in $(1+1)$ dimensions we have[3]

$$\left[\chi' + \frac{q}{\sqrt{M}}\,\partial_{x'}\phi', \dot{\chi} + \frac{p}{\sqrt{M}}\,\partial_x\phi\right] = i\delta(x-x'), \tag{2.7}$$

with

$$[q,p] = i. \tag{2.8}$$

Here and in the following, the prime index stands only for the space argument and the constant M is the mass of the classical extended object defined by

$$M = \int (\partial_x\phi)^2\,dx. \tag{2.9}$$

With the use of the recursion formula, it has been shown that we can check the dependence of the ψ on the quantum coordinates to all order.[3,5] The solution ψ of the Euler equation (2.1) thus obtained is not unique. The ψ given in Refs. 4 and 5 is the one obtained by replacing in (2.3) the coordinates (x,t) by the generalized quantum coordinates (X,T), which are given by[4,5]

$$X = x\cosh A + t\sinh A + B, \tag{2.10}$$

$$T = t\cosh A + x\sinh A + B\tanh A, \tag{2.11}$$

where the time independent operators A and B are defined, in terms of the covariant quantum coordinate[4] q, by

$$\cosh A = (1-\dot{q}^2)^{-1/2}, \tag{2.12}$$

$$\sinh A = \dot{q}(1-\dot{q}^2)^{-1/2}, \tag{2.13}$$

$$B = q(0)(1-\dot{q}^2)^{-1/2}. \tag{2.14}$$

We can obtain another solution by adopting the following T given in (13.18) of Ref. 6, instead of (2.11).

$$T = t\cosh A + x\sinh A. \tag{2.15}$$

In fact there are infinitely many other solutions whose T's differ from (2.11) and (2.15). Then we cannot expect that all of these solutions can satisfy the canonical commutation relation. Only some of them should do. In this paper, however, we expand the ψ into a power series on A, B, and χ, and limit ourselves to the second order. Then, in this paper we find that all of the canonical solutions should be identical in this approximation.

III. SECOND ORDER EQUAL TIME COMMUTATION RELATION OF THE QUANTUM MECHANICAL AND FIELD OPERATORS

The Taylor expansion of the dynamical map (2.3) in terms of the generalized quantum coordinates (2.10), (2.11) or alternatively (2.10), (2.15), up to the third order in the quantum mechanical and field operators, is given by

$$\Psi(X,T) = \phi(x) + \Psi^{(1)}(x,t) + \Psi^{(2)}(x,t), \tag{3.1}$$

where for convenience we have defined the first and second order pieces of the dynamical map as[6]

$$\Psi^{(1)}(x,t) = C\partial_x\phi + \chi, \tag{3.2}$$

$$\Psi^{(2)}(x,t) = \tfrac{1}{2}C^2\partial_x^2\,\phi + \tfrac{1}{2}xA^2\partial_x\phi + C\partial_x\chi + xA\dot{\chi} + \tfrac{1}{2}\psi^{(2)}. \tag{3.3}$$

Here the quantum mechanical operator C, defined as

$$C = B + At, \tag{3.4}$$

with the use of the commutation relation (2.8) and the definitions (2.12), (2.14) of the operators A and B, is seen to satisfy

$$[C,A] = i/M. \tag{3.5}$$

Now if we assume that the equal time commutation relation holds for the dynamical map (3.1), we must have

$$[\Psi',\dot{\Psi}] = i\delta(x-x'), \tag{3.6}$$

$$[\Psi',\Psi] = 0. \tag{3.7}$$

Actually, with the use of the commutation relations (2.7) and (3.5), it is easy to check that the first order piece $\Psi^{(1)}$ of the dynamical map satisfies these two commutation relations so that their expansion yields two sets of expressions. One set is given by their first order vanishing piece and the other is obtained in keeping only the second order piece, which is proportional to the quantum mechanical operator C.

The first order piece leads to

$$\frac{i}{M}\partial_{x'}\phi'(\partial_x\chi + x\ddot{\chi}) + \tfrac{1}{2}[\chi',\dot{\psi}^{(2)}] + \frac{i}{M}\partial_x\phi\partial_{x'}\chi' + \tfrac{1}{2}[\psi^{(2)\prime},\dot{\chi}] = 0, \tag{3.8}$$

$$\frac{i}{M}x\partial_{x'}\phi'\dot{\chi} + \tfrac{1}{2}[\chi',\psi^{(2)}] - \frac{i}{M}x'\partial_x\phi\dot{\chi}' + \tfrac{1}{2}[\psi^{(2)\prime},\chi] = 0. \tag{3.9}$$

As for the second order piece we have

$$C\partial_{x'}\left(\frac{i}{M}\partial_{x'}\phi'(\partial_x\chi + x\ddot{\chi}) + \tfrac{1}{2}[\chi',\dot{\psi}^{(2)}]\right) + C\partial_x\left(\frac{i}{M}\partial_x\phi\partial_{x'}\chi' + \tfrac{1}{2}[\psi^{(2)\prime},\dot{\chi}]\right), \tag{3.10}$$

$$C\partial_{x'}\left(\frac{i}{M}x\partial_{x'}\phi'\dot{\chi} + iAx\delta + \tfrac{1}{2}[\chi',\psi^{(2)}]\right) + C\partial_x\left(-\frac{i}{M}x'\partial_x\phi\dot{\chi}' - iAx'\delta + \tfrac{1}{2}[\psi^{(2)\prime},\chi]\right). \tag{3.11}$$

The fact that the expression (3.9) does not appear exactly in (3.11) does not matter here because to have an exact second order piece, we must deal with the third order expansion of the dynamical map $\Psi(X,T)$.

It follows from the two latter sets of relations that the different equal time commutation relations are

$$[\dot\psi^{(2)},\chi'] = \frac{2i}{M}\partial_{x'}\phi'(\partial_x\chi + x\ddot\chi), \tag{3.12}$$

$$[\dot\chi',\psi^{(2)}] = \frac{2i}{M}\partial_{x'}\phi'\partial_x\chi, \tag{3.13}$$

$$[\psi^{(2)},\chi'] = \frac{2i}{M}x\partial_{x'}\phi'\dot\chi. \tag{3.14}$$

It is easy to verify that these three commutation relations are self-consistent.

As expected, the time derivative of the commutator (3.14) is given by the difference of the commutators (3.12) and (3.13). Also summing the time derivatives of the two first commutators (3.12), (3.13) and using the equations of motion (2.5), (2.6) we obtain the condition

$$\left\{\left(\partial_x^2 - m^2 - \frac{\delta^2 V}{\delta^2\phi}\right) - \left(\partial_{x'}^2 - m^2 - \frac{\delta^2 V'}{\delta^2\phi'}\right)\right\}[\psi^{(2)},\chi']$$
$$= \frac{2i}{M}\partial_{x'}\phi'(2\partial_x\dot\chi + x\dddot\chi), \tag{3.15}$$

which by virtue of the equation of motion (2.5) is exactly fulfilled by the expression for the commutator (3.14).

In addition we can notice that the last commutator is proportional to the space argument of the two particle field $\psi^{(2)}$, so that near the origin this commutator behaves like that of the free fields. Moreover in this case the first two commutators are equal.

At present, we can calculate the generators of the infinitesimal Lorentz transformation to second order in the quantum mechanical and field operators.

IV. LORENTZ TRANSFORMATION

A. The second order field stress tensor

In (1 + 1) dimension scalar field theory, the generator of the infinitesimal Lorentz transformation is given by

$$M_{0x} = \int (xT_{00} + tT_{0x})\,dx, \tag{4.1}$$

where $T_{\mu\nu}$ is the usual energy stress tensor. If we take the second order expression (3.1) for the dynamical map and write the potential as a Taylor expansion in terms of the classical field ϕ, then by use of the equation of motion (2.4) and the asymptotical properties of this field [due to the finiteness of the classical part of M_{0x},$x(\partial_x\phi)^2$ vanishes as x goes to infinity], the generator of the Lorentz transformation becomes in the tree approximation

$$M_{0x} = M_{0x}^{(0)} + M_{0x}^{(1)} + M_{0x}^{(2)}. \tag{4.2}$$

Here the zeroth, first, and second order pieces in the quantum mechanical and field operators are defined by

$$M_{0x}^{(0)} = \int x\left[\tfrac{1}{2}(\partial_x\phi)^2 + \frac{m^2}{2}\phi^2 + V(\phi)\right]dx, \tag{4.3}$$

$$M_{0x}^{(1)} = -MB + \int x\partial_x(\partial_x\phi\chi)\,dx + t\int\chi\partial_x\phi\,dx, \tag{4.4}$$

$$M_{0x}^{(2)} = \int x\left[\tfrac{1}{2}\dot\Psi^{(1)^2} + \tfrac{1}{2}(\partial_x\Psi^{(1)})^2 + \frac{1}{2}\left(m^2 + \frac{\delta^2 V}{\delta^2\phi}\right)\Psi^{(1)^2}\right.$$
$$\left. + \partial_x(\Psi^{(2)}\partial_x\phi)\right]dx + t\int(\dot\Psi^{(1)}\partial_x\Psi^{(1)}$$
$$+ \dot\Psi^{(2)}\partial_x\phi)\,dx, \tag{4.5}$$

M being the mass of the classical extended object. The $M_{0x}^{(0)}$ is merely a constant. In addition the first term of expression (4.4) is the generator of the Lorentz transformation in the no-particle sector.[4]

B. The infinitesimal Lorentz transformation of the full first order dynamical map

Using the equation of motion (2.5) of the one particle field, it is easy to see that

$$[\chi',\ddot\chi] = 0. \tag{4.6}$$

Therefore, since this field also commutes with $\Psi^{(1)}$ and all the quantum mechanical operators, it follows that

$$[\chi',M_{0x}] = \int x\,[\dot\Psi^{(1)}[\chi',\dot\Psi^{(1)}] + \partial_x([\chi',\Psi^{(2)}]\,\partial_x\phi)]\,dx$$
$$+ t\int([\chi',\dot\Psi^{(1)}]\,\partial_x\Psi^{(1)} + [\chi',\dot\Psi^{(2)}]\,\partial_x\phi)\,dx. \tag{4.7}$$

The quantum mechanical operators A and B are time independent, so from the expression of the first and second order pieces (3.2), (3.3) of the dynamical map and the equal time commutation relations (2.7) and (3.12), (3.13), keeping in mind the asymptotical property of the classical field ϕ, we can verify that

$$[\chi,M_{0x}] = i(x\dot\chi + t\partial_x\chi)$$
$$- \frac{i}{M}\partial_x\phi\,[x\partial_x\phi\,(x\dot\Psi^{(1)} + t\partial_x\Psi^{(1)}]_{-\infty}^{+\infty}. \tag{4.8}$$

Here the term proportional to the time variable in the bracket comes from the expression

$$\int(x\ddot\chi + 2\partial_x\chi)\partial_x\phi\,dx, \tag{4.9}$$

by using the equation of motion (2.4), (2.5). As for $\Psi^{(1)}$, it is the first order piece of the dynamical map defined in (3.2). The last commutation relation shows that in the Lorentz transformation the one particle quantum field, the classical field ϕ, and the quantum mechanical operator A mix together in a coherent form. In the asymptotical limit it becomes

$$[\chi,M_{0x}]_{x\to\infty} = i(x\dot\chi + t\partial_x\chi)_{x\to\infty} \tag{4.10}$$

which, in this case, gives as expected the infinitesimal Lorentz transformation of the asymptotical free field. This result is not surprising since the asymptotical limit of the equal time commutation relation (2.7) has exactly the canonical free field form. If the space derivative of the classical field decreases more rapidly than $1/x^2$ as x goes to infinity, then the bracket of the right-hand side of (4.8) vanishes, and there is no mixing in the Lorentz transform of the one particle field. This is the case for the sine-Gordon model.[7]

To calculate the infinitesimal Lorentz transform of the quantum mechanical operators A and B, it is better to modify the expression $T_{00}^{(2)}$ of the energy part of the generator $M_{0x}^{(2)}$.

From the equation of motion (2.5) it follows that

$$T_{0x}^{(2)} = \tfrac{1}{2}[\dot{\Psi}^{(1)^2} + \partial_x(\Psi^{(1)}\partial_x\Psi^{(1)} + 2\partial_x\phi\Psi^{(2)}) - \ddot{\chi}\Psi^{(1)}]. \tag{4.11}$$

Then with the help of the commutation relation (3.5) and the asymptotical property of the classical field, we obtain

$$[A,M_{0x}] = i - \frac{i}{2M}[x(2\partial_x\phi\partial_x\chi + \partial_x^2\phi\chi)]_{-\infty}^{+\infty}, \tag{4.12}$$

$$[B,M_{0x}] = \frac{it}{2M}[x(2\partial_x\phi\partial_x\chi + \partial_x^2\phi\chi)]_{-\infty}^{+\infty} + \frac{i}{M}[x\partial_x\phi(x\dot{\Psi}^{(1)} + t\partial_x\Psi^{(1)})]_{-\infty}^{+\infty}. \tag{4.13}$$

Thus the first order Lorentz transform of the quantum mechanical operators A and B has a one-particle free-field part. From these last commutation relations, the generator M_{0x} being written to second order in the quantum mechanical and field operator, we can obtain the exact Lorentz transform of the generalized quantum coordinates (2.10), (2.11) up to the second order terms. They are

$$[X,M_{0x}] = iT^{(1)} + \frac{i}{M}[x\partial_x\phi(x\dot{\Psi}^{(1)} + t\partial_x\Psi^{(1)})]_{-\infty}^{+\infty}, \tag{4.14}$$

$$[T,M_{0x}] = iX^{(1)} - \frac{ix}{2M}[x(2\partial_x\phi\partial_x\chi + \partial_x^2\phi\chi)]_{-\infty}^{+\infty}, \tag{4.15}$$

where, as usual, $T^{(1)}$ and $X^{(1)}$ stand for the first order expansion of the generalized quantum coordinates. If the term $x\partial_x\phi$ vanishes as x tends to infinity, then we see that the generalized quantum coordinates play the role of the space-time variable for the extended object in the no particle sector.

What about the first order infinitesimal Lorentz transform of the dynamical map? In order to have the exact first order piece of the Lorentz transform of the dynamical map, we must deal with its first and second order pieces (3.1). Owing to the expression (4.8) and (4.12), (4.13) of the Lorentz transform of the one particle field χ and the operators A and B and to the fact that the sum of the three last terms of the second order piece (3.3) of the dynamical map commute with the first order piece (4.4) of the energy stress tensor, it follows that

$$[\phi + \Psi^{(1)} + \Psi^{(2)},M_{0x}] = ix(\dot{\phi} + \dot{\Psi}^{(1)}) + it\partial_x(\phi + \Psi^{(1)}). \tag{4.16}$$

The first and second order pieces $\Psi^{(1)}$ and $\Psi^{(2)}$ are defined in (3.2) and (3.3). So up to the second order the dynamical map, which describes the classical, quantum mechanical, and one partical field properties of the extended object, is at least in the tree approximation, a relativistic covariant entity. But if we dissociate the extended object into its classical and quantum mechanical pieces on the one hand and its quantum field piece on the other, then the transformation relations (4.8) and (4.12), (4.13) show that in a change of reference frame the quantum mechanical piece mixes with the quantum field and that the quantum field piece mixes with classical field and quantum mechanical operators. In this sense the quantum mechanical or field aspect of the extended object depends on the frame of reference, just as the magnetic or electric aspect of the electromagnetic field.

V. SUMMARY

If we construct the dynamical map of a static scalar extended object using the Yang–Feldman method, the dynamical map is not determined uniquely at each order of the perturbation expansion due to the zero energy modes of the soliton solutions. In Ref. 5, it was shown that if we assume the canonical commutation relations to be held for the set of quantum mechanical and quantum field operators, then the dynamical map can be given exactly, at least to the second order, by $\Psi(X,T)$.

In this paper we have shown in the tree approximation that the first order piece of the dynamical map is Lorentz covariant, assuming that the static scalar extended object is described by the dynamical map $\Psi(X,T)$ and that the equal time canonical commutation relations work. Thus the static scalar extended object in $(1 + 1)$ dimensions is fully described by this dynamical map: In order to have the dynamical map of the nonstatic extended object, we have to perform a Lorentz transformation. Moreover, the main result is that the quantum mechanical and quantum field aspects of the extended object depend on the frame of reference if the space derivatives of the classical field soliton solution do not rapidly fall off to zero. If it appears that this fact is a general feature of an extended object with more degrees of freedom, as for instance spin and colors, then this could be very attractive for our understanding of hadron physics.

[1] J. L. Gervais and A. Neveu, Phys. Rep. **23**, 237 (1976); R. Jackiw, Rev. Mod. Phys. **49**, 681 (1977); for a review of the quantum theory of solitons see, for example, L. D. Faddeev and V. E. Korepin, Phys. Rep. **42**, 1 (1978).

[2] L. Leplae, F. Mancini, and H. Umezawa, Phys. Rev. B **2**, 3594 (1970); H. Matsumoto, N. Y. Papastamatiou, and H. Umezawa, Nucl. Phys. B **82**, 45 (1974); **97**, 90 (1975); H. Matsumoto, P. Sodano, and H. Umezawa, Phys. Rev. D **19**, 511 (1979); H. Matsumoto, G. Oberlechner, H. Umezawa, and M. Umezawa, J. Math Phys. **20**, 2088 (1979); H. Matsumoto, G. Semenoff, H. Umezawa, and M. Umezawa, *ibid.* **21**, 1761 (1980).

[3] See, for example, the two last references of Ref. 2.

[4] H. Matsumoto, N. J. Papastamatiou, H. Umezawa, and M. Umezawa, Phys. Rev. D **23**, 1339 (1981).

[5] G. Semenoff, H. Matsumoto, and H. Umezawa, J. Math. Phys. **22**, 2208 (1981).

[6] M. Umezawa, Phys.Rev. D **24**, 1548 (1981).

[7] G. Oberlechner, M. Umezawa, and Ch. Zenses, Lett. Nuovo Cimento **23**, 641 (1978), and references contained therein.

Scalar interactions of supersymmetric relativistic spinning particles

Peter Van Alstine
Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37235
Horace Crater
Department of Physics, The University of Tennessee Space Institute, Tullahoma, Tennessee 37388

(Received 18 January 1982; accepted for publication 15 March 1982)

We introduce scalar interactions for the relativistic spinning particle in such a way as to preserve a supersymmetry that leaves a special position-variable invariant. This generates systems of particles in scalar interaction with a supersymmetry for each spinning particle. For two-particle systems the supersymmetry eliminates all spin complications and reduces consistency problems to those of a purely bosonic system. Once the latter are disposed of, our approach leads to consistent systems of quantum mechanical wave equations.

PACS numbers: 11.30.Pb, 11.10.Qr

In the last six years, many authors have applied Dirac's constrained Hamiltonian mechanics to interacting spinless particles to obtain consistent systems of relativistic wave equations.[1] In a previous paper,[2] we even used such techniques to extend a nonrelativistic heavy quark potential developed by Richardson[3] (for the ψ and Υ families) to the relativistic domain of the light and intermediate mass vector mesons. Although our formalism gave a good account of the ground states and observed radial excitations, it neglected quark spin from the start. Quantum mechanical descriptions derived from classical constraint systems are not easily extended to include spin, however, without upsetting the delicate consistency of the original classical dynamics. One way to avoid this difficulty is to build a consistent version of classical or "pseudoclassical" spin into the canonical formalism before quantization. For the free particle in an external field,[4,5] and system of particles with a collective spin,[6] these problems have been overcome by other authors through the use of one-dimensional, locally supersymmetric actions analogous to those for supergravity.[7] However, in order to deal with the more complicated system of two interacting particles, each with its own constituent spin, we need to find a way to introduce a supersymmetry[8] for each spinning particle and preserve all of them against breaking induced by interaction.

We shall see that when an initially free supersymmetric spinning particle is put in scalar interaction with an external agent, a spinless particle, or a second supersymmetric spinning particle, the requirement that the interacting system remain supersymmetric determines the spin-dependence of the potential. The resulting supersymmetric actions eventually lead to first-class Hamiltonian constraint systems suitable for quantization.

First, we remind the reader of the corresponding treatment for spinless particles. A free particle is described by the arc length action

$$S=\int L\,d\tau=\int -m(-\dot{x}^2)^{1/2}d\tau, \qquad (1)$$

which leads to the mass shell constraint on the particle's four-momentum,

$$\mathscr{H}=p^2+m^2\approx 0. \qquad (2)$$

One introduces interaction with an external scalar field by letting $m\to m(x)$ in (1) and (2). A similar Lagrangian for two spinless particles leads to two mass shell constraints

$$\mathscr{H}_1=p_1{}^2+m_1{}^2\approx 0,\quad \mathscr{H}_2=p_2{}^2+m_1{}^2\approx 0. \qquad (3)$$

Todorov[1] found that such constraints become compatible (i.e., $\{\mathscr{H}_1,\mathscr{H}_2\}=0$) when the potentials m_i obey

$$m_1{}^2(x)-m_2{}^2(x)=\text{const}, \qquad (4)$$

where $x=x_1-x_2$ depends only on the components of x perpendicular to the total four-momentum. That is, $m_i=m_i(x_\perp{}^2/2)$,[9] where

$$x_\perp^\mu=(g^{\mu\nu}-P^\mu P^\nu/P^2)x_\nu, \qquad (5)$$

with $P=p_1+p_2$.

We introduce spin at the pseudoclassical level[4-6] through anticommuting degrees of freedom (elements of a Grassmann algebra) that, together with ordinary degrees of freedom, satisfy

$$AB=(-1)^{\epsilon_A\epsilon_B}BA;\quad \epsilon_{\text{odd}}=1,\quad \epsilon_{\text{even}}=0. \qquad (6)$$

We describe a single free spinning particle by modifying a Lagrangian recently proposed by Galvao and Teitelboim[5] so that its odd Lagrange multiplier v is a coordinate (instead of a velocity).

$$S=\int L\,d\tau=\int\Big[-m(-\dot{x}^2)^{1/2}[1+iv(\hat{u}\cdot\theta+\theta_5)] + \frac{i}{2}\dot{\theta}\cdot\theta+\frac{i}{2}\dot{\theta}_5\theta_5\Big]d\tau, \qquad (7)$$

where $\hat{u}=\dot{x}/(-\dot{x}^2)^{1/2}$, while θ, θ_5, and v are odd Grassmann functions of τ. This action consists of a bosonic length of world–line piece, plus an odd constraint that leads to the Dirac equation, plus a kinetic term for the spin degrees of freedom. It leads directly to Galvao and Teitelboim's Dirac and mass shell constraints in phase space

$$\mathscr{S}=p\cdot\theta+m\theta_5\approx 0,\quad \mathscr{H}=p^2+m^2\approx 0. \qquad (8)$$

This action is left invariant (on shell[10]) by the supersymmetry transformation

$$\delta x=\epsilon(\theta-\hat{p}\theta_5),\quad \delta\theta=-i\epsilon p,$$
$$\delta\theta_5=-i\epsilon(-p^2)^{1/2},\quad \delta v=-i\dot{\epsilon}/(-\dot{x}^2)^{1/2} \qquad (9)$$

0022-2488/82/091697-03$02.50

[where $p \equiv \partial L/\partial \dot{x} = m\hat{u}(1 + iv\theta_5) - imv\theta$, $\hat{p} = p/(-p^2)^{1/2}$] whose Noether generator is

$$G = p\cdot\theta + (-p^2)^{1/2}\theta_5. \tag{10}$$

That is, $\delta S = \frac{1}{2}\int d\tau\,(d/d\tau)\,(\epsilon G) \approx 0$. Since our generator G differs from S, our supersymmetry transformations (9) and their consequences are to be contrasted[11] with those in Refs. 4–7.

With pseudoclassical Dirac brackets replacing Poisson brackets,[4–7,12]

$$\{\ ,\ \} = \frac{\overleftarrow{\partial}}{\partial x^\mu}\frac{\overrightarrow{\partial}}{\partial p_\mu} - \frac{\overleftarrow{\partial}}{\partial p^\mu}\frac{\overrightarrow{\partial}}{\partial x_\mu} + i\frac{\overleftarrow{\partial}}{\partial \theta^\mu}\frac{\overrightarrow{\partial}}{\partial \theta_\mu} + i\frac{\overleftarrow{\partial}}{\partial \theta_5}\frac{\overrightarrow{\partial}}{\partial \theta_5}, \tag{11}$$

the constraints $\mathscr{S}$ and $\mathscr{H}$ in (7) are consistent,

$$\{\mathscr{S},\mathscr{S}\} = i\mathscr{H} \approx 0,$$
$$\{\mathscr{S},\mathscr{H}\} = \frac{1}{i}\{\mathscr{S},\{\mathscr{S},\mathscr{S}\}\} = 0, \tag{12}$$

and are left (weakly) invariant by G. G has a strongly vanishing bracket with itself, and hence might be termed an abelian supersymmetry generator. Two supersymmetry transformations generated by G do not produce a reparametrization of the world–line, in contrast to those generated by $\mathscr{S}$.[4–7]

We start with this description of a free spinning particle and add interactions in such a way as to retain supersymmetry. An important consequence of the transformations (9) is that they leave invariant a special position variable,

$$\tilde{x} = x + i\theta\theta_5/m. \tag{13}$$

For the free particle, $\tilde{x}$ has linear τ development for arbitrary v and is a position variable with v-dependent *Zitterbewegung* subtracted out.[13] Thus, if we start with an initially supersymmetric Lagrangian (7) for each particle and insert an x dependence for each through $\tilde{x}$, we will obtain a supersymmetric Lagrangian that describes interaction.

To introduce scalar interactions, we first modify each mass m to a mass potential $m(x)$ just as we would for spinless particles. Then we replace x by its supersymmetric counterpart $\tilde{x}$ wherever x appears. This prescription leads to a self-referent definition of x in the interacting case

$$\tilde{x}_i = x_i + i\theta_i\theta_{5i}/\tilde{m}_i, \quad \tilde{m}_i = m_i(\{\tilde{x}_j\}). \tag{14}$$

Since the θ's belonging to a given particle anticommute,[14] (14) leads to a terminating Taylor-series expansion that determines $\tilde{x}$ in terms of θ_i, θ_{5i}, m, and m's ordinary derivatives with respect to the x_i. The resulting Lagrangians give constraints of the form

$$\mathscr{S}_j = p_j\cdot\theta_j + \tilde{m}_j\theta_{5j} \approx 0 \tag{15}$$

for each spinning particle, and

$$\mathscr{H}_j = p_j^2 + \tilde{m}_j^2 \approx 0 \tag{16}$$

for each particle (boson or fermion). For each fermion, $\{\mathscr{S}_j,\mathscr{S}_j\} = i\mathscr{H}_j \approx 0$ and $\{\mathscr{S}_j,\mathscr{H}_j\} \equiv 0$, so that the constraints for each particle are self-compatible.

Mutual (first-class) compatibility of interacting particles requires that $\{\mathscr{S}_i,\mathscr{S}_j\} \approx 0$, $\{\mathscr{S}_i,\mathscr{H}_j\} \approx 0$, and $\{\mathscr{H}_i,\mathscr{H}_j\} \approx 0$ for all $i \neq j$. The Jacobi condition for the pseudoclassical Dirac bracket

$$\sum_{\text{cyclic}} \eta_{\alpha\gamma}\{A_\alpha,\{A_\beta,A_\gamma\}\} \equiv 0, \tag{17}$$

where $\eta_{\alpha\gamma} = \Pi_i(-)^{\epsilon_{\alpha_i}\epsilon_{\gamma_i}}$, relates all constraint brackets to $\{\mathscr{S}_i,\mathscr{S}_j\}$ for two fermions, or $\{\mathscr{S}_i,\mathscr{H}_j\}$ for one fermion and one boson. Hence, if these brackets vanish strongly, then so do all the others. For the two-body case, this will lead to the same problems as that solved by Todorov for the interacting system of two spinless particles as in Eqs. (3)–(5).

For a supersymmetric spinning particle in scalar interaction with a spinless one, our constraints become

$$\mathscr{S}_1 = p_1\cdot\theta_1 + \tilde{m}_1\theta_{51} \approx 0, \quad \mathscr{H}_2 = p_2^2 + \tilde{m}_2^2 \approx 0,$$
$$\tilde{m}_i = \tilde{m}_i(\tilde{x}_\perp^2/2).^{15} \tag{18}$$

Here

$$\tilde{x}_\perp = (\tilde{x}_1 - x_2)_\perp = x_\perp + i\theta_{1\perp}\theta_{51}/\tilde{m}_1. \tag{19}$$

Using the Grassmann–Taylor expansion, we find

$$\tilde{m}_1 = m_1 + ix_\perp\cdot\theta_1\theta_{51}m_1'/m_1,$$
$$\tilde{m}_2 = m_2 + ix_\perp\cdot\theta_1\theta_{51}m_2'/m_1, \tag{20}$$

where the prime denotes derivative with respect to argument. Then the constraints take the form

$$\mathscr{S}_1 = p_1\theta_1 + \tilde{m}_1\theta_5 = p_1\theta_1 + m_1\theta_5, \tag{21a}$$
$$\mathscr{H}_1 = -i\{\mathscr{S}_1,\mathscr{S}_1\} = p_1{}^2 + \tilde{m}_1{}^2, \tag{21b}$$
$$\mathscr{H}_2 = p_2{}^2 + \tilde{m}_2{}^2, \tag{21c}$$

where

$$\tilde{m}_1{}^2 = m_1{}^2 + 2ix_\perp\cdot\theta_1\theta_{51}m_1',$$
$$\tilde{m}_2{}^2 = m_2{}^2 + 2ix_\perp\cdot\theta_1\theta_{51}m_2{}'m_2/m_1. \tag{22}$$

Application of the Jacobi identity (17) leads to

$$\{\mathscr{H}_2,\mathscr{H}_1\} = -i\{\mathscr{H}_2,\{\mathscr{S}_1,\mathscr{S}_1\}\}$$
$$= -2i\{\mathscr{S}_1,\{\mathscr{H}_2,\mathscr{S}_1\}\}.$$

But, if $m_1{}^2 - m_2{}^2 =$ constant, then

$$\{\mathscr{H}_2,\mathscr{S}_1\} = 2x_\perp\cdot(p_1 - p_2)(m_1^2 - m_2^2)'\theta_{15}/m_1 = 0. \tag{23}$$

So, the constraints $\mathscr{S}_1,\mathscr{H}_1,\mathscr{H}_2$, are all first-class. In the static limit $m_1 \rightarrow \infty$, m_2 finite or $m_2 \rightarrow \infty$, m_1 finite, (21a)–(21c) reduce, respectively, to the correct constraints for a spinless particle or for a supersymmetric spinning particle in an external scalar potential.

For two supersymmetric spinning particles in scalar interaction, the constraints are

$$\mathscr{S}_1 = p_1\cdot\theta_1 + \tilde{m}_1\theta_{51} \approx 0, \quad \mathscr{S}_2 = p_2\cdot\theta_2 + \tilde{m}_2\theta_{52} \approx 0. \tag{24}$$

Once again $\tilde{m} \equiv m(\tilde{x}_\perp^2/2)$,[15] where $\tilde{x}_\perp$ now becomes

$$\tilde{x}_\perp = (\tilde{x}_1 - \tilde{x}_2)_\perp = x_\perp + i\theta_{1\perp}\theta_{51}/\tilde{m}_1 - i\theta_{2\perp}\theta_{52}/\tilde{m}_2. \tag{25}$$

A Taylor expansion makes the spin content of $\tilde{x}_\perp$ explicit:

$$\tilde{x}_\perp = x_\perp + i\theta_{1\perp}\theta_{51}/m_1 - i\theta_{2\perp}\theta_{52}/m_2$$
$$- \theta_{1\perp}\theta_{51}x_\perp\cdot\theta_2\theta_{52}m_1'/m_1^2m_2$$
$$- \theta_{2\perp}\theta_{52}x_\perp\cdot\theta_1\theta_{51}m_2'/m_2{}^2m_1. \tag{26}$$

Thus, we find that

$$\begin{aligned}\tilde{m}_1 = m_1 + ix_\perp\cdot\theta_1\theta_{51}m_1'/m_1 - ix_\perp\cdot\theta_2\theta_{52}m_1'/m_2 \\ + \theta_{1\perp}\cdot(\theta_{2\perp}\theta_{51}\theta_{52}/m_1m_2)m_1' \\ + x_\perp\cdot\theta_1\theta_{51}x_\perp\cdot\theta_2\theta_{52}(m_1'/m_1m_2).\end{aligned} \tag{27}$$

Then the constraints become

$$\mathscr{S}_1 = p_1\cdot\theta_1 + m_1\theta_{51} - ix_\perp\cdot\theta_2\theta_{52}\theta_{51}m_1'/m_2 \approx 0, \tag{28a}$$

$$\mathscr{S}_2 = p_2\cdot\theta_2 + m_2\theta_{52} + ix_\perp\cdot\theta_1\theta_{51}\theta_{52}m_2'/m_1 \approx 0, \tag{28b}$$

$$\mathscr{H}_1 = \frac{1}{i}\{\mathscr{S}_1,\mathscr{S}_1\} = p_1^2 + \tilde{m}_1^2 \approx 0, \tag{28c}$$

$$\mathscr{H}_2 = \frac{1}{i}\{\mathscr{S}_2,\mathscr{S}_2\} = p_2^2 + \tilde{m}_2^2 \approx 0. \tag{28d}$$

Once again we find that $\{\mathscr{S}_1,\mathscr{S}_2\} = 0$ provided that $m_1^2 - m_2^2 =$ constant. Jacobi identities lead to the vanishing of all other brackets, so that all are compatible. For example, $\{\mathscr{H}_1,\mathscr{H}_2\} = -2i\{\mathscr{S}_1,\{\mathscr{H}_2,\mathscr{S}_1\}\} = -4\{\mathscr{S}_1,\{\mathscr{S}_2,\{\mathscr{S}_1,\mathscr{S}_2\}\}\}$. This time the two static limits each produce a single supersymmetric spinning particle in an external scalar potential. So, once spin complications have been disposed of by the supersymmetric structure of (28), compatibility problems reduce to those of the purely bosonic case. In fact, our system (28) might be regarded as the "square root" of Todorov's bosonic one.[1,5]

We quantize these systems by turning brackets into canonical commutators (anticommutators),[16]

$$\{\ ,\ \} \to \frac{1}{i\hbar}[\ ,\ ]_{\pm}\,. \tag{29}$$

The γ matrices emerge as the operator versions of the fermonic variables

$$\theta^\mu \to i(\hbar/2)^{1/2}\gamma_5\gamma^\mu, \quad \theta_5 \to i(\hbar/2)^{1/2}\gamma_5. \tag{30}$$

Then the constraints $\mathscr{S}_i$ (and $\mathscr{H}_i$) become a consistent set of coupled Dirac (and Klein–Gordon) equations. For example, in the spin-one-half–spin-one-half system $\mathscr{S}_1 \approx 0$ and $\mathscr{S}_2 \approx 0$ become

$$\mathscr{S}_1\psi = (\gamma_{51}\gamma_1\cdot p_1 + m_1\gamma_{51} - (i\hbar/2)x_\perp\cdot\gamma_2\gamma_{51}m_1'/m_2)\psi = 0 \tag{31}$$

and

$$\mathscr{S}_2\psi = (\gamma_{52}\gamma_2\cdot p_2 + m_2\gamma_{52} + (i\hbar/2)\,x_\perp\cdot\gamma_1\gamma_{52}m_2'/m_1)\psi = 0,$$

which reduce to ordinary Dirac equations in either static limit. The quantum consistency condition $[\mathscr{S}_1,\mathscr{S}_2]_-\,\psi = 0$ can be verified in direct analogy to the classical condition $\{\mathscr{S}_1,\mathscr{S}_2\} \approx 0$. Because of the equivalence in form of the Poisson bracket relation

$$\{A_\alpha A_\beta, A_\gamma\} = A_\alpha\{A_\beta, A_\gamma\} + \eta_{\beta\gamma}\{A_\alpha, A_\beta\}A_\gamma \tag{32}$$

and the quantum commutation (anticommutation) relation

$$\begin{aligned}[A_\alpha A_\beta, A_\gamma]_{-\eta_{\alpha\gamma}\eta_{\beta\gamma}} = A_\alpha[A_\beta, A_\gamma]_{-\eta_{\beta\gamma}} \\ + \eta_{\beta\gamma}[A_\alpha, A_\gamma]_{-\eta_{\alpha\gamma}}A_\beta,\end{aligned} \tag{33}$$

the verifications of classical and quantum consistency are virtually identical.[17]

To summarize, we have extended two-body relativistic constraint mechanics and its quantization to include constituent spin. The introduction of supersymmetry is crucial for eliminating spin complications. This is achieved through the replacement of the relative coordinate x with an $\tilde{x}$ variable that induces a particular spin dependence for direct scalar interactions whenever fermions are present.

We can extend our approach to include interactions other than the scalar. The resulting wave equations may lead to realistic spectra for confined systems of spinning quarks.

We wish to thank Professor Ingram Bloch for useful discussions on various aspects of this work.

[1] I. T. Todorov, JINR Report #E2-10125, Dubna (1976); V. V. Molotkov and I. T. Todorov, JINR Report #E2-12270, Dubna (1979). Closely related constraint approaches have been developed by M. Kalb and P. Van Alstine, Yale Reports, C00-3075-146 (1976), C00-3075-156 (1976); P. Van Alstine, Ph.D. Dissertation, Yale University 1976; Ph. Droz-Vincent, Phys. Rev. D **19**, 702 (1979); A. Komar, Phys. Rev. D **18**, 1881, 1887, 3617 (1978); F. Rohrlich, Ann. Phys. (New York) **117**, 292 (1979); M. King and F. Rohrlich, Phys. Rev. Lett. **44**, 621, (1980); T. Takabayasi, Prog. Theor. Phys. **54**, 1235 (1979); K. Rafanelli, S. Orenstein, and V. M. Penafiel N. "Relativistic Dynamics of N Interacting Particles," Queens College Preprints (1980).

[2] H. W. Crater and P. Van Alstine, Phys. Lett. B **100**, 166 (1981).

[3] J. Richardson, Phys. Lett. B **82**, 272 (1979).

[4] R. Casalbuoni, Phys. Lett. B **62**, 49 (1976); R. Casalbuoni, Nuovo Cimento A **33**, 389 (1975); F. A. Berezin and M. S. Marinov, JETP Lett. **21**, 678 (1975); Ann. Phys. (New York) **104**, 336 (1977); A. Barducci, R. Casalbuoni, and L. Lusanna, Nuovo Cimento A **32**, 377 (1976); F. Ravndal, Phys. Rev. D **21**, 2823 (1980).

[5] C. A. P. Galvao and C. Teitelboim, J. Math. Phys. **21**, 1863 (1980); C. Teitelboim, Phys. Rev. Lett. **38**, 1106 (1977).

[6] D. Dominici and G. Longhi, Nuovo Cimento A **46**, 213 (1978).

[7] P. Fayet and S. Ferrara, Phys. Rep. **32**, 249 (1977); L. Brink, P. DiVecchia, and P. Howe, Nucl. Phys. B **118**, 76 (1977).

[8] A. Neveu and J. Schwartz, Nucl. Phys. B **31**, 86 (1971); P. Ramond, Phys. Rev. D **3**, 2415 (1971); J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974).

[9] I. T. Todorov (Ref. 1) includes momentum dependence by taking $m_i = m_i(x_\perp^2/2, x_\perp\cdot p, p^2, P^2)$.

[10] To obtain invariance of the action under local (but not global) transformations we need to use the $\mathscr{S}$ constraint. Similar effects have been discussed by Galvao and Teitelboim[5] for a different particle supersymmetry and by F. Deser and P. K. Townsend, Phys. Lett. B **98**, 188 (1981) in supergravity.

[11] Note that this transformation leaves the three pieces of the action separately invariant for solutions of the equations of motion. Furthermore, while mixing fermionic and bosonic variables in the usual way, it leaves invariant $(-\dot{x}^2)^{1/2}$, the magnitude of the particles' four-velocity. This property of (9) allows us to separate problems associated with the introduction of spin from those that would arise even in a purely bosonic system.

[12] P. A. M. Dirac, *Lectures in Quantum Mechanics*, Belfer Graduate School of Science (Yeshiva Univ., New York, 1964).

[13] An $\tilde{x}$-like variable has been used for spinning systems without Grassmann numbers by M. H. L. Pryce, Proc. R. Soc. (London) **195**, 62, (1948); F. Halbwachs, Nuovo Cimento **176**, 832 (1964), and K. Rafanelli, Can. J. Phys. **50**, 2489 (1972). Other $\tilde{x}$'s have been used with Grassmann variables by Barducci, Casalbuoni, and Lusanna, Ref. 4, and by F. Ravndal, Ref. 4 [the latter's version with an independent Grassmann variable θ replacing $\theta_5(\tau)$].

[14] Although θ's for a given particle anticommute among themselves they commute with the θ's belonging to another particle. That is, each particle has an independent Grassmann space. All the dynamical variables we deal with have definite even or odd character with respect to each space. The Jacobi identity (17) holds for such objects.

[15] We could include momentum dependence as Todorov does for m_i by taking $\tilde{m}_i = m_i(\tilde{x}_\perp{}^2/2, \tilde{x}_\perp\cdot p, p^2, P^2)$. He also includes spin dependence but without Grassmann variables and supersymmetry. More recent work by Todorov presented in lectures at the Scuola Internazionale Superiore di Studi Avanzati and at the ICTP in Trieste treats mass and spin shell constraints of a type different from ours and includes second as well as first class constraints.

[16] Quantizations of related one-body systems have been given by Ravndal, Ref. 4, and also by J. W. F. Valle, Int. J. Theor. Phys. **18**, 923 (1979).

[17] The classical and quantum consistency conditions are algebraically identical if the squares of the θ's for each particle are set equal to some common nonzero value in analogy with the Dirac matrices of the quantum case.

Coulomb-modified nuclear scattering. I

B. Talukdar and D. K. Ghosh
Department of Physics, Visva-Bharati University, Santiniketan, West Bengal, India
T. Sasakawa
Department of Physics, Tohoku University, Sendai 980, Japan

(Received 15 October 1981; accepted for publication 2 April 1982)

We contemplate deriving a wavefunction approach to Coulomb-distorted nuclear scattering. The theory of ordinary differential equations supplemented by certain well-known properties of higher transcendental functions has been found adequate for the purpose if the nuclear potential is a nonlocal separable one with exponential form factors. The method presented will work for potentials of arbitrary rank. We have derived specific results for Jost function and Fredholm determinants for scattering by (i) Coulomb plus Yamaguchi and (ii) Coulomb plus Mongan case IV potentials.

PACS numbers: 24.10. – i, 24.90. + d

I. INTRODUCTION

Experiments which involve scattering by additive interactions are analyzed by the use of the Gell-Mann–Goldberger scattering-by-2-potential theorem[1] (GG theorem). Applicability of the GG theorem is directly related to the existence and/or completeness of the wave operators for the scattering system.[2] The wave operators exist under strong limits when each of the associated interactions is of short range, but they do not exist in the presence of a Coulomb force. To deal with long-range interactions, the wave operators are judiciously modified by relaxing some requirements. Recently, the situation with regard to this has been nicely summarized by Chandler.[3] On several occasions van Haeringen,[4] van Haeringen and van Wageningen,[5] and Kok and van Haeringen[6] have used the GG theorem based on modified wave operators to derive the basic statement of the scattering theory for Coulomb-modified nuclear potentials.

The purpose of the present paper is to develop a mathematical framework for the Coulomb–nuclear problem which does not make explicit use of the GG theorem. For our development we shall use only the theory of ordinary differential equations together with certain properties of the higher transcendental functions. We shall see in the course of our study that the merit of the present approach is its simplicity. For the nuclear part of the interaction we use nonlocal separable potentials. This is justified by the observation that the short-range local potentials can be approximated by finite-rank separable potentials[7] and also that the nonlocal potentials can describe a much wider variety of phenomena than that encompassed with short-range local potential.[8] The method proposed will be applicable for Coulomb plus separable potentials of arbitrary rank. The plan of the present paper is as follows.

In Sec. II we briefly describe the conventional method of treating the separable nuclear interaction with emphasis on the Yamaguchi potential[9] and judiciously modify the approach to develop a mathematical framework which is adequate for dealing with the Coulomb–nuclear problem. In Sec. III we present results for the Jost function and associate Fredholm determinants for the Coulomb plus Yamaguchi potential. Similar results are presented in Sec. IV for the Coulomb distorted Mongan case IV potential.[10] For simplicity of presentation we consider the s-wave case only, with the subscript $l = 0$ omitted, and work in units in which $\hbar^2/2m$ is unity. The higher partial-wave generalization of our results is really trivial. We present some concluding remarks in Sec. V.

II. WAVEFUNCTIONS FOR NONLOCAL SEPARABLE POTENTIALS

Yamaguchi[9] has introduced a one-term separable potential to describe the nucleon–nucleon scattering. We begin by describing a method to calculate the wavefunctions for this potential, which is a little unconventional compared to what exists in the literature. The Schrödinger equation for the Yamaguchi potential can be written in the form

$$\left(\frac{d^2}{dr^2} + k^2\right)\psi(k,r) = d(k)e^{-\alpha r}, \tag{1}$$

where

$$d(k) = \lambda \int_0^\infty e^{-\alpha s}\psi(k,s)ds, \tag{2}$$

with λ and α, the strength and range parameters of the potential. We shall solve (1) by treating the integral in (2) as a constant. The unknown constant which appears will be determined by substituting the solution back in the defining equation for $d(k)$ and matching the desired boundary conditions. For example, if we are interested in the Jost solution[11] $f(k,r)$, we write from (1)

$$\psi(k,r) = f(k,r) = e^{ikr} + [d(k)/(\alpha^2 + k^2)]e^{-\alpha r}. \tag{3}$$

Using (3) in (2), we get

$$d(k) = \lambda/(\alpha - ik)D(k), \tag{4}$$

where the Fredholm determinant

$$D(k) = 1 - \lambda/2\alpha(\alpha^2 + k^2). \tag{5}$$

Substituting (4) in (3), we get

$$f(k,r) = e^{ikr} + [\lambda/(\alpha - ik)(\alpha^2 + k^2)D(k)]e^{-\alpha r}. \tag{6}$$

Clearly, $f(k,r)$ in (6) satisfies the Jost boundary condition since

$$f(k,r) \underset{r\to\infty}{\sim} e^{ikr}. \tag{7}$$

From (6) the Jost function is

$$f(k)(=f(k,0))=D^{+}(k)/D(k), \tag{8}$$

where

$$D^{+}(k)=D(k)+\lambda(\alpha+ik)/(\alpha^2+k^2)^2 \tag{9}$$

is the Fredholm determinant associated with the physical (outgoing wave) solution. The method outlined above has been used by one of us[12] to construct analytic expression for off-shell Jost and physical solutions for realistic separable nucleon–nucleon interactions. We indicate below how this method should be suitably modified to treat the Coulomb–nuclear problem.

We change the dependent and independent variables in (1) by substituting

$$\psi(k,r)=\phi(k,r)=re^{ikr}g(r), \qquad r=-z/2ik \tag{10}$$

and get

$$z\frac{d^2g(z)}{dz^2}+(c-z)\frac{dg(z)}{dz}-ag(z)=-\frac{d(k)}{2ik}e^{\rho z}, \tag{11}$$

where

$$a=1, \quad c=2, \quad \text{and} \quad \rho=(\alpha+ik)/2ik. \tag{12}$$

Complementary functions of (11) are given by the confluent hypergeometric functions

$$\Phi(a,c;z)=\frac{\Gamma(c)}{\Gamma(a)}\sum_{n=0}^{\infty}\frac{\Gamma(a+n)z^n}{\Gamma(c+n)n!} \tag{13}$$

and

$$\overline{\Phi}(a,c;z)=z^{1-c}\Phi(a-c+1,2-c;z). \tag{14}$$

Note that,for $c=2$, (14) is not an acceptable solution of (1). However, $\overline{\Phi}$ tends towards the solution[13] of (1) when c approaches 2. In our subsequent discussions we shall always mean that limit. This is no loss of generalization. See, for example, the treatment of Coulomb field by Newton.[14] Another solution of (1) defined within the framework of the same limiting procedure is

$$\Psi(a,c;z)=\frac{\Gamma(1-c)}{\Gamma(a-c+1)}\Phi(a,c;z)+\frac{\Gamma(c-1)}{\Gamma(a)}\overline{\Phi}(a,c;z). \tag{15}$$

Given Φ and $\overline{\Phi}$, we find a particular solution of (1) by the method of variation of parameters.[15] Thus we have

$$g_p(z)=-\frac{d(k)}{2ik}\left[-\Phi(a,c;z)\int\frac{\overline{\Phi}(a,c;z)}{zW}e^{\rho z}dz+\overline{\Phi}(a,c;z)\int\frac{\Phi(a,c;z)}{zW}e^{\rho z}dz\right], \tag{16}$$

where the Wronskian

$$W=W(\Phi,\overline{\Phi})=-(c-1)z^{-c}e^{z}. \tag{17}$$

The integrals in (16) can be performed by expanding $e^{\rho z}$ in powers of z and making use of the integrals[16]

$$\begin{aligned}\int e^{-z}z^{\sigma+c-2}\Phi(a,c;z)dz&=z^ce^{-z}[(\sigma-1)\Phi(a,c;z)\\&\quad\times\theta_{\sigma-1}(a+1,c+1;z)\\&\quad-(a/c)\theta_\sigma(a,c;z)\Phi(a+1,c+1;z)]\end{aligned} \tag{18}$$

and

$$\begin{aligned}\int e^{-z}z^{\sigma+c-2}\overline{\Phi}(a,c;z)dz&=z^ce^{-z}[(\sigma-1)\overline{\Phi}(a,c;z)\\&\quad\times\theta_{\sigma-1}(a+1,c+1;z)\\&\quad+(c-1)\theta_\sigma(a,c;z)\overline{\Phi}(a+1,c+1;z)],\end{aligned} \tag{19}$$

where

$$\begin{aligned}\theta_\sigma(a,c;z)&=z^\sigma\sum_{n=0}^{\infty}\frac{\Gamma(\sigma+a+n)\Gamma(\sigma)\Gamma(\sigma+c-1)}{\Gamma(\sigma+a)\Gamma(\sigma+n+1)\Gamma(\sigma+c+n)}z^n\\&=\frac{z^\sigma}{\sigma(\sigma+c-1)}\,{}_2F_2(1,\sigma+a;\sigma+1,\sigma+c;z).\end{aligned} \quad (20$$

Finally we obtain

$$g_p(z)=-\frac{d(k)}{2ik}\sum_{n=0}^{\infty}\frac{\theta_{n+1}(1,2;z)}{n!}\rho^n. \tag{21}$$

To write (21), we have made use of the Wronskian relation (17). Combining (10), (13), and (21), the solution regular at the origin will be given by

$$\phi(k,r)=re^{ikr}\Phi(1,2;-2ikr)-re^{ikr}\frac{d(k)}{2ik}\times\sum_{n=1}^{\infty}\frac{\theta_n(1,2;-2ikr)}{(n-1)!}\rho^{n-1}. \tag{22}$$

Substituting (22) in (2), we have

$$d(k)=\frac{\lambda/(\alpha^2+k^2)}{1+\lambda\,[(\alpha^2+k^2)(\alpha-ik)]^{-1}\Sigma_{n=1}^{\infty}(-1)^n[(\alpha+ik)/(\alpha-ik)]^n{}_1F_0(1;-2ik/(\alpha-ik))}. \tag{23}$$

In deriving (23) we have used the following integrals:

$$\int_0^\infty e^{-\lambda z}z^\nu\Phi(a,c;pz)=[\Gamma(\nu+1)/\lambda^{\nu+1}]\,{}_2F_1(a,\nu+1;c;p/\lambda) \tag{24}$$

and

$$\int_0^\infty e^{-\lambda z}z^\nu\theta_\sigma(\alpha,c;pz)dz=\frac{\Gamma(\nu+\sigma-1)}{\sigma(\sigma+c+1)}\frac{p^\sigma}{\lambda^{\nu+\sigma+1}}\times{}_3F_2(1,\sigma+\alpha,\nu+\sigma+1;\sigma+1,\sigma+c;p/\lambda) \tag{25}$$

together with the reduction formula[17]

$$ {}_pF_q(\alpha_1,\beta_1,\gamma_1,\cdots;\alpha_2,\beta_1,\gamma_2,\cdots;z) $$
$$ = {}_{p-1}F_{q-1}(\alpha_1,\gamma_1,\cdots;\alpha_2,\gamma_2,\cdots;z) \tag{26} $$

for the generalized hypergeometric function ${}_pF_q$. The integral in (24) has been given by Landau and Lifshitz[18] while that in (25) can be proved by expanding $\theta_\sigma(\alpha,c;pz)$ as a power series in z and integrating term by term.

Since ${}_1F_0(1;-2ik/(\alpha-ik)) = [(\alpha-ik)/(\alpha+ik)]$, it can easily be shown that the denominator in (23) is $D(k)$ given in (5). This result is quite expected because the Fredholm determinants for the regular and Jost solutions are equal for a symmetric nonlocal potential.[19] Thus we write (22) in the form

$$ \phi(k,r) = re^{ikr}\Phi(1,2;-2ikr) $$
$$ - re^{ikr}\frac{\lambda}{(\alpha^2+k^2)D(k)} $$
$$ \times\frac{1}{2ik}\sum_{n=1}^{\infty}\frac{\theta_n(1,2;-2ikr)}{(n-1)!}\rho^{n-1}. \tag{27} $$

The integral representation of the Jost function $f(k)$ in terms of the regular solution $\phi(k,r)$ is given by

$$ f(k) = 1+\lambda\int_0^\infty e^{-\alpha s}e^{iks}ds\int_0^\infty e^{-\alpha r}\phi(k,r)dr. \tag{8$'$} $$

Applying the results in (24) to (26), it is easy to see that $f(k)$ in Eq. (8′) is in exact agreement with that given by Eq. (8) obtained earlier by a rather simple technique. However, we note that only a relatively complicated formulation of the problem outlined above can be extended to treat the Coulomb–nuclear interaction.

III. JOST FUNCTION FOR COULOMB PLUS YAMAGUCHI POTENTIAL

The radial Schrödinger equation for the Coulomb plus Yamaguchi potential is given by

$$ \left(\frac{d^2}{dr^2}+k^2-\frac{2\eta k}{r}\right)\phi(k,r) = d(k)e^{-\alpha r} \tag{28} $$

with

$$ d(k) = \lambda\int_0^\infty e^{-\alpha s}\phi(k,s)ds. \tag{29} $$

Here η is the well-known Coulomb parameter

$$ \eta = z_1z_2e^2/v. \tag{30} $$

The transformations in (10) reduce (28) in the form

$$ z\frac{d^2g(z)}{dz^2}+(2-z)\frac{d(z)}{dz}-(1+i\eta)g(z) = -\frac{d(k)}{2ik}e^{\rho z}, \tag{31} $$

where

$$ \rho = (\alpha+ik)/2ik. \tag{32} $$

The complementary functions of (31) are

$$ g_1(z) = \Phi(1+i\eta,2;z) \tag{33a} $$

and

$$ g_2(z) = \overline{\Phi}(1+i\eta,2;z). \tag{33b} $$

As in (16), the particular integral of (28) is given by

$$ g_p(z) = -\frac{d(k)}{2ik}\left[-g_1(z)\int\frac{g_2(z)}{zW}e^{\rho z}dz\right. $$
$$ \left.+g_2(z)\int\frac{g_1(z)}{zW}e^{\rho z}dz\right]. \tag{34} $$

Following the procedure outlined for the pure Yamaguchi potential, $g_p(z)$ can be written in the closed form

$$ g_p(z) = -\frac{d(k)}{2ik}\sum_{n=0}^{\infty}\frac{\theta_{n+1}(1+i\eta,2;z)}{n!}\rho^n. \tag{35} $$

In terms of (35) the regular solution for (28) is obtained as

$$ \phi(k,r) = re^{ikr}\Phi(1+i\eta,2;-2ikr) $$
$$ -\frac{d(k)}{2ik}re^{ikr} $$
$$ \times\sum_{n=1}^{\infty}\frac{\theta_n(1+i\eta,2;-2ikr)}{(n-1)!}\rho^{n-1}. \tag{36} $$

The unknown constant $d(k)$ is obtained by substituting $\phi(k,r)$ from (36) in (29). Thus we have

$$ d(k) = \lambda e^{2\eta y}/(\alpha^2+k^2)D(k), \tag{37} $$

where the Fredholm determinant $D(k)$ associated with the regular solution is given by

$$ D(k) = 1+\frac{\lambda}{(\alpha^2+k^2)(\alpha-ik)}\sum_{n=1}^{\infty}(-1)^n\left(\frac{\alpha+ik}{\alpha-ik}\right)^n $$
$$ \times{}_2F_1\left(1,1+n+i\eta;1+n;-\frac{2ik}{\alpha-ik}\right). \tag{38} $$

Here

$$ y = \tan^{-1}k/\alpha. \tag{39} $$

Combining (36) and (37), we get

$$ \phi(k,r) = re^{ikr}\left[\Phi(1+i\eta,2;-2ikr)\right. $$
$$ -\frac{\lambda e^{2\eta y}}{2ik(\alpha^2+k^2)D(k)} $$
$$ \left.\times\sum_{m=1}^{\infty}\theta_m(1+i\eta,2;-2ikr)\frac{\rho^{m-1}}{(m-1)!}\right]. \tag{40} $$

In Appendix A we present a Laplace transform method to solve (31) and arrive at (40). This method appears to be simpler than the approach outlined above.

The appropriate Jost function $f(k)$ can be obtained from the following integral representation[14,20,21]

$$ f(k) = f_c(k)+\lambda\int_0^\infty e^{-\alpha s}f_c(k,s) $$
$$ \times ds\int_0^\infty e^{-\alpha r}\phi(k,r)\,dr, \tag{41} $$

where the Coulomb Jost solution and Jost function are

$$ f_c(k,r) = (-2ik)e^{\pi\eta/2}re^{ikr}\Psi(1+i\eta,2;-2ikr) \tag{42} $$

and

$$ f_c(k) = \frac{e^{\pi\eta/2}}{\Gamma(1+i\eta)}, \tag{43} $$

respectively. Using (40), (42), and (43) in (41), we have

$$f(k)=f_c(k)\left[1-2ik\lambda\,\frac{\Gamma(1+i\eta)e^{2\eta y}}{(\alpha^2+k^2)D(k)}\,I\right]$$

$$=\frac{D^+(k)}{D(k)},\tag{44}$$

where

$$D^+(k)=D(k)\frac{e^{\pi\eta/2}}{\Gamma(1+i\eta)}+\lambda e^{2\eta y}\frac{e^{\pi\eta/2}}{\Gamma(1+i\eta)(\alpha^2+k^2)^2}$$

$$\times[R(\alpha,y)+\beta(y)]\tag{45}$$

with

$$R(\alpha,y)=\alpha+2\eta k e^{2\eta y}\{-i\pi/2+i/2\eta$$

$$+\psi(1+i\eta)-\psi(1)$$

$$+\ln(2k/y)-\tfrac{1}{2}\ln[1+(k/y)^2]\}\tag{46}$$

and

$$\beta(y)=2\eta k e^{2\eta y}\sum_{p=1}^{\infty}\frac{(-2\eta y)^p}{p!}$$

$$\times\sum_{l=0}^{\infty}(-1)^l\frac{2^{2l}B_{2l}y^{2l}}{(2l)!(2l+p)}.\tag{47}$$

In the above B_{2l} and ψ stand for the Bernoulli numbers and logarithmic derivative of the gamma function, respectively. To write (44), we have used the result of the integral

$$I=\int_0^\infty s e^{-\alpha s}e^{iks}\Psi(1+i\eta,2;-2iks)\,ds\tag{48}$$

in addition to those in (24) and (25). The method for the evaluation of I has been shown in Appendix B.

IV. JOST FUNCTION FOR COULOMB PLUS A RANK-2 SEPARABLE POTENTIAL

A rank-2 separable potential has been introduced by Mongan[10] in fitting the ${}^1s_0$ nucleon–nucleon phase shifts. The Schrödinger equation for the Coulomb plus Mongan case IV potential can be written in the form

$$\left(\frac{d^2}{dr^2}+k^2-\frac{2\eta k}{r}\right)\phi(k,r)$$

$$=d_1(k)e^{-\alpha_1 r}+d_2(k)e^{-\alpha_2 r}\tag{49}$$

with

$$d_1(k)=\lambda_1\int_0^\infty e^{-\alpha_1 s}\phi(k,s)\,ds\tag{50a}$$

and

$$d_2(k)=\lambda_2\int_0^\infty e^{-\alpha_2 s}\phi(k,s)\,ds.\tag{50b}$$

The regular solution of (49) is obtained as

$$\phi(k,r)=re^{ikr}\Phi(1+i\eta,2;-2ikr)$$

$$-\frac{1}{2ik}re^{ikr}$$

$$\times\sum_{n=1}^{\infty}\frac{\theta_n(1+i\eta,2;-2ikr)}{(n-1)!}$$

$$\times[d_1(k)\rho_1^{n-1}+d_2(k)\rho_2^{n-1}].\tag{51}$$

The unknown constants $d_1(k)$ and $d_2(k)$ are obtained by substituting $\phi(k,r)$ in (50a) and (50b) and solving the resulting simultaneous equations. We have

$$d_i(k)=\{(\alpha_j^2+k^2)+\lambda_j[(\alpha_j^2+k^2)X_n(\alpha_j)$$

$$-(\alpha_i^2+k^2)Y_n(\alpha_i,\alpha_j)]\}$$

$$\times\frac{\lambda_i e^{2\eta y_i}}{D(k)(\alpha_i^2+k^2)(\alpha_j^2+k^2)},\tag{52}$$

$$i=1,2,\quad j=1,2,\quad i\neq j,$$

where the Fredholm determinant $D(k)$ associated with the regular solution is given by

$$D(k)=1+\lambda_1X_n(\alpha_1)+\lambda_2X_n(\alpha_2)$$

$$+\lambda_1\lambda_2[X_n(\alpha_1)X_n(\alpha_2)$$

$$-Y_n(\alpha_1,\alpha_2)Y_n(\alpha_2,\alpha_1)].\tag{53}$$

In Eqs. (52) and (53)

$$X_n(\alpha_i)=\frac{1}{(\alpha_i-ik)(\alpha_i^2+k^2)}\sum_{n=1}^{\infty}(-1)^n\left(\frac{\alpha_i+ik}{\alpha_i-ik}\right)^n$$

$$\times{}_2F_1\left(1,1+n+i\eta;1+n;\frac{-2ik}{\alpha_i-ik}\right),\tag{54a}$$

$$Y_n(\alpha_i,\alpha_j)=\frac{1}{(\alpha_j+ik)(\alpha_i-ik)^2}\sum_{n=1}^{\infty}(-1)^n\left(\frac{\alpha_j+ik}{\alpha_i-ik}\right)^n$$

$$\times{}_2F_1\left(1,1+n+i\eta;1+n;\frac{-2ik}{\alpha_i-ik}\right),\tag{54b}$$

$$i=1,2,\quad j=1,2,\quad i\neq j.$$

The appropriate Jost function $f(k)$ can be obtained from the integral representation

$$f(k)=f_c(k)+\lambda_1\int_0^\infty e^{-\alpha_1 s}f_c(k,s)\,ds$$

$$\times\int_0^\infty e^{-\alpha_1 r}\phi(k,r)\,dr$$

$$+\lambda_2\int_0^\infty e^{-\alpha_2 s}f_c(k,s)\,ds$$

$$\times\int_0^\infty e^{-\alpha_2 r}\phi(k,r)\,dr.\tag{55}$$

The Jost function comes out to be

$$f(k)=D^+(k)/D(k),\tag{56}$$

where

$$D^+(k)=D(k)\frac{e^{\pi\eta/2}}{\Gamma(1+i\eta)}+\frac{e^{\pi\eta/2}}{\Gamma(1+i\eta)}$$

$$\times\sum_{i=1}^{2}\lambda_i[R(\alpha_i,y_i)+\beta(y_i)]$$

$$\times\left\{e^{2\eta y_i}D(k)-\frac{1}{(\alpha_j^2+k^2)}\right.$$

$$\left.\times[M_n(\alpha_i,\alpha_j)X_m(\alpha_i)+M_n(\alpha_j,\alpha_i)Y_m(\alpha_i,\alpha_j)]\right\}\tag{57}$$

with

$$M_n(\alpha_i,\alpha_j) = \lambda_i e^{2\eta y_i}\{(\alpha_j^2 + k^2) + \lambda_j[(\alpha_j^2 + k^2)X_n(\alpha_j) - (\alpha_i^2 + k^2)Y_n(\alpha_i,\alpha_j)]\}, \tag{58}$$

$$R(\alpha_i, y_i) = \alpha_i + 2\eta k e^{2\eta y_i} \times \{-i\pi/2 + i/2\eta + \psi(1 + i\eta) - \psi(1) + \ln(2k/y_i) - \tfrac{1}{2}\ln[1 + (k/y_i)^2]\}, \tag{59}$$

$$\beta(y_i) = 2\eta k e^{2\eta y_i} \sum_{p=1}^{\infty} \frac{(-2\eta y_i)^p}{p!} \times \sum_{l=0}^{\infty} (-1)^l \frac{2^{2l} B_{2l} y_i^{2l}}{(2l)!(2l + p)}, \tag{60}$$

and

$$y_i = \tan^{-1} k/\alpha_i, \quad i = 1,2, \quad j = 1,2, \quad i \neq j. \tag{61}$$

The quantity $D^+(k)$ in Eq. (56) is the Fredholm determinant associated with the physical solution. A useful check on the fairly complicated formulas for $D(k)$ and $D^+(k)$ consists in showing that when the Coulomb field is turned off by allowing $\eta \to 0$, the well-known results for Mongan case IV potential are reproduced. It is about trivial to see that the appropriate results of Mulligan *et al.*[8] are obtained from our expressions for $D(k)$ and $D^+(k)$ as $\eta \to 0$.

V. CONCLUDING REMARKS

In this work we have proposed a wavefunction method to calculate Jost functions and other appropriate Fredholm determinants for scattering by a Coulomb-modified nuclear separable potential. The method proposed depends only on the theory of ordinary differential equations and is quite general. The specific results presented refer to scattering by (i) Coulomb plus Yamaguchi and (ii) Coulomb plus Mongan case IV potentials. To our belief, most of these results have not appeared before. Some applications of our results will be the following:

(i) The Jost function $f(k)$ in Eq. (56) can be used in $\cot\delta = [f(k) + f(-k)]/i[f(k) - f(-k)]$ to compute the scattering length and effective range for the Coulomb-modified potential considered in this paper. We have already checked that the results of van Haeringen[4] are reproduced for $\lambda_2 = 0$.

(ii) The Mongan case IV potential can support spurious states and bound states embedded in the continuum for some selected values of the parameters. As discussed by Mulligan *et al.*,[8] these states are analyzed in terms of zeros of the Fredholm determinants associated with the regular and physical (outgoing wave) solutions of the Schrödinger equation. Our expressions for $D(k)$ and $D^+(k)$ may be used to examine the effect of the Coulomb potential on these states.

APPENDIX A

This appendix derives a Laplace transform method for solving the inhomogeneous differential equation

$$z\frac{d^2g(z)}{dz^2} + (c - z)\frac{dg(z)}{dz} - ag(z) = -\frac{d(k)}{2ik}e^{\rho z}. \tag{A1}$$

Since confluent hypergeometric functions are of exponential order, and the right-hand side of (A1) also is exponential, the Laplace transform method is expected to serve as one of the best techniques to solve this. If Re $c < 2$, both parts of the complementary functions have transforms, if $c \geqslant 2$ only one part has. Taking the Laplace transform of (A1), we get

$$\frac{d}{ds}[s(1 - s)\bar{g}(s)] + (cs - a)\bar{g}(s) = (c - 1)g(0) - \frac{d(k)}{2ik}\frac{1}{s - \rho}, \tag{A2}$$

where $\bar{g}(s) = \mathscr{L}\{g(z)\}$. This is a first-order differential equation in $\bar{g}(s)$ and can easily be written in the form

$$\frac{d}{ds}[(s - 1)^{1 + a - c}s^{1 - a}\bar{g}(s)] = -\frac{(c - 1)g(0)}{s^a(s - 1)^{c - a}} + \frac{d(k)}{2ik} \times \frac{1}{s^a(s - 1)^{c - a}(s - \rho)}. \tag{A3}$$

Integrating (A3) between the limits s to ∞ we write

$$\bar{g}(s) = s^{a - 1}(s - 1)^{c - a - 1}\left[A + (c - 1)g(0)\int_s^\infty \frac{d\omega}{\omega^a(\omega - 1)^{c - a}} - \frac{d(k)}{2ik}\int_s^\infty \frac{d\omega}{\omega^a(\omega - 1)^{c - a}(\omega - \rho)}\right], \tag{A4}$$

where A is a constant. The first two terms on right-hand side of (A3) gives the complementary functions of (A1) while the last term gives the particular integral. This can be seen as follows.

Consider the standard integral given in (24). For $\nu = 0$ this reads

$$\mathscr{L}\{\Phi(a,c;z)\} = (1/s)_2F_1(a,1;c;1/s). \tag{A5}$$

The Euler representation for the Gaussian hypergeometric functions ${}_2F_1(\alpha,\beta;\gamma;r)$ is

$${}_2F_1(\alpha,\beta;\gamma;r) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma - \beta)} \times \int_0^1 t^{\beta - 1}(1 - t)^{\gamma - \beta - 1}(1 - tr)^{-\alpha}\,dt. \tag{A6}$$

Using (A6) in (A5), we get

$$\mathscr{L}\{\Phi(a,c;z)\} = s^{a - 1}(c - 1)\int_0^1 (1 - t)^{c - 2}(s - t)^{-a}\,dt. \tag{A7}$$

The transformation

$$\omega = (s - t)/(1 - t) \tag{A8}$$

reduces (A7) to the form

$$\mathscr{L}\{\Phi(a,c;z)\} = (c - 1)s^{a - 1}(s - 1)^{c - a - 1} \times \int_s^\infty \frac{d\omega}{\omega^a(\omega - 1)^{c - a}}. \tag{A9}$$

Thus

$$(c-1)\mathscr{L}^{-1}\left\{s^{a-1}(s-1)^{c-a-1}\int_s^\infty \frac{d\omega}{\omega^a(\omega-1)^{c-a}}\right\}$$
$$=\Phi(a,c;z). \tag{A10}$$

Similarly

$$\mathscr{L}^{-1}\{s^{a-1}(s-1)^{c-a-1}\}=\frac{1}{\Gamma(2-c)}\overline{\Phi}(a,c;z). \tag{A11}$$

To deal with the third term in (A4), we restrict ourselves to the half-plane Re $s>$ Re ρ and Re $s>1$. Thus

$$\int_s^\infty \frac{d\omega}{\omega^a(\omega-1)^{c-1}(\omega-\rho)}=\sum_{n=0}^\infty \rho^n\int_s^\infty \frac{d\omega}{\omega^{a+n+1}(\omega-1)^{c-a}}. \tag{A12}$$

As we shall see presently, the above justifies the validity of expansions in (27) or (36).

Allowing $a\rightarrow a+n+1$ and $c\rightarrow c+n+1$, we get from (A9)

$$(c+n)s^{a+n}(s-1)^{c-a-1}\int_s^\infty d\omega/\omega^{a+n+1}(\omega-1)^{c-a}$$
$$=\mathscr{L}\{\Phi(a+n+1,c+n+1;z)\}$$
$$=(1/s)_2F_1(1,a+n+1;c+n+1;1/s). \tag{A13}$$

Using the expansion

$$\frac{1}{s}\,{}_2F_1\left(1,a+n+1;c+n+1;\frac{1}{s}\right)$$
$$=\frac{\Gamma(c+n+1)}{\Gamma(a+n+1)}\sum_{m=0}^\infty\frac{\Gamma(a+n+m+1)}{\Gamma(c+n+m+1)}$$
$$\times\frac{\Gamma(m+1)}{m!}s^{-m-1}, \tag{A14}$$

we write (A13) as

$$(c+n)s^{a-1}(s-1)^{c-a-1}\int_s^\infty\frac{d\omega}{\omega^{a+n+1}(\omega-1)^{c-a}}$$
$$=\frac{\Gamma(c+n+1)}{\Gamma(a+n+1)}\sum_{m=0}^\infty\frac{\Gamma(a+n+m+1)}{\Gamma(c+n+m+1)}$$
$$\times\frac{\Gamma(m+1)}{m!}s^{-n-m-2}. \tag{A15}$$

Therefore,

$$\mathscr{L}^{-1}\left\{s^{a-1}(s-1)^{c-a-1}\int_s^\infty\frac{d\omega}{\omega^{a+n+1}(\omega-1)^{c-a}}\right\}$$
$$=\frac{1}{c+n}\frac{1}{n!}\frac{\Gamma(c+n+1)}{\Gamma(a+n+1)}\sum_{m=0}^\infty\frac{\Gamma(+n+m+1)}{\Gamma(c+n+m+1)}$$
$$\times\frac{\mathrm{B}(n+1,m+1)}{m}z^{n+m+1}, \tag{A16}$$

where $\mathrm{B}(p,q)$ is a beta function, written as

$$\mathrm{B}(p,q)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}. \tag{A17}$$

In views of (20), (A10), (A11), (A12), (A14), and (A16),the inverse transform of (A4) can easily be taken to write (40). Note that for regular boundary condition $A=0$ and $g(0)=1$.

APPENDIX B

In this appendix we describe the method for evaluating

$$I=\int_0^\infty se^{-(\alpha-ik)s}\Psi(1+i\eta,2;\;-2iks)\,ds. \tag{B1}$$

This is facilitated by expressing Ψ in terms of the irregular Whittaker function W as

$$\Psi(1+i\eta,2;-2iks)=(-2iks)^{-1}e^{-iks}W_{-i\eta,1/2}(-2iks). \tag{B2}$$

Using (B2) in (B1), we get

$$I=-\frac{1}{2ik}\int_0^\infty ds\,e^{-\alpha s}W_{-i\eta,1/2}(-2iks). \tag{B3}$$

To transform this integral to a form suitable for our calculations, we use the integral representation for the Whittaker function

$$W_{-i\eta,1/2}(-2iks)=\frac{-2ikse^{iks}}{\Gamma(1+i\eta)}\int_0^{\infty e^{i\theta}}e^{2ikst}$$
$$\times t^{i\eta}(1+t)^{-i\eta}\,dt \tag{B4}$$

with $0<\theta<\pi$. Here θ is the measure of rotation of the path of integration for extending the domain of validity of the Laplace integral.

Substituting (B4) in (B3) and interchanging orders of integration, we have after carrying out the integration

$$I=\frac{1}{\Gamma(1+i\eta)}\int_0^{\infty e^{i\theta}}\left(\frac{t}{1+t}\right)^{i\eta}\frac{dt}{[\alpha-ik(1+2t)]^2}. \tag{B5}$$

Making the substitution of variable

$$z=t/(1+t), \tag{B6}$$

(B5) can be written in the form

$$I=\frac{1}{\Gamma(1+i\eta)(\alpha-ik)}\int_0^1 z^{i\eta}\,dz$$
$$\times\frac{d}{dz}\left[\frac{z}{(\alpha-ik)-(\alpha+ik)z}\right]. \tag{B7}$$

After partial integration, (B7) yields

$$I=\frac{1}{-2ik}\frac{1}{(\alpha-ik)\Gamma(1+i\eta)}$$
$$\times\left[1-2\eta k\int_0^1\frac{z^{i\eta}dz}{(\alpha-ik)-(\alpha+ik)z}\right]. \tag{B8}$$

Further change of variable

$$z=[(\alpha-ik)/(\alpha+ik)]u \tag{B9}$$

reduces (B8) to

$$I=\frac{1}{-2ik}\frac{1}{(\alpha-ik)\Gamma(1+i\eta)}$$
$$\times\left[1-2\eta k\frac{e^{2\eta y}}{\alpha+ik}\int_0^{z_0}\frac{u^{i\eta}du}{1-u}\right], \tag{B10}$$

where

$$z_0=e^{2iy}\quad\text{with}\quad y=\tan^{-1}k/\alpha. \tag{B11}$$

For the integral on the right-hand side of equation (B10), we separate the pole term in the integrand at $u=1$

from the more complicated part of the integrand by writing

$$\int_0^{z_0} \frac{u^{i\eta}}{1-u}\,du = \int_0^{z_0} \frac{du}{1-u} + \int_0^1 \frac{u^{i\eta}-1}{1-u}\,du - \int_{z_0}^1 \frac{u^{i\eta}-1}{1-u}\,du. \tag{B12}$$

Of the three integrals

$$\int_0^{z_0} du/(1-u) = -\ln(1-z_0), \quad |\arg(1-z_0)| < \pi. \tag{B13}$$

Combining (B11) and (B13), we get

$$\int_0^{z_0} du/(1-u) = i\pi/2 - iy - \ln(2k/y) + \tfrac{1}{2}\ln[1+(k/y)^2]. \tag{B14}$$

The second integral in (B12) can be obtained from the limit

$$\int_0^1 \frac{u^{i\eta}-1}{1-u}\,du = \lim_{\epsilon\to 0+}\left[\int_0^1 \frac{u^{i\eta}}{(1-u)^{1-\epsilon}}\,du - \int_0^1 \frac{du}{(1-u)^{1-\epsilon}}\right]. \tag{B15}$$

Using the integral representation for the beta function,

$$\mathrm{B}(r,s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} = \int_0^1 x^{r-1}(1-x)^{s-1}\,dx, \tag{B16}$$

we get (B15) in the form

$$\int_0^1 \frac{u^{i\eta}-1}{1-u}\,du = \lim_{\epsilon\to 0+}\left\{\left[\frac{\Gamma(1+i\eta)\Gamma(1+\epsilon)}{\Gamma(1+i\eta+\epsilon)} - 1\right]\Big/\epsilon\right\}. \tag{B17}$$

Straightforward application of the l'Hospital's rule converts (B17) in the form

$$\int_0^1 \frac{u^{i\eta}-1}{1-u}\,du = \psi(1) - \psi(1+i\eta), \tag{B18}$$

where ψ stands for the logarithmic derivative of the gamma function,

$$\psi(z) = \Gamma'(z)/\Gamma(z). \tag{B19}$$

For the last integral in (B12) we note that both limits of integration are in the unit circle. Thus we change the variable by

$$u = e^{2i\xi} \tag{B20}$$

and get

$$\int_{z_0}^1 \frac{u^{i\eta}-1}{1-u}\,du = -\int_0^y \frac{(1-e^{-2\eta\xi})}{\sin\xi}\,d\xi. \tag{B21}$$

The integral on the right-hand side of the above equation has the advantage that real and imaginary parts can be separated in a straightforward manner and the imaginary part evaluated in closed form. Thus

$$\int_{z_0}^1 \frac{u^{i\eta}-1}{1-u}\,du = -\frac{i}{2\eta}(2\eta y - 1 + e^{-2\eta y}) - \int_0^y (1-e^{-2\eta\xi})\cot\xi\,d\xi. \tag{B22}$$

Finally we consider the integral

$$\int_0^y (1-e^{-2\eta\xi})\cot\xi\,d\xi = y\int_0^1 (1-e^{-2\eta y s})\cot(ys)\,ds. \tag{B23}$$

with $\xi = ys$. Expanding both the exponential and cotangent in power series,[22] we get

$$\cot z = \sum_{n=0}^{\infty}(-1)^n 2^{2n}[B_{2n}/(2n)!]z^{2n-1}, \tag{B24}$$

where B_{2n} are the Bernoulli numbers; and

$$1 - e^{-2\eta y s} = -\sum_{p=1}^{\infty}\frac{(-2\eta y s)^p}{p!}. \tag{B25}$$

Using the results given in (B24) and (B25), we obtain from (B23)

$$\int_0^y (1-e^{-2\eta\xi})\cot\xi\,d\xi = -\sum_{p=1}^{\infty}\frac{(-2\eta y)^p}{p!} \times \sum_{n=0}^{\infty}(-1)^n\frac{2^{2n}B_{2n}y^{2n}}{(2n)!(2n+p)}. \tag{B26}$$

Combining the results in (B11), (B12), (B14), (B18), (B22), and (B26), we have the value of the integral I as

$$I = -\frac{1}{2ik}\frac{1}{\Gamma(1+i\eta)(\alpha^2+k^2)}\Bigg(\alpha + 2\eta k e^{2\eta y} \times\Bigg\{-\frac{i\pi}{2} + \frac{i}{2\eta} + \psi(1+i\eta) - \psi(1) + \ln\left(\frac{2k}{y}\right) - \frac{1}{2}\ln[1+(k/y)^2] + \sum_{p=1}^{\infty}\frac{(-2\eta y)^p}{p!} \times \sum_{n=0}^{\infty}(-1)^n\frac{2^{2n}B_{2n}y^{2n}}{(2n)!(2n+p)}\Bigg\}\Bigg). \tag{B27}$$

Equation (B27) is our desired result and has been used in the text. A more convenient form than the double series in (B27) can be derived by exploiting the relation between the incomplete beta function and the Gaussian hypergeometric function is

$$\mathrm{B}_x(a,b) = \int_0^x t^{a-1}(1-t)^{b-1}\,dt = a^{-1}x^a\,{}_2F_1(a,1-b;a+1;x). \tag{B28}$$

Note that relation (B28) is valid for $\mathrm{Re}\,a > 0$, which is true for our case. In (B28) we use $a = 1 + i\eta$, $b = 0$, $x = z_0$, and $t = u$ and get

$$\int_0^{z_0}\frac{u^{i\eta}}{1-u}\,du = \frac{z_0^{1+i\eta}}{(1+i\eta)}\,{}_2F_1(1+i\eta,1;2+i\eta;z_0). \tag{B29}$$

Combining (B10) and (B29), we have

$$I = -\frac{1}{2ik}\frac{1}{(\alpha-ik)\Gamma(1+i\eta)}\left[1 - 2\eta k\frac{e^{2\eta y}}{\alpha+ik}\frac{z_0^{1+i\eta}}{(1+i\eta)} \times {}_2F_1(1+i\eta,1;2+i\eta;z_0)\right]. \tag{B30}$$

[1] M. Gell-Mann and M. L. Goldberger, Phys. Rev. **91**, 398 (1953).

[2] Z. Bajzer, Nuovo Cimento **224**, 300 (1974).

[3] C. Chandler, Proceedings of the Ninth International Conference on the Few Body Dynamics (1980); Nucl. Phys. A **353**, 129C (1981).

[4] H. van Haeringen, Nucl. Phys. A **253**, 355 (1975); J. Math. Phys. **17**, 995 (1976); J. Math. Phys. **18**, 927 (1977).

[5] H. van Haeringen and R. van Wageningen, J. Math. Phys. **16**, 1441 (1975).

[6] L. P. Kok and H. van Haeringen, Phys. Rev. C **21**, 512 (1980).

[7] F. Tabakin, Phys. Rev. **174**, 1208 (1968).

[8] B. Mulligan, L. G. Arnold, B. Bagchi, and T. O. Krause, Phys. Rev. C **13**, 2131 (1976).

[9] Y. Yamaguchi, Phys. Rev. **95**, 1628 (1954).

[10] T. R. Mongan, Phys. Rev. **178**, 1597 (1969).

[11] R. Jost, Helv. Phys. Acta **20**, 256 (1947).

[12] (a) B. Talukdar, U. Das, and S. Chakravarty, Phys. Rev. C **19**, 322 (1979); (b) B. Talukdar and U. Das, Pramana **13**, 525 (1979).

[13] A. Erdelyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1.

[14] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).

[15] E. Butkov, *Mathematical Physics* (Addison-Wesley, Reading, Mass., 1973).

[16] A. W. Babister, *Transcendental Functions Satisfying Nonhomogeneous Linear Differential Equations* (MacMillan, New York, 1967).

[17] Y. L. Luke, *The Special Functions and their Approximation* (Academic, New York, 1969), Vol. 1.

[18] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, London, 1959), Vol. III.

[19] M. Coz, L. G. Arnold, and A. D. MacKellar, Ann. Phys. (N.Y.) **59**, 219 (1970).

[20] V. de Alfaro and T. Regge, *Potential Scattering* (North-Holland, Amsterdam, 1965).

[21] L. G. Arnold and R. G. Seyler, Phys. Rev. C **7**, 574 (1973).

[22] J. Mathews and R. L. Walker, *Mathematical Methods of Physics* (Benjamin, Menlo Park, CA, 1973), 2nd ed.

On the representation of electromagnetic fields in gyrotropic media in terms of scalar Hertz potentials

S. Przeździecki and W. Laprus
Institute of Fundamental Technological Research, Polish Academy of Sciences, Świętokrzyska 21, 00-049 Warsaw, Poland

(Received 27 March 1981; accepted for publication 12 June 1981)

A proof is given of the following theorem: An arbitrary sourceless electromagnetic field determined in a region of a gyrotropic medium can be represented therein in terms of two scalar functions, called the scalar Hertz potentials, that fulfill a system of two second order partial differential equations. Some restrictions are imposed on the region, and their implications are discussed.

PACS numbers: 41.10.Hv, 42.10.Qj

I. INTRODUCTION

This paper is a modified version of a report.[1] The scalar Hertz potentials for gyrotropic media have been introduced in Ref. 2. The authors have presented there the following theorem.

Theorem 1: An electromagnetic field **E**, **H** generated in an arbitrary region D of a gyrotropic medium from two scalar functions u, v via formulas

$$\mathbf{E} = \boldsymbol{\epsilon}^{-1}\cdot\nabla\times\boldsymbol{\epsilon}\cdot\nabla\times u\mathbf{z}_0 + i\omega\mu\epsilon^{-1}\tilde{\boldsymbol{\epsilon}}\cdot\nabla\, v\,\mathbf{z}_0, \tag{1a}$$

$$\mathbf{H} = -i\omega\epsilon\mu^{-1}\tilde{\boldsymbol{\mu}}\cdot\nabla\times u\mathbf{z}_0 + \boldsymbol{\mu}^{-1}\cdot\nabla\times\boldsymbol{\mu}\cdot\nabla\times\, v\mathbf{z}_0, \tag{1b}$$

satisfies in D the homogeneous set of Maxwell's equations

$$\nabla\times\mathbf{H} = -i\omega\boldsymbol{\epsilon}\cdot\mathbf{E}, \tag{2a}$$

$$\nabla\times\mathbf{E} = i\omega\boldsymbol{\mu}\cdot\mathbf{H}, \tag{2b}$$

if the functions u and v fulfill in D the system of equations

$$\left(\nabla_t^2 + \frac{\epsilon_a}{\epsilon}\frac{\partial^2}{\partial z^2} + k_e^2\right)u = -\omega\mu\tau_g\frac{\epsilon_a}{\epsilon}\frac{\partial v}{\partial z}, \tag{3a}$$

$$\left(\nabla_t^2 + \frac{\mu_a}{\mu}\frac{\partial^2}{\partial z^2} + k_m^2\right)v = \omega\epsilon\tau_g\frac{\mu_a}{\mu}\frac{\partial u}{\partial z}. \tag{3b}$$

The time dependence is assumed to be given by the factor $\exp(-i\omega t)$ which is suppressed throughout. A system of Cartesian coordinates x,y,z is introduced in which the permittivity and permeability tensors have the following forms:

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon & -i\epsilon_g & 0 \\ i\epsilon_g & \epsilon & 0 \\ 0 & 0 & \epsilon_a \end{pmatrix}, \tag{4a}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \mu & -i\mu_g & 0 \\ i\mu_g & \mu & 0 \\ 0 & 0 & \mu_a \end{pmatrix}. \tag{4b}$$

The other symbols are defined as follows:

$$k_e^2 = \omega^2\epsilon_a\frac{\mu^2 - \mu_g^2}{\mu}, \quad k_m^2 = \omega^2\mu_a\frac{\epsilon^2 - \epsilon_g^2}{\epsilon},$$

$$\tau_g = \frac{\epsilon_g}{\epsilon} + \frac{\mu_g}{\mu}, \quad \nabla_t = \nabla - \mathbf{z}_0\frac{\partial}{\partial z},$$

where $\mathbf{z}_0$ is a unit vector directed along the z axis, and the tilde denotes the transpose of a matrix.

The proof of this theorem follows just from a substitution of (1) into (2).

We recall that a medium is said to be gyrotropic if in an appropriate system of Cartesian coordinates the tensors $\boldsymbol{\epsilon}$ and $\boldsymbol{\mu}$ have the forms given by (4), where not both ϵ_g and μ_g are zero. The z axis is called the distinguished axis of the medium. If $\epsilon_g = \mu_g = 0$ but not both $\epsilon_a = \epsilon$ and $\mu_a = \mu$, the medium is uniaxial.

The scalar Hertz potentials defined by Theorem 1 have been employed to good purpose in solving a half-plane diffraction problem for a gyrotropic medium.[3] It is believed that they will prove useful in many more electromagnetic problems for gyrotropic media. However, similarly as for isotropic media, there immediately arises the question of how general is the class of electromagnetic fields generated via (1) in a region D by the set of all functions u, v satisfying in D the system (3); or, more specifically, does this class coincide with the class of all sourceless fields determined in D? Though this question was already posed in Ref. 2, it was left open, and it is the aim of the present paper to provide an answer to it.

II. FORMULATION OF THE PROBLEM

The problem of the generality of representation (1) may be formulated as follows:

Can an arbitrary electromagnetic field satisfying the set (2) in a region D of a gyrotropic medium be represented via (1) in terms of two scalar functions u, v that fulfill the system (3)?

For isotropic media this question was formulated and answered by Bochenek[4] (for a class of regions).

In this paper we present a generalization of Bochenek's result to the case of the scalar Hertz potentials for gyrotropic media. We also extend his answer to a somewhat broader class of regions.

The following representation theorem summarizes the results of the present paper.

Theorem 2 (representation theorem): An arbitrary electromagnetic field **E**, **H** determined in a sufficiently simple region D of a gyrotropic medium and satisfying therein the set (2) can be represented in D in terms of two scalar functions u, v in form (1) with u, v satisfying the system (3).

The restriction on the region D to be of sufficiently simple shape means that any straight line parallel to $\mathbf{z}_0$ must not have more than one interval in common with D. Alternatively, we shall formulate this property by saying that D has to be

0022-2488/82/091708-05$02.50

convex with respect to the z axis as shown in Fig. 1.

We denote by D_0 the projection of D on a plane $z = \text{const}$.

In order to simplify the proof, we shall assume that there exists in D a surface S, given by the equation $z = z_0(x,y)$, whose projection on the plane $z = \text{const}$ coincides with D_0 and $z_0(x,y)$ has continuous second derivatives (see Fig. 1).

The role of the convexity restriction will be discussed after the proof has been presented.

Theorems 1 and 2 can be considered as mutually inverse provided Theorem 1 is confined to regions for which Theorem 2 holds.

Let us rewrite the system (3) in the form

$$\mathscr{H}\mathbf{w} = (\mathscr{T} + \mathscr{L})\mathbf{w} = 0, \tag{5}$$

where

$$\mathscr{T} = \begin{pmatrix} \nabla_t^2, & 0 \\ 0 & \nabla_t^2 \end{pmatrix},$$

$$\mathscr{L} = \begin{pmatrix} \dfrac{\epsilon_a}{\epsilon}\dfrac{\partial^2}{\partial z^2} + k_e^2; & \omega\mu\tau_g\dfrac{\epsilon_a}{\epsilon}\dfrac{\partial}{\partial z} \\ -\omega\epsilon\tau_g\dfrac{\mu_a}{\mu}\dfrac{\partial}{\partial z}, & \dfrac{\mu_a}{\mu}\dfrac{\partial^2}{\partial z^2} + k_m^2 \end{pmatrix},$$

$$\mathbf{w} = \begin{pmatrix} u \\ v \end{pmatrix}.$$

The letter $\mathscr{H}$ is used in (5) to stress that the system (3) is a generalization of the Helmholtz equation satisfied by the Hertz potentials in isotropic media. The operators $\mathscr{T}$ and $\mathscr{L}$ are the transverse and longitudinal parts of $\mathscr{H}$, respectively.

From (1) we have

$$\mathbf{F} = -\mathbf{K}\mathscr{T}\mathbf{w}, \tag{6}$$

where

$$\mathbf{K} = \begin{pmatrix} \epsilon/\epsilon_a & 0 \\ 0 & \mu/\mu_a \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} E_z \\ H_z \end{pmatrix}.$$

In view of (5) we also have

$$\mathbf{F} = \mathbf{K}\mathscr{L}\mathbf{w}$$

or (7)

$$\mathbf{K}^{-1}\mathbf{F} = \mathscr{L}\mathbf{w}.$$

The basic idea in the proof of Theorem 2 is to find for a given $\mathbf{F}$ such a solution to (7) that it simultaneously satisfies the set (5). It may seem surprising that a solution can fulfill two systems but this is possible because, as shown in Ref. 2, $\mathbf{F}$ fulfills the set

$$\mathscr{H}\mathbf{K}^{-1}\mathbf{F} = 0. \tag{8}$$

An alternative proof could be given by looking for a solution to (6) that would fulfill (5).

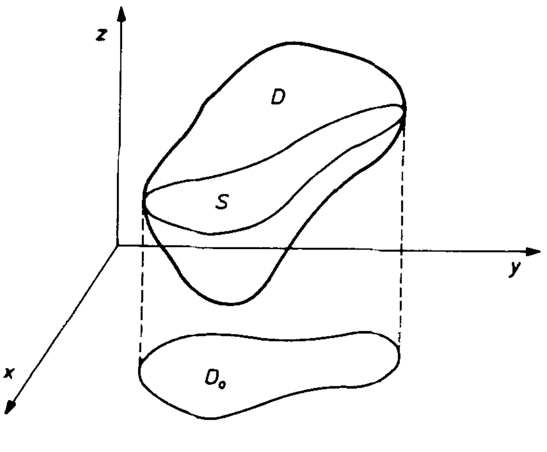


FIG. 1. Region D, its projection D_0 on xy plane, and surface S. Projection of S coincides with D_0.

III. THE PROOF

A. Formal scheme

To highlight the structure of the proof, we first present its basic scheme in a formal way. The feasibility of the relevant steps will be demonstrated further on.

We consider an arbitrary field $\mathbf{E}$, $\mathbf{H}$ given in D and satisfying therein the system (2). For this field we find $\mathbf{w}_1$ from relation (7) thus

$$\mathscr{L}\mathbf{w}_1 = \mathbf{K}^{-1}\mathbf{F}. \tag{9}$$

Only as an exception could it happen that $\mathbf{w}_1$ would fulfill (5); this would essentially end the proof. However, with no loss of generality, we may write

$$\mathscr{H}\mathbf{w}_1 = \mathbf{p}, \tag{10}$$

where $\mathbf{p}$ is some function. In the next subsection we show that

$$\mathscr{L}\mathbf{p} = 0. \tag{11}$$

We now construct $\mathbf{w}_2$ so that

$$\mathscr{T}\mathbf{w}_2 = -\mathbf{p}, \tag{12a}$$

$$\mathscr{L}\mathbf{w}_2 = 0. \tag{12b}$$

Finding a function $\mathbf{w}_2$ that satisfies simultaneously the two systems of equations constitutes the crucial step in the proof.

For $\mathbf{w}_{12} = \mathbf{w}_1 + \mathbf{w}_2$ we have

$$\mathbf{F} = \mathbf{K}\mathscr{L}\mathbf{w}_{12} \tag{13}$$

and

$$\mathscr{H}\mathbf{w}_{12} = 0. \tag{14}$$

Let us denote by $\mathbf{E}'$, $\mathbf{H}'$ the field corresponding to $\mathbf{w}_{12}$ via (1). Then the field $\mathbf{E} - \mathbf{E}'$,$\mathbf{H} - \mathbf{H}'$is of TEM type with respect to $\mathbf{z}_0$ (in particular it could be zero). It can be shown that any TEM field in D can be represented in terms of scalar Hertz potentials (see subsection 3E). Let us denote by $\mathbf{w}_t$ the potentials for the considered TEM field. Then the sum

$$\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_t \tag{15}$$

constitutes the potentials for the considered field $\mathbf{E}$, $\mathbf{H}$, which ends the proof.

B. Equation (11)

$$\begin{aligned}\mathscr{L}\mathbf{p} &= \mathscr{L}\mathscr{H}\mathbf{w}_1 = \mathscr{L}(\mathscr{T} + \mathscr{L})\mathbf{w}_1 = (\mathscr{T} + \mathscr{L})\mathscr{L}\mathbf{w}_1 \\ &= \mathscr{H}\mathscr{L}\mathbf{w}_1 = \mathscr{H}\mathbf{K}^{-1}\mathbf{K}\mathscr{L}\mathbf{w}_1 = \mathscr{H}\mathbf{K}^{-1}\mathbf{F} = 0. \end{aligned} \tag{16}$$

The final equality in (16) follows from (8).

C. The function $\mathbf{w}_1$

An explicit form for $\mathbf{w}_1$ is not necessary for the validity of our proof. What we need to know is that such a function exists and fulfills Eq. (10). However, for the considered case of a homogeneous medium we can easily construct an explicit solution (see the Appendix). It has the form

$$\mathbf{w}_1 = \frac{-1}{2k^2\tau_g\gamma_1}\int_{z_0}^{z}\mathbf{g}(x,y,z')\sin\gamma_1(z-z')\,dz' + \frac{1}{2k^2\tau_g\gamma_2}\int_{z_0}^{z}\mathbf{g}(x,y,z')\sin\gamma_2(z-z')\,dz', \tag{17}$$

where

$$\mathbf{g} = \mathscr{N}\mathbf{F},$$

$$\mathscr{N} = \begin{pmatrix} \dfrac{\partial^2}{\partial z^2} + \dfrac{\mu}{\mu_a}k_m^2, & -\omega\mu\tau_g\dfrac{\partial}{\partial z} \\ \omega\epsilon\tau_g\dfrac{\partial}{\partial z}, & \dfrac{\partial^2}{\partial z^2} + \dfrac{\epsilon}{\epsilon_a}k_e^2 \end{pmatrix}, \quad k^2 = \omega^2\epsilon\mu,$$

and γ_1, γ_2 are determined by

$$\gamma_n^2 = \omega^2(\epsilon \pm \epsilon_g)(\mu \pm \mu_g),$$

the upper sign is for $n = 1$ and the lower for $n = 2$.

D. Construction of $\mathbf{w}_2$

The general solution to (11) has in D the following form:

$$\mathbf{p} = \sum_{n=1}^{n=4} p_n(x,y)\begin{pmatrix}1\\ \zeta_n\end{pmatrix}e^{i\gamma_n z} \tag{18}$$

where

$$\gamma_3 = -\gamma_1, \quad \gamma_4 = -\gamma_2,$$

$$\zeta_1 = -\zeta_3 = -i\frac{\epsilon}{\mu}\left(\frac{\mu+\mu_g}{\epsilon+\epsilon_g}\right)^{1/2},$$

$$\zeta_2 = -\zeta_4 = i\frac{\epsilon}{\mu}\left(\frac{\mu-\mu_g}{\epsilon-\epsilon_g}\right)^{1/2},$$

and p_n are arbitrary functions.

We now solve (12a) assuming $\mathbf{p}$ to be given by (18). We get

$$\mathbf{w}_2 = \sum_{n=1}^{n=4} q_n(x,y)\begin{pmatrix}1\\ \zeta_n\end{pmatrix}e^{i\gamma_n z}, \tag{19}$$

where

$$q_n(x,y) = \frac{1}{2\pi}\iint_{D_0} p_n(\xi,\eta)\ln\frac{1}{\rho}\,d\xi d\eta,$$

$$\rho = [(x-\xi)^2 + (y-\eta)^2]^{1/2}. \tag{20}$$

Obviously $\mathbf{w}_2$ also satisfies Eq. (12b).

E. TEM field

For TEM fields the system of Maxwell's equations (2) reduces to the following system [cf. (8) and (9) in Ref. 2]:

$$\frac{\partial\mathbf{E}}{\partial z} = -i\omega\mathbf{z}_0\times\boldsymbol{\mu}\cdot\mathbf{H}, \quad \frac{\partial\mathbf{H}}{\partial z} = i\omega\mathbf{z}_0\times\boldsymbol{\epsilon}\cdot\mathbf{E},$$
$$\nabla\cdot\mathbf{E} = 0, \qquad \nabla\cdot\mathbf{H} = 0. \tag{21}$$

It can be shown that the general solution to (21) has in D the form

$$\mathbf{E} = \sum_{n=1}^{n=4}[\mathbf{e}_n(x,y) - (-1)^n i\mathbf{z}_0\times\mathbf{e}_n(x,y)]e^{i\gamma_n z},$$

$$\mathbf{H} = \sum_{n=1}^{n=4}[\mathbf{h}_n(x,y) - (-1)^n i\mathbf{z}_0\times\mathbf{h}_n(x,y)]e^{i\gamma_n z}, \tag{22}$$

where $\mathbf{e}_n\cdot\mathbf{z}_0 = 0$ and the fields $\mathbf{e}_n$ fulfill the conditions

$$\nabla\cdot\mathbf{e}_n = 0, \quad \nabla\cdot(\mathbf{z}_0\times\mathbf{e}_n) = 0, \tag{23}$$

$\mathbf{h}_n$ are given by $\mathbf{h}_n = (\omega/\gamma_n)\mathbf{z}_0\times\boldsymbol{\epsilon}\cdot\mathbf{e}_n$.

The planar fields $\mathbf{e}_n$ can now be represented in D_0 in the following way:

$$\mathbf{e}_n = (-1)^n\omega\mu\zeta_n\nabla\varphi_n + i\omega\mu\nabla\times\psi_n\mathbf{z}_0. \tag{24}$$

This representation follows as a particular case from the Helmholtz theorem (see, for example, Ref. 5).

From (23) we get

$$\nabla_t^2\varphi_n = 0, \tag{25a}$$

$$\nabla_t^2\psi_n = 0. \tag{25b}$$

We can now define the scalar potentials u_t, v_t for the fields (22) in the form

$$\mathbf{w}_t = \sum_{n=1}^{n=4}\left[\varphi_n(x,y) + \frac{1}{\zeta_n}\psi_n(x,y)\right]\begin{pmatrix}1\\ \zeta_n\end{pmatrix}e^{i\gamma_n z}. \tag{26}$$

It can be easily checked that

$$\mathscr{H}\mathbf{w}_t = 0 \tag{27}$$

and that the fields (22) are determined from (26) via formulas (1).

A more detailed discussion of TEM fields will be presented in Ref. 6.

F. The convexity condition for the region D

We shall demonstrate by way of an example that the convexity condition imposed on D is necessary for the validity of Theorem 2 in the case of an isotropic or uniaxial medium. Since the proof for a gyrotropic medium is exactly patterned after the isotropic case, this example suggests that the restriction on D is essential in general.

Consider the electromagnetic field of an electric dipole of moment $\mathbf{p}$ perpendicular to $\mathbf{z}_0$. We denote by x_d, y_d, z_d the coordinates of the dipole. For an isotropic or uniaxial medium (z axis being distinguished) the scalar Hertz potentials for the considered field are given in each of the half-spaces $z < z_d$ and $z > z_d$ by the functions Π, M determined in Ref. 7. These functions, however, are not the potentials for the whole space since their z derivatives are discontinuous across the plane $z = z_d$. Denote these discontinuities by

$$\left[\frac{\partial\Pi}{\partial z}\right] = \frac{\partial\Pi}{\partial z}\bigg|_{z_d+0} - \frac{\partial\Pi}{\partial z}\bigg|_{z_d-0}$$

and similarly for $[\partial M/\partial z]$. Let us note moreover that these discontinuities are singular at the point $x = x_d$, $y = y_d$.

If we now consider a convex region not containing the dipole but crossed by the plane $z = z_d$, then by virtue of Theorem 2 there must exist a way to compensate the discontinuities of $\partial\Pi/\partial z$ and $\partial M/\partial z$. It is indeed so, since the Hertz potentials for an electromagnetic field are not determined uniquely. In other words, there exist Hertz potentials different from zero that generate the zero field. To demonstrate this, we observe first that these potentials must have the form of the potentials for a TEM field, given by (26), since obviously for the zero field we have $E_z = H_z = 0$. Secondly, one can easily find transverse fields $\mathbf{e}_n$ in formulas (22) that lead to the zero TEM field. The Hertz potentials $\mathbf{w}_t$ corresponding to these fields $\mathbf{e}_n$ yield the field equal to zero, such potentials will be called null potentials (ghost potentials in Ref. 3).

Let us now state explicitly that by virtue of formulas (26) the discontinuities $[\partial \Pi/\partial z]$, $[\partial M/\partial z]$ determine the corrective potentials uniquely in that part of the considered convex region ($z < z_d$ or $z > z_d$) that is not pierced by the line $x = x_d, y = y_d$. Consequently, we compensate these discontinuities by adding these potentials to Π and M in that part of the convex region.

However, this corrective technique, which is the only possible one, must fail for a concave region D' shown in Fig. 2 when the line $x = x_d, y = y_d$ pierces the region D' on both sides of the point $z = z_d$. In this case we are not able to construct the corrective null potentials with no singularity along the line $x = x_d, y = y_d$. The singularity present in $\partial \Pi/\partial z$, $\partial M/\partial z$ would be carried along this line spoiling the corrective null potentials.

IV. CONCLUDING REMARKS

Theorem 2 clarifies some of the basic facts connected with the representation of electromagnetic fields in terms of the scalar Hertz potentials. Its main significance stems from the information it provides about the generality of this representation. For example, in solving an electromagnetic problem with the aid of the potentials we now can avoid the unpleasant situation in which we would not know whether the electromagnetic field to be determined can be represented via potentials.

Somewhat philosophically, one might also remark that Theorem 2 explains why it is not possible to forget about Maxwell's equations altogether and employ only the system of equations for the potentials (restrictions on D).

It has been shown in Ref. 2 that the idea of auxiliary functions for electromagnetic fields can be extended further by introducing the so-called superpotentials which generate the scalar potentials. Analogously to Theorem 2, it can be shown that an arbitrary pair of scalar potentials given in the considered simple region D and satisfying therein system (3) can be represented in terms of one superpotential satisfying its fourth-order equation. Thus, indirectly, Theorem 2 leads to a remarkable result that an arbitrary electromagnetic field in a gyrotropic medium can be derived from only one scalar function while in an isotropic medium two functions are necessary. This result becomes less surprising if we note that in a gyrotropic medium the components E_z and H_z are coupled by system (8) while in an isotropic medium E_z and H_z can be two arbitrary, independent solutions of the Helmholtz equation.

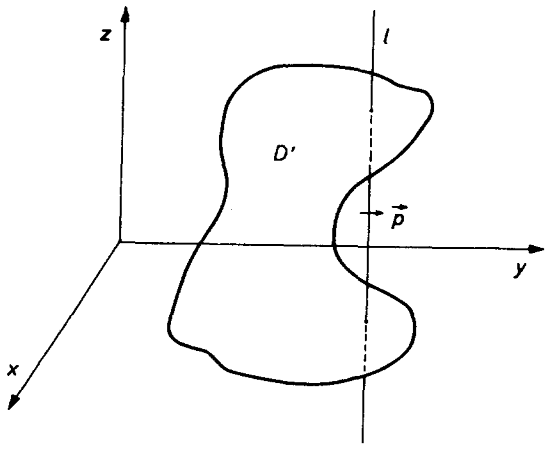


FIG. 2. Location of the dipole with respect to the concave region D'.

Finally we have two more remarks on possible applications of Theorem 2 to two cases of some practical significance. The first one concerns cylindrical regions whose axes are perpendicular to the z axis. The second one involves media stratified along z.

(1) For cylindrical regions and fields harmonic in one of the transverse coordinates [e.g., given by $\exp(i\alpha x)$ or $\exp(i\beta y)$] the convexity condition on the region D becomes unnecessary and the theorem can be proved for any region D that can be divided by planes $z = z_\nu$ ($\nu = 1,2,\cdots$) into subregions convex with respect to z in such a way that each of the planes $z = z_\nu$ has only one strip in common with D.

(2) The theorem can be extended to gyrotropic media stratified along the distinguished axis, i.e., for ϵ and μ varying with z. The formal scheme of the proof and relation (16) remain valid in this case though the operators $\mathscr{H}$ and $\mathscr{L}$ and the matrix $\mathbf{K}$ take other forms than used here [cf. formulas (28), (29), and (30) in Ref. 2]. As was already observed, the closed form for $\mathbf{w}_1$ is redundant and we can content ourselves only with the existence and regularity of $\mathbf{w}_1$. What we really need in subsection 3D is the existence of four linearly independent solutions for Eqs. (11) and (12b). This follows from the relevant theorems on the systems of ordinary differential equations[8] (nonsingular and continuous). More essential modifications are necessary concerning the representation of TEM fields. In this case the definition of the potentials via (26) has to be changed and the proof of the relation (27) becomes much more complicated. These considerations will be carried out elsewhere. Jump discontinuities in ϵ,μ can be dealt with by an appropriate division of D.

APPENDIX

Let us change in the system (9) the variable z to $s = z - z_0(x,y)$ and consider this system for $s > 0$. The Laplace transform of (9) takes the form

$$\hat{\mathscr{L}}\hat{\mathbf{w}}_1 = \mathbf{K}^{-1}\hat{\mathbf{F}}^{-1} \tag{A1}$$

where $\hat{\mathscr{L}}$ is the Laplace transform of $\mathscr{L}$ and $\hat{\mathbf{w}}_1(x,y,p)$ and $\hat{\mathbf{F}}(x,y,p)$ are the Laplace transforms of $\mathbf{w}_1(x,y,s+z_0)$ and $\mathbf{F}(x,y,s+z_0)$, respectively, i.e.,

$$\hat{\mathbf{w}}_1 = \int_0^\infty \mathbf{w}_1(x,y,s+z_0)e^{-ps}\,ds,$$

similarly for $\hat{\mathbf{F}}$.

From (A1) we get

$$\hat{\mathbf{w}}_1 = (\det\mathbf{K}\hat{\mathscr{L}})^{-1}\hat{\mathbf{g}}, \tag{A2}$$

with $\hat{\mathbf{g}} = \hat{\mathscr{N}}\hat{\mathbf{F}}$, where $\hat{\mathscr{N}} = (\mathbf{K}\hat{\mathscr{L}})^{-1}\det\mathbf{K}\hat{\mathscr{L}}$ is the Laplace transform of $\mathscr{N}$, and $\hat{\mathbf{g}}(x,y,p)$ is the Laplace transform of

$\mathbf{g}(x,y,s+z_0)$ (see subsection 3C).

The inverse of the determinant in (A2) can be rewritten as follows:

$$(\det \mathbf{K}\hat{\mathscr{L}})^{-1} = \frac{1}{(p^2+\gamma_1^2)(p^2+\gamma_2^2)} = \frac{1}{2k^2\tau_g}\left(\frac{-1}{p^2+\gamma_1^2} + \frac{1}{p^2+\gamma_2^2}\right). \quad \text{(A3)}$$

In order to find the inverse Laplace transform of $\hat{\mathbf{w}}_1$, we make use of (A3) and apply to (A2) the convolution theorem. We obtain

$$\mathbf{w}_1 = \frac{-1}{2k^2\tau_g\gamma_1}\int_0^s \mathbf{g}(x,y,s+z_0)\sin\gamma_1(s-s')\,ds' + \frac{1}{2k^2\tau_g\gamma_2}\int_0^s \mathbf{g}(x,y,s+z_0)\sin\gamma_2(s-s')\,ds' \quad \text{(A4)}$$

or returning to the variable $z = s + z_0$

$$\mathbf{w}_1 = \frac{-1}{2k^2\tau_g\gamma_1}\int_{z_0}^z \mathbf{g}(x,y,z')\sin\gamma_1(z-z')\,dz' + \frac{1}{2k^2\tau_g\gamma_2}\int_{z_0}^z \mathbf{g}(x,y,z')\sin\gamma_2(z-z')\,dz', \quad \text{(A5)}$$

which coincides with (17).

One may easily check that (A5) constitutes a solution also for $z < z_0$.

[1] S. Przeździecki and W. Laprus, "Representation of electromagnetic fields in gyrotropic media in terms of scalar Hertz potentials," Inst. Fund. Tech. Res. Polish Acad. Sci. Report 17 (1978) (in Polish).

[2] S. Przeździecki and R. A. Hurd, "A note on scalar Hertz potentials for gyrotropic media," Appl. Phys. **20**, 313–7 (1979).

[3] S. Przeździecki and R. A. Hurd, "Diffraction by a half-plane perpendicular to the distinguished axis of a gyrotropic medium (arbitrary incidence)," Can. J. Phys. **59**, 403–24 (1981).

[4] K. Bochenek, *Methods of Analysis of Electromagnetic Fields* (PWN, Warsaw, 1961) (in Polish), p. 100.

[5] R. Plonsey and R. E. Collin, *Principles and Applications of Electromagnetic Fields* (McGraw-Hill, New York, 1961), Sec. 1.18.

[6] W. Laprus and S. Przeździecki, "TEM fields in gyrotropic media and their representation in terms of scalar Hertz potentials" (in preparation).

[7] P. C. Clemmow, *The Plane Wave Spectrum Representation of Electromagnetic Fields* (Pergamon, Oxford, 1966), Chap. VIII.

[8] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).

The pattern space factor and quality factor of cylindrical source antennas

John M. Jarem
Department of Electrical Engineering, University of Texas, El Paso, Texas 79968

(Received 2 December 1980; accepted for publication 15 January 1982)

For the first time the quality factor of cylindrical source antennas is derived by a plane wave expansion. The evanescent energy (and therefore the quality factor) as defined by a plane wave expansion is shown to be different from Collin and Rothschild's [IEEE Trans. Antennas Propagation **AP-12**, 23 (1964)] quality factor.

PACS numbers: 41.10.Hv, 84.40. – x, 02.30.Mv

I. INTRODUCTION

The problem of defining a suitable quality factor (ratio of 2 times frequency times the greater of the electric and magnetic energies in the evanescent fields to the power radiated in the antenna system[1]) for planar, cylindrical, and spherical antenna sources has received considerable attention in recent years.[1-8] In this problem of defining a suitable quality factor, basically two methods have been used to define the energies in the evanescent fields. The first method is to express the EM fields in the system as an infinite sum of propagating (visible region of **k**-space) and nonpropagating (invisible region of **k**-space) plane waves, find the electric and magnetic energies associated with each plane wave, and then sum these energies only over the evanescent or nonpropagating values of the wavenumbers. This method has been used by Rhodes[2] and Collin and Rothschild[3] for planar aperture antennas.

The second method introduced by Collin and Rothschild[1] for finding the energy in the evanescent field is based on the recognition that there exists an energy density in the evanescent field which is given by the difference between the total electric or magnetic energy densities $\frac{1}{4}\epsilon_0\mathbf{E}\cdot\mathbf{E}^*$ and $\frac{1}{4}\mu_0\mathbf{H}\cdot\mathbf{H}^*$ in space and the energy density associated with radiated power flow namely $P_{\text{rad}}/v_{\text{energy flow}} = (\mu_0\epsilon_0)^{1/2}P_{\text{rad}}$. In this method the total energy in the evanescent field is given as an integral over all space of this evanescent energy density. This method has been used by Collin and Rothschild,[1] Kalafus,[4] and Fante[5] to find the quality factor of cylindrical and spherical antenna sources.

At the present time the first method (which is to define the evanescent energies as a sum of nonpropagating plane waves) has only been applied to finding the Q for planar aperture antennas and not for cylindrical or spherical antennas. This investigation will be concerned with showing that the quantities defined by Collin and Rothschild[1] as the evanescent energies of a cylindrical antenna system are not the same as the evanescent energies as defined by a plane wave sum over the nonpropagating waves as is done in the case of a planar aperture. This will be shown by calculating the plane wave evanescent energy in the region $y \geqslant 0$ which results from a source located in the region $y < 0$, $r < a$. This energy will be shown to be infinite for arbitrary source distributions. This then shows the difference between the two evanescent energies since Collin and Rothschild[1] evanescent energy was finite.

II. ANALYSIS

The subsequent analysis will be concerned with finding the quality factor for cylindrical antenna sources where (for simplicity) the radiating source is taken to be an axial magnetic current source which excites only TE modes given by (this is the same source as used by Ref. 1).

$$\begin{aligned} &\mathbf{M} = M_z(x,y)e^{j(-\beta_z z + \omega t)}\hat{z} \quad \text{for} \\ &r = (x^2 + y^2)^{1/2} < a, y < 0, \\ &\mathbf{M} = 0 \quad \text{for} \quad r > a \quad \text{or} \quad y \geqslant 0 \\ &\beta_z < (\mu_0\epsilon_0)^{1/2}\omega = k, \end{aligned} \tag{1}$$

and where the quality factor of the antenna system is given by[1]

$$Q = 2\omega[W_E, W_M]/P,$$

where $[W_E, W_M]$ represents the greater of the evanescent electric and magnetic energies, ω is the operating frequency, and P is the total power radiated in the radial direction. As mentioned in the Introduction, the electric and magnetic evanescent energies will be defined in terms of a wavenumber summation over the invisible region of the antenna system.

The first step in the analysis will be to obtain expressions for the evanescent electric and magnetic fields in the system. This may be accomplished by expanding the fields in a plane wave expansion over the visible and invisible wavenumber and then keeping only that portion of the fields which have resulted from summation over the invisible wavenumber region. To this end, we note, as shown in Appendix A, that a plane wave expansion of the **E** and **H** fields in the region $y > 0$, due to the source **M** of Eq. (1), is given by ($E_z = 0$)

$$H_z = H_0 H = H_0 \int_{-\infty}^{\infty} \frac{F_+(u)}{(1-u^2)^{1/2}} \times e^{-j(uR\cos\phi + (1-u^2)^{1/2}R\sin\phi)}du, \tag{2}$$

$$0 < \phi < \pi,$$

$$H_r = \frac{-j\beta_z H_0}{\kappa}\frac{\partial H}{\partial R}, \tag{3}$$

$$H_\phi = \frac{-j\beta_z H_0}{\kappa R}\frac{\partial H}{\partial \phi}, \tag{4}$$

$$E_r = \frac{-j\omega\mu_0 H_0}{\kappa R}\frac{\partial H}{\partial\phi}, \tag{5}$$

$$E_\phi = \frac{j\omega\mu_0 H_0}{\kappa}\frac{\partial H}{\partial R}, \tag{6}$$

where

$$(1-u^2)^{1/2} = -j(u^2-1)^{1/2} \quad \text{when } |u|>1,$$

$$\kappa = (\omega^2/c^2 - \beta_z{}^2)^{1/2}, \quad H_0 = 1A/m,$$

$$R = \kappa r, \quad X = \kappa x = R\cos\phi, \quad Y = \kappa y = R\sin\phi.$$

In this expression $F_+(u)$ is the pattern space factor of the system and is given in Appendix A in terms of M_z.

Following our previous statements that the evanescent fields are defined by keeping only that portion of the wavenumber summation which is over the invisible region (in this case $|u|>1$), we find that the evanescent fields of the system $\mathbf{H}^e$, $\mathbf{E}^e$ may be found by calculating

$$H^e = \int_{|u|>1}\frac{F_+(u)}{-j(u^2-1)^{1/2}}e^{-j(uR\cos\phi)-(u^2-1)^{1/2}R\sin\phi}du \tag{7}$$

and substituting this function H^e in place of H in Eqs. (2)–(6).

In the source free region $Y<0$, $R>\kappa a$, the H_z field may found from the plane wave expansion

$$H_z = H_0 H = \int_{-\infty}^{\infty}\frac{F_-(u)}{(1-u^2)^{1/2}}e^{j(uR\cos\phi + (1-u^2)^{1/2}R\sin\phi)}du,$$
$$\pi<\phi<2\pi, \quad R>\kappa a,$$

where $F_-(u)$ is the pattern space factor of this region. In this expression $(1-u^2)^{1/2} = -j(u^2-1)^{1/2}$ when $|u|>1$.

Now that the evanescent fields have been defined, we may now calculate the evanescent energies of the system. As mentioned in the Introduction, it is only necessary to calculate the evanescent energy in the region $y\geqslant 0$ to show the difference between the evanescent energy as defined by Collin and Rothschild[1] and as defined by a plane wave expansion. Calling the evanescent energies in the region $y\geqslant 0$ $W_E{}^+$ and $W_M{}^+$, we find that the expressions for these quantities are given by

$$W_E{}^+ = \lim_{\substack{R_0\to\infty\\ \epsilon\to 0}}\frac{\mu_0 H_0{}^2}{4\kappa^2}\left[\frac{k^2}{\kappa^2}\int_{0+\epsilon}^{\pi-\epsilon}d\phi\int_0^{R_0}R\,dR \times\left(\frac{1}{R^2}\frac{\partial H^e}{\partial\phi}\frac{\partial H^{e*}}{\partial\phi}+\frac{\partial H^e}{\partial R}\frac{\partial H^{e*}}{\partial R}\right)\right], \tag{8}$$

$$W_M{}^+ = \lim_{\substack{R_0\to\infty\\ \epsilon\to 0}}\frac{\mu_0 H_0{}^2}{4\kappa^2}\left\{\int_{0+\epsilon}^{\pi-\epsilon}d\phi\int_0^{R_0}R\,dR \times\left[\frac{\beta_z{}^2}{\kappa^2}\left(\frac{1}{R^2}\frac{\partial H^e}{\partial\phi}\frac{\partial H^{e*}}{\partial\phi} + \frac{\partial H^e}{\partial R}\frac{\partial H^{e*}}{\partial R}\right)+H^e H^{e*}\right]\right\}. \tag{9}$$

To proceed further, we differentiate H^e in (7) with respect to ϕ and R as indicated in (8) and (9) and substitute into (8) and (9). At this point, following a procedure exactly analogous to that of Ref. 2 (pp. 64,65), we interchange spatial and wavenumber integrals to obtain the expression:

$$W_E{}^+ = \frac{\mu_0 H_0^2}{4\kappa^2}\left(\frac{k^2}{\kappa^2}\int_{|u|>1}du\int_{|u'|>1}du'\,F_+(u)F_+^*(u') \times\frac{[uu'+(u^2-1)^{1/2}(u'^2-1)^{1/2}]}{(u^2-1)^{1/2}(u'^2-1)^{1/2}}K(u,u')\right), \tag{10}$$

$$W_M{}^+ = \frac{\mu_0 H_0^2}{4\kappa^2}\int_{|u|>1}du\int_{|u'|>1}du'\,F_+(u)F_+^*(u') \times\frac{\{(\beta_z{}^2/\kappa^2)[uu'+(u^2-1)^{1/2}(u'^2-1)^{1/2}]+1\}}{(u^2-1)^{1/2}(u'^2-1)^{1/2}} \times K(u,u'), \tag{11}$$

where

$$K(u,u') = \lim_{\substack{R_o\to\infty\\ \epsilon\to 0}}\int_0^{R_0}R\,dR\int_{0+\epsilon}^{\pi-\epsilon}e^{(j\alpha\cos\phi-\beta\sin\phi)R}d\phi,$$

where $\alpha = u'-u$ and $\beta = (u'^2-1)^{1/2}+(u^2-1)^{1/2}$. The integration interchange is justified since the interchange has been made before letting $R_0\to\infty$, $\epsilon\to 0$.

If we change from polar to rectangular coordinates with $X = R\cos\phi$ and $Y = R\sin\phi$, $K(u,u')$ becomes after letting $R_0\to\infty$,

$$K(u,u') = \int_0^\infty dY\,e^{-\beta Y}\int_{-\infty}^{\infty}dX\,e^{j\alpha X} = (2\pi/\beta)\delta(u-u'). \tag{12}$$

This expression has been obtained by realizing that the integral over x is 2π times the delta function and the integral over y is $1/\beta$ when the integration over the u' variable in Eqs. (10) and (11) is carried out. The final expression for $W_E{}^+$ and $W_M{}^+$ is found.

Instead of presenting $W_E{}^+$ and $W_M{}^+$ directly we will form the expressions $Q_E{}^+ = \omega W_E{}^+/P$ and $Q_M{}^+ = \omega W_M{}^+/P$ in order that a comparison can be made with the quality factor as obtained by a plane wave expansion and as obtained by Ref. 1. In these expressions P is the real radiated power and, as shown in Appendix B, may be expressed in terms of the pattern space factors $F_\pm$ integrated over the visible region $-1\leqslant u\leqslant 1$. The final expression for $Q_E{}^+$ and $Q_M{}^+$ is given by

$$Q_E{}^+ = \frac{\frac{1}{2}(k^2/\kappa^2)\int_{|u|>1}|F_+(u)|^2[(2u^2-1)/(u^2-1)^{3/2}]du}{\int_{-1}^{1}(|F_+|^2+|F-|^2)du/(1-u^2)^{1/2}}, \tag{13}$$

$$Q_M{}^+ = \frac{\frac{1}{2}\int_{|u|>1}|F_+(u)|^2\{[(\beta_z{}^2/\kappa^2)(2u^2-1)+1]/(u^2-1)^{3/2}\}du.}{\int_{-1}^{1}(|F+|^2+|F-|^2)du/(1-u^2)^{1/2}} \tag{14}$$

Of course it is necessary to recognize that to form a full expression for the Q ($Q = \max[Q_E, Q_M]$) as defined by a plane wave expansion, terms which represent the evanescent electric and magnetic energy in the region $R>\kappa a$, $Y<0$ must be added to the numerators of Eqs. (13) and (14) to form the full expressions for Q_E and Q_M. These expressions have not been derived since Eqs. (13) and (14) are sufficient to show the difference between Ref. 1 and the Q as defined by a plane wave expansion.

Clearly from Eqs. (13) and (14) for the those source distributions M_z for which $F_+(u)$ does not vanish at $u = \pm 1$, the numerators of Eqs. (13) and (14) diverge. On the other hand, the evanescent energies as defined by Collin and Rothschild[1] converge for all source distributions confined to the region $R<\kappa a$ which produce single modes (or a finite number of single modes).

As a specific example, let us calculate the evanescent energy as defined by Ref. 1 of a delta magnetic current source $M_z = M_0\delta(X)\delta(Y+\kappa a/2)$. It is easy to see that in a coordinate system $X', Y' (X' = X, Y' = Y + \kappa a/2)$, located at the center of the delta source, the only mode that will be excited is the $H_0^{(2)}(kr')$ mode [see Eq. (A2) of Appendix A], where $r' = \kappa(x'^2 + y'^2)^{1/2}$. The evanescent energy for this mode as given by Ref. 1 (pp. 25,26) will be finite when calculated over the region $R' \geqslant R_0'$, where R_0' is a nonzero distance satisfying $0 < R_0' < \kappa a/2$.

At this point let us calculate the evanescent energy of the above delta source using a plane wave expansion. We, first of all, find for the above delta function that the pattern space factor $F_+(u)$ does not vanish $u = \pm 1$ [it is proportional to $e^{-(u^2-1)^{1/2}\kappa a/2}$ for $|u| > 1$; see Eq. (A5)], and, since it does not vanish at $u = \pm 1$, we conclude, according to our earlier statements, that the evanescent energies $W_E^{\ +}$ and $W_M^{\ +}$ calculated in the region $Y \geqslant 0$, must approach infinity [see Eqs. (13) and (14)].

Thus we see, in this example, that the evanescent energy, as calculated by a plane wave expansion, is infinite in the region $Y \geqslant 0$ whereas the evanescent energy as calculated by Ref. 1 in a region which includes the region $Y \geqslant 0$ turned out to be finite. If we recall that the evanescent energy density is greater than or equal to zero at each point in space, either as calculated by Ref. 1 or as calculated by a plane wave expansion, we then clearly see, at least for this example, that the evanescent energy as calculated by Ref. 1 is not the same because it is finite, as the evanescent energy as calculated by a plane wave expansion because it is infinite. We have thus completed a major objective of the investigation which is to show that the evanescent energies as defined by Ref. 1 are not the same as the evanescent energy as defined by a plane wave expansion.

Several additional interesting statements can be made about the above expressions. We first of all note that the form of the Eqs. (13) and (14) is in direct analogy to the quality factors derived by Rhodes[2] and Collin and Rothschild[3] for planar aperture antennas. The evanescent energy is expressed as an integral of the pattern space factor squared over the invisible region, and the power is expressed as a pattern space factor squared over the visible region. Furthermore, the pattern space factor is in direct analogy with the pattern space factor as derived by Refs. 2 and 3 in that it is expressed as an integral transform of the source field [using Eq. (A5)] whereas the integral transform of Refs. 2 and 3 for the planar aperture was just the finite Fourier transform of the aperture field. Another interesting feature of comparison is the fact that both of the numerators of $Q_E^{\ +}, Q_M^{\ +}$ have terms which are proportional to $2/(u^2-1)$ and $1/(u^2-1)^{3/2}$ [note that $(2u^2-1)/(u^2-1)^{3/2} = 2/(u^2-1) + 1/(u^2-1)^{3/2}$], which are precisely the same terms which multiply the squared pattern space factor of the H-plane strip source antenna and the E-plane strip source antenna of the planar aperture case. See Refs. 2 and 3.

III. CONCLUSION

In conclusion, the main contribution this author has tried to make in this paper is the fact that the expression for the evanescent energy of a cylindrical radiator as defined by a plane wave expansion is not necessarily the same as the evanescent energy as defined by Collin and Rothschild[1] (the difference between the total energy and an energy of power flow). This contribution was found by showing that the evanescent energy in the region $Y \geqslant 0$ due to a source in the region $R < \kappa a, Y < 0$ was in some cases infinite (when the pattern space factor of the source did not vanish at $u = \pm 1$), whereas the evanescent energy found by Ref. 1 was always finite for sources which produce single modes. An example of the evanescent energy of a delta source was given, and the evanescent energy was found to be infinite for a plane wave expansion but finite when calculated by Ref. 1.

APPENDIX A

The purpose of this appendix is to derive the pattern space factor $F_+(u)$ from the magnetic current source M_z. We first note that the H_z component of the fields satisfies the following wave equation (all coordinates are unnormalized):

$$(\nabla_t^2 + \kappa^2)H_z(x,y) = (j\omega\epsilon_0\kappa^2/k^2)M_z(x,y). \tag{A1}$$

If we set the rhs of (A1) to a delta function $\delta(x)\delta(y)$, then the solution of (A1) will be given by (Ref. 10, p. 823),

$$r = (x^2+y^2)^{1/2},$$

$$g(x,y) = \frac{j}{4} H_0^{(2)}(\kappa r)$$

$$= \int_{-\infty}^{\infty} \left(\frac{j}{4\pi}\right) \frac{e^{-j[k_x x + (\kappa^2 - k_x^2)^{1/2} y]}}{(\kappa^2 - k_x^2)^{1/2}}\, dk_x. \tag{A2}$$

If the rhs of (A1) is set equal to $(j\omega\epsilon_0\kappa^2/k^2)M_z(x_0,y_0)$ $\delta(x-x_0)\delta(y-y_0)$, then by simple translation the solution for the function at x_0, y_0 will be

$$\psi(x,y|x_0,y_0) = (j\omega\epsilon_0\kappa^2/k^2)M_z(x_0,y_0)\, g(x-x_0, y-y_0)$$

$$= \int_{-\infty}^{\infty} \left(\frac{-\omega\epsilon_0\kappa^2 M_z(x_0,y_0)}{4\pi k^2} e^{j[k_x x_0 + (\kappa^2 - k_x^2)^{1/2} y_0]} \right)$$

$$\times \frac{e^{-j[k_x x + (\kappa^2 - k_x^2)^{1/2} y]}}{(\kappa^2 - k_x^2)^{1/2}}\, dk_x. \tag{A3}$$

If the above ψ are added everywhere that the $M_z(x_0,y_0)$ is not zero, then the superposition of these ψ will be H_z

$$H_z = \int_{-\infty}^{\infty} \left\{ \int\!\!\int_{V_0} dx_0 dy_0 \frac{-\omega\epsilon_0\kappa^2 M_z(x_0,y_0)}{4\pi k^2} \right.$$

$$\left. \times -e^{j[k_x x_0 + (\kappa^2 - k_x^2)^{1/2} y_0]} \right\}$$

$$\times \frac{e^{-j[k_x x + (\kappa^2 - k_x^2)^{1/2} y]}}{(\kappa^2 - k_x^2)^{1/2}}\, dk_x. \tag{A4}$$

The expression in curly brackets is the unnormalized pattern space factor of the system. If we made the change of variables $X_0 = \kappa x_0, Y_0 = \kappa y_0$, and $u = k_x/\kappa$ in both the volume and wavenumber integrals, and also divided H_z by $H_0 = 1$ A/m, the following expression results:

$$H = \int_{\infty}^{\infty} \left\{ \frac{-\omega\epsilon_0}{4\pi k^2 H_0} \int_{V_0} dX_0\, dY_0\, M_z(X_0,Y_0) e^{j[uX_0 + (1-u^2)^{1/2} Y_0]} \right\}$$

$$\times \frac{e^{-j[uX + (1-u^2)^{1/2} Y]}}{(1-u^2)^{1/2}}\, du, \quad Y_0 < 0, \ Y \geqslant 0. \tag{A5}$$

The expression in curly brackets is the pattern space factor $F_+(u)$ of Eq. (A3).

APPENDIX B

The power may be obtained from the Poynting vector $\frac{1}{2}\text{Re}(\mathbf{E}\times\mathbf{H}^*)$ evaluated on a surface $r\rightarrow\infty$ surrounding the source. The term $\frac{1}{2}\mathbf{E}\times\mathbf{H}^*$ becomes entirely real as $r\rightarrow\infty$ and we have

$$P=\lim_{r\rightarrow\infty}\tfrac{1}{2}\int_0^{2\pi}\mathbf{E}\times\mathbf{H}^*\cdot rd\phi\hat{r}=\lim_{r\rightarrow\infty}\frac{r}{2}\int_0^{2\pi}E_\phi H_z^* d\phi. \tag{B1}$$

The expressions for H_z and E_ϕ as $r\rightarrow\infty$ are found an asymptotic expansion of Eq. (7) which turns out to be $(R=\kappa r)$

$$H_z=H_0F_\pm(\cos\phi)(2\pi/R)^{1/2}e^{-j(R-\pi/4)} \tag{B2}$$

and a similar expression for E_ϕ as given in (3e). Substitution of (A2) into (A1) yields, after letting $u=\cos\phi$,

$$P=(\pi H_0^2\omega\mu_0/\kappa^2)\int_{-1}^{1}[|F_+(u)|^2+|F_-(u)|^2]du/(1-u^2)^{1/2}.$$

[1] R. E. Collin and S. Rothschild, "Evaluation of Antenna Q," IEEE Trans. Antennas Propagation **AP-12**(1), 23–27 (1964).

[2] D. R. Rhodes, *Synthesis of Planar Aperture Antennas* (Oxford U.P., Oxford, 1974).

[3] R. E. Collin and S. Rothschild, "Reactive Energy in Aperture Fields and Aperture Q," Can. J. Phys. **41**, 1967–1979 (1963).

[4] R. M. Kalafus, On the Evaluation of Antenna Quality Factors," IEEE Trans. Antennas Propagation **AP-17**(6), 729–32 (1969).

[5] R. L. Fante, "Quality Factor of General Ideal Antennas," IEEE Trans. Antennas Propagation **AP-17**(2), 151–5 (1969).

[6] G. V. Borgiotti, "On the Reactive Energy of an Aperture," IEEE Trans. Antennas Propagation **AP-15**, 565–6 (1967).

[7] R. E. Collin and D. R. Rhodes, "Stored Energy, Q, and Frequency Sensitivity of Planar Aperture Antennas," IEEE Trans. Antennas Propagation, **AP-15**, 567–9 (1967).

[8] D. R. Rhodes "Observable Stored Energies of Electromagnetic Systems," J. Franklin Inst. **302**(3), 225–237 (1976).

[9] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., Cambridge, 1966).

[10] D. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).

On the phase transition of the one-dimensional Percus–Yevick equation for an arbitrary potential of finite range

M. Chen
Department of Mathematics, Vanier College, 821 Ste. Croix Blvd., St. Laurent, Quebec, H4L 3X9, Canada

(Received 23 February 1981; accepted for publication 12 June 1981)

A qualitative investigation of the one-dimensional Percus–Yevick integral equation by perturbation method is discussed for an arbitrary potential of finite range l. When the particle density ρ is restricted to the interval (0,1) it is proved that every order of perturbation has a unique continuous and bounded solution, which can be expressed as a convergent generalized Fourier series. The perturbation series is absolutely and uniformly convergent if the supremum norm of the nth order solution is less than or equal to $n!$. Under the assumptions (i), $0<\rho<1$ and (ii), the absolute and uniform convergence of the perturbation series, it can be proved that the Percus–Yevick equation cannot exhibit a phase transition.

PACS numbers: 64.60.Cn, 34.20. – b

I. INTRODUCTION

The problem of phase transition for the one-dimensional system of classical fluids has been of great interest in the past. For the nearest neighbor interaction, Gürsey[1] had shown that there was no phase transition for the one-dimensional system. This conclusion was further generalized by Van Hove[2] to an arbitrary attractive potential $v(x)$ of finite range l. On the other hand, Kac *et al.*[3] has investigated the parametric limit of the potential and shown that there indeed does exist a phase transition as $l \rightarrow \infty$ in the van der Waals limit. The delicate nature of phase transition depends markedly upon the model potentially employed.[4]

In 1958 Percus and Yevick proposed an approximate theory for the pair distribution function[5] in classical fluids. Since then much work has been developed in its various applications. It is generally agreed that the Percus–Yevick (PY) approximation has been quite successful, particularly, in the one-dimensional case where it becomes an exact theory for the hard rod potential. Recently Wertheim[6] had studied the PY equation for the nearest neighbor interaction. He concluded that the one-dimensional PY equation could not exhibit a phase transition. The purpose of this paper is to further generalize Wertheim's conclusion for an arbitrary attractive potential of finite range by a perturbative method.[7]

In this paper we assume that the intermolecular potential consists of a hard rod potential $u_0(x)$ of diameter 1 and an arbitrary attractive potential $v(x)$ of finite length l. By considering $v(x)$ as a perturbation on $u_0(x)$ we can obtain a set of coupled integral-differential equations from the PY equation which can then be transformed into a set of differential-difference equations of advanced and retarded types. When the particle density ρ is restricted to the interval (0, 1), every order of perturbation for the pair distribution function has a unique continuous and bounded solution which can be expressed as a generalized fourier series expansion. The perturbation series can be shown to be absolutely and uniformly convergent if the supremum norm of the nth order solution is less than or equal to $n!$. Finally we prove that the PY equation cannot exhibit a phase transition under the assumptions $0<\rho<1$ and the absolute and uniform convergence of the perturbation series.

II. PERTURBATION SERIES

Consider the intermolecular potential

$$u(x) = u_0(x) - \lambda \xi v(x),$$

where

$$u_0(x) = \begin{cases} \infty, & |x| < 1, \\ 0, & |x| \geqslant 1, \end{cases}$$

$$v(x) = \begin{cases} 0, & |x| \leqslant 1 \text{ or } |x| \geqslant l, \\ \text{a positive smooth function for } |x| \in [1,\ l,]. \end{cases}$$

ξ denotes the maximum of the physical tail potential so that $\text{Max}|v(x)| = 1$ and $0 \leqslant |\lambda| \leqslant 1$. For convenience, we set l to be a positive integer. Let $\beta = 1/KT$, where K is the Boltzmann constant, and T is the temperature. We define

$$f(x) = e^{-\beta u(x)} - 1,$$

$$y(x) = e^{\beta u(x)} g(x),$$

$$h(x) = g(x) - 1,$$

where $f(x)$ is the Mayer's function, $g(x)$ is the pair distribution function and $h(x)$ is the total correlation function. Following Ornstein and Zernike,[8] the total correlation function can be written as a sum of the direct correlation function $c(x)$ and an indirect correlation function by a convolution as follows:

$$h(x) = c(x) + \rho \int_{-\infty}^{\infty} h(x - x')c(x')\, dx'. \tag{1}$$

The convolution relation (1) is usually called the Ornstein–Zernike (OZ) relation, which can be considered as the definition of $c(x)$. Suppose $c(x) = 0$ for $|x| \geqslant l$. It has been proved[9] that the one-dimensional OZ relation (1) can be transformed into the following equivalent Baxter's relations:[10]

$$c(x) = Q(x) - \rho \int_x^l Q(x')Q(x' - x)\, dx', \quad 0 < x < l, \tag{2}$$

$$h(x) = Q(x) + \rho \int_0^l h(x - x')Q(x')\, dx', \quad x \geqslant 0, \tag{3}$$

where $Q(x)$ is a real bounded function for $0 \leqslant x \leqslant l$, and $Q(x) = 0$ for $x < 0$, or $x \geqslant l$. The Percus–Yevick approximation assumes $c(x) = f(x)y(x)$, so that $c(x) = 0$ for $|x| \geqslant l$. The one-dimensional PY integral equation can be written as a coupled integral-differential equation by supplementing the Baxter's relations (2) and (3) with the PY approximation. For hard sphere potential the PY equation can easily be solved with exact solution. However, the problem becomes very difficult for any realistic potential.

If the attractive potential $-\xi v(x)$ is considered as a perturbation on the hard rod potential $u_0(x)$, we can obtain a series expansion in $\lambda\beta\xi$ for $f(x)$,

$$\begin{aligned} f(x) &= e^{-\beta u_0(x)} e^{\lambda\beta\xi v(x)} - 1 \\ &= f_0(x) + \sum_{n=1}^{\infty} \frac{\lambda^n}{n!} f_n(x), \end{aligned}$$

where

$$f_0(x) = e^{-\beta u_0(x)} - 1,$$

$$f_n(x) = e^{-\beta u_0(x)} [\beta\xi v(x)]^n.$$

Similarly we can have the following perturbation series expansions:

$$Q(x) = Q_0(x) + \sum_{n=1}^{\infty} \frac{1}{n!} (\lambda\beta\xi)^n Q_n(x), \tag{4}$$

$$y(x) = y_0(x) + \sum_{n=1}^{\infty} \frac{1}{n!} (\lambda\beta\xi)^n y_n(x), \tag{5}$$

$$h(x) = h_0(x) + \sum_{n=1}^{\infty} \frac{1}{n!} (\lambda\beta\xi)^n h_n(x), \tag{6}$$

$$c(x) = c_0(x) + \sum_{n=1}^{\infty} \frac{1}{n!} (\lambda\beta\xi)^n c_n(x), \tag{7}$$

where the subscript "0" in $Q_0(x)$, $y_0(x)$, $h_0(x)$ and $c_0(x)$ denotes the unperturbed system with hard rod potential $u_0(x)$, and

$$h_n(x) = e^{-\beta u_0(x)} \sum_{i=0}^{n} \binom{n}{i} [v(x)]^i y_{n-i}(x), \quad n \geqslant 1,$$

$$c_n(x) = e^{-\beta u_0(x)} \sum_{i=0}^{n} \binom{n}{i} [v(x)]^i y_{n-1}(x) - y_n(x), \quad n \geqslant 1,$$

$$Q_0(x) = Q = -1/(1-\rho).$$

From the PY integral equation and Eqs. (4)–(7), it then follows that

$$\begin{cases} c_0(x) = Q_0(x) - \rho \int_x^1 Q_0(x') Q_0(x'-x)\, dx', & 0 \leqslant x \leqslant 1, \\ h_0(x) = Q_0(x) + \rho \int_0^1 h_0(x-x') Q_0(x')\, dx', & x \geqslant 0, \\ c_0(x) = f_0(x) y_0(x), & x \geqslant 0, \end{cases} \tag{8}$$

and

$$\begin{aligned} c_n(x) = Q_n(x) - \int_x^l [Q_0(x') Q_n(x'-x) + Q_n(x') Q_0(x'-x)]\, dx' \\ - \rho \sum_{i=1}^{n-1} \int_x^l Q_i(x') Q_{n-i}(x'-x)\, dx', \quad 0 \leqslant x \leqslant l, \end{aligned} \tag{9}$$

$$h_n(x) = Q_n(x) + \rho \sum_{i=1}^{n} \binom{n}{i} \int_o^l h_{n-i}(x-x') Q_i(x')\, dx', \quad x \geqslant 0. \tag{10}$$

Note that Eq. (8) is the PY equation for the hard rod potential, whose solution is well known.[11] Due to the nature of the intermolecular potential, Eqs. (9) and (10) can be further simplified. After some lengthy derivations, we finally obtain the following results:

$$-Q_n(x)$$

$$= \begin{cases} A_n(x) + \rho \int_{x+1}^{l} h_0(x'-x) Q_n(x')\, dx', & 0 < x < 1, \quad (11) \\ B_n(x) - \rho Q \int_x^{x+1} Q_n(x')\, dx', & 1 < x \leqslant l-1, \quad (12) \\ B_n(x) - \rho Q \int_x^{l} Q_n(x')\, dx', & l-1 \leqslant x < l, \quad (13) \end{cases}$$

$$y_n(x) = \begin{cases} E_n(x) + \rho Q \int_x^{x+1} Q_n(x')\, dx', & 0 < x < 1, \quad (14) \\ F_n(x) + \rho Q \int_1^{x} y_n(x')\, dx', & 1 < x < 2, \quad (15) \\ G_n(x) + \rho Q \int_{x-1}^{X} y_n(x')\, dx', & 2 < x < \infty, \quad (16) \end{cases}$$

where A_n, B_n are functions of Q_m and y_m for $m < n$, whereas E_n, F_n, and G_n are functions of Q_m, y_m and Q_n, so that in the nth order perturbation they can be considered as known functions. Because of their complexity, the detailed expressions of A_n, $B_n \cdots G_n$ are omitted since we will not need them in the subsequent discussions.

It is interesting to note that, if the attractive potential is considered as a perturbation on the hard sphere potential, we can then express $Q(x)$, $C(x)$, and $y(x)$ in a series expansion of $\beta\xi$, which is similar to the ordinary density expansion. Moreover, the PY equation can be reduced to a set of coupled linear differential-integral equations (11)–(16).

III. SOLUTIONS OF PERTURBATION SERIES

Equations (11)–(16) are related. However, Eq. (13) is self-contained and can therefore be solved first. With the solution of Q_n in $[l-1, l]$ we can successively solve Eqs. (12), (11), (14), (15), and finally Eq. (16).

(a) Solution of $Q_n(x)$ for $l - 1 \leqslant x \leqslant l$

$$Q_n(x) = -B_n(x) + \rho Q \int_x^l Q_n(x')\, dx'. \tag{13}$$

For the hard rod potential it is known that $y_0(x)$ is a real entire function on $(m, m+1)$, but of class C^{m-2} on $[m, m+1]$ where m is a positive integer greater than or equal to 2. By induction we can then deduce that $B_n(x)$ is at least of class C^{l-3} on $[l-1, l]$ depending on the smoothness of $v(x)$. Since $Q_n(l) = 0$, Eq. (13) can be transformed into a differential equation.

$$Q_n'(x) + \rho Q Q_n(x) = -B_n'(x) = s(x)$$

with solution

$$Q_n(x) = -\int_x^l e^{-\rho Q(x-t)} s(t)\, dt.$$

(b) Solution of $Q_n(x)$ for $1 \leqslant x \leqslant l-1$.

$$Q_n(x) = -B_n(x) + \rho Q \int_x^{x+1} Q_n(t)\,dt. \tag{12}$$

Equation (12) can be transformed into a differential-difference equation of the advanced type

$$Q'_n(x) + \rho Q Q_n(x) - \rho Q Q_n(x+1) = -B'_n(x) = b_n(x), \tag{17}$$

with the initial condition given by the solution of Eq. (13) for $l-1 \leqslant x \leqslant l$. By the standard continuation method the following theorem can easily be proved.

Theorem 1: There exists a unique continuous function $Q_n(x)$ which satisfies Eq. (17) for $x \in [1, l-1]$ and the initial condition given by the solution of Eq. (13) for $x \in [l-1, l]$. Furthermore, Q_n is a real entire function in each subinterval $(i, i+1)$, and is at least of class C^1 on $[i, i+1]$, $i = 1, 2, \ldots, l-2$, except possibly at $x = 2$.

In order to further study the properties of $Q_n(x)$ we next consider the Laplace transform of Eq. (17).
Let

$$\widetilde{Q}_n(t) = \int_1^{l-1} Q_n(x)\, e^{-tx}\, dx,$$

$$B(t) = \int_{l-1}^{l} Q_n(x)\, e^{-tx}\, dx - \int_1^2 Q_n(x)\, e^{-tx}\, dx,$$

$$\tilde{s}(t) = \int_1^{l-1} b_n(x)\, e^{-tx}\, dx,$$

$$F(t) = e^{-t}\tilde{s}(t) + b_0 B(t) - Q_n(l-1)\, e^{-tl} + Q_n(1) e^{-2t},$$

$$H(t) = te^{-t} + b_0\, e^{-t} - b_0,$$

$$b_0 = \rho Q.$$

The Laplace transform of Eq. (17) yields

$$\widetilde{Q}_n(t) = H^{-1}(t) F(t). \tag{18}$$

In taking the Laplace transform it must be assumed that Q_n is known on [1, 2] from the continuation method. The information of Q_n on [1, 2] and $[l-1, l]$ appears in $B(t)$.

Before we employ the inverse Laplace transform, we first consider the distribution of zeros of $H(t)$.

Lemma 1: All roots of $H(t) = 0$ are simple.

Proof:

From $H(t)$ we can obtain $H'(t) = e^{-t}(1 - b_0 - t)$ and $h''(t) = -e^{-t}(2 - b_0 - t)$. Let $H'(t) = 0$. Then $t = 1 - b_0$. But $h''(1 - b_0) \neq 0$, whereas $H(1 - b_0) = e^{-(1-b_0)} - b_0$. Consider $u(x) = e^{-(1-x)} - x$. It is evident that $u(x) \geqslant 0$ and $u(x) = 0$ at $x = 1$. Hence H has a double root if and only if $b_0 = 1$, which implies $b_0 = \rho Q = -\rho/(1-\rho) = 1$, i.e., $\rho = \pm\infty$.

Let $G(t) = e^t H(t) = -b_0 e^t + b_0 + t$. $G(t)$ and $H(t)$ have the same roots. Since $G(t) = (-b_0 e^t + t)[1 + \epsilon(t)]$, where $\epsilon(t) \to 0$ as $|t| \to \infty$, for large $|t|$ the roots of $G(t)$ are asymptotic to the comparison function

$$G_c(t) = -b_0 e^t + t.$$

We now consider the distribution of roots of the exponential polynomial $G_c(t)$, which can be recast in the form

$$G_c(t) = p_0 t^{m_0} e^{\beta_0 t} + p_1 t^{m_1} e^{\beta_1 t},$$

with $p_0 = 1, m_0 = 1, \beta_0 = 0, p_1 = -b_0, m_1 = 0, \beta_1 = 1$. The distribution diagram of the exponential polynomial $G_c(t)$ consists of a line L passing through two points (β_0, m_0) and (β_1, m_1) with slope -1 in the β-m plane.

Define the curvilinear strip V_1 by

$$V_1 : |\mathrm{Re}(t - \ln(t))| \leqslant c_1,$$

where c_1 is a constant to be specified later. The strip V_1 is bounded by a curve $\mathrm{Re}(t - \ln(t)) =$ const with the following characteristics[12]:

(i) If $t = x + iy$ lies on the curve, then $|y/x| \to \infty$ and $|\arg(t)| \to \pi/2$ as $|t| \to \infty$, i.e., $|t| = y(1 + O(1))$ as $|t| \to \infty$.

(ii) The curve is asymptotic to the curve $x - \ln(|y|) = \text{const}$.

(iii) The curve lies entirely in a right half-plane and $\mathrm{Re}(t) \to \infty$ as $t \to \infty$.

By Theorems 12.9 and 12.10 of Bellman–Cooke,[13] all zeros of large modulus of $G(t)$ lie within V_1, and the zeros in V_1 are asymptotically the same as those of $G_c(t)$ comprised of the terms associated with points on the line L of the distribution diagram.

In order to consider the distribution of zeros in V_1 we let

$$G_c(t) = tG_1(t),$$

where

$$G_1(t) = 1 - b_0 t^{-1} e^t = 1 - b_0\, e^{t - \ln(t)}.$$

By the transformation T: $t \to z$ defined by

$$z = t - \ln(t),$$

$G_1(t)$ is transformed into $f(z)$ given by

$$f(z) = 1 - b_0 e^z,$$

with roots

$$\begin{aligned} z &= \ln(|1/b_0|) + 2n\pi i \\ &= \ln(|1 - \rho)/\rho|) + 2n\pi i, \quad n = 0, \pm 1, \pm 2, \cdots. \end{aligned} \tag{19}$$

Since T is a one-to-one transformation,[14] there exists a one-to-one correspondence between the zeros of large modulus of $G_1(t)$ and the zeros of large modulus of $f(z)$. By the inverse transformation of T the zeros of $G_c(t)$ lie along a curve $\mathrm{Re}(t - \ln(t)) = c_1 = \ln(|1 - \rho)/\rho|)$. In fact, if we let $t = x + iy$, then

$$x - \ln(|t|) = \mathrm{Re}(z) = \ln(|1 - \rho)/\rho|),$$

$$y - \arg(t) = \mathrm{Im}(z) = 2n\pi.$$

But $|\arg(t)| \to \pi/2$ and $|t| = |y|(1 + O(1))$; consequently, we have

$$\begin{aligned} y &= \arg(t) + \mathrm{Im}(z) = (2n \pm \tfrac{1}{2})\pi, \quad n\text{: large integer}, \\ x &= \ln(|t|) + \ln(|(1 - \rho)/\rho|) \\ &= \ln(|1 - \rho)/\rho|) + \ln(|(2n \pm \tfrac{1}{2})\pi|) + O(1). \end{aligned} \tag{20}$$

Summarizing our result, we have

Theorem 2: The zeros of $G(t)$ form a root chain of advanced type lying asymptotically along a curve $|t^{-1} e^t| = \ln(|(1 - \rho)/\rho|)$. For large modulus of t the roots have the form given by Eq. (20).

We can now employ the inverse Laplace transform of Eq. (18) and obtain

$$Q_n(x) = \frac{1}{2\pi i}\int_\tau H^{-1}(t)F(t)\,e^{tx}\,dt$$
$$= \mathrm{Res}[H^{-1}(t)F(t)\,e^{tx}] = \sum_{n=1}^{\infty}\frac{F(t_n)}{H'(t_n)}e^{t_n x},$$

where t_n is a root of $H(t)$ and the summation is taken over all characteristic roots of $H(t)$.

By Theorem 6.10 of Bellman–Cooke,[15] the generalized Fourier series expansion given above is uniformly convergent for $2\leqslant x\leqslant l-1$. As emphasized in Sec. 6.10 of Bellman–Cooke, this finite transform method is valid only for finite l. As $l\to\infty$, the Laplace integral $\int_1^{l-1}Q_n(x)e^{-tx}\,dx$ diverges and the method breaks down subsequently.

(c) Solution of $Q_n(x)$ for $0<x<1$.

$$Q_n(x) = -A_n(x) - \rho\int_{x+1}^{l} h_0(t-x)Q_n(t)\,dt. \tag{11}$$

Once we know the solution of Eq. (12), the solution of Eq. (11) will follow immediately by a simple integration.

(d) Solution of $y_n(x)$ for $0<x<1$.

$$y_n(x) = E_n(x) + \rho Q\int_0^1 Q_n(x+t)\,dt. \tag{14}$$

Again the solution of Eq. (14) can be obtained easily by simple integration.

(e) Solution of $y_1(x)$ for $1\leqslant x\leqslant 2$.

$$y_n(x) = F_n(x) + \rho Q\int_1^x y_n(t)\,dt. \tag{15}$$

We can transform Eq. (15) into a differential equation

$$y_n'(x) - \rho Q y_n(x) = F_n'(x),$$

which has the solution

$$y_n(x) = e^{+\rho Qx}\left\{e^{-\rho Q}y_n(1^-) + \int_1^x e^{-\rho Qt}F_n'(t)\,dt\right\}, \tag{21}$$

where we have made use of $y_n(1^-)$ as the boundary condition so that y_n is continuous at $x=1$.

(f) Solution of $y_n(x)$ for $2\leqslant x<\infty$.

$$y_n(x) = G_n(x) + \rho Q\int_{x-1}^{x} y_n(t)\,dt. \tag{16}$$

Equation (16) can be transformed into a first order retarded differential-difference equation

$$y_n' - \rho Q y_n(x) + \rho Q y_n(x-1) = G_n'(x) = \alpha(x) \tag{22}$$

with the initial condition given by Eq. (21).

Since y_n is continuous on [1, 2] and α is at least of class C^1 on $(2,\infty)$, by the continuation method we can obtain the following result:

Theorem 3: There exists a unique continuous function y_n which satisfies Eq. (22) for $x\geqslant 2$ and the initial condition given by Eq. (21) for $1\leqslant x\leqslant 2$. Moreover, y_n is at least of class C^1 on $(2,\infty)$ and at least of class C^2 on $(3,\infty)$.

Due to the fact that both Q_n and $v(x)$ vanish for $x\geqslant l$, by generalizing Theorem 3.5 of Bellman–Cooke[16] we can obtain an exponential bound for $y_n(x)$, $|y_n(x)|\leqslant K_1\,e^{k_2(x-2)}$, $x\geqslant 2$, K_1, K_2,: positive constants. We can now take the Laplace transform of Eq. (22). Let

$$\tilde{y}_n(t) = \int_2^\infty y_n(x)e^{-tx}\,dx,$$
$$\tilde{\alpha}(t) = \int_2^\infty \alpha(x)e^{-tx}\,dx,$$
$$\beta(t) = \tilde{\alpha}(t) + y_n(2)\,e^{-2t} - \rho Q\,e^{-t}\int_1^2 y_n(x)\,e^{-tx}\,dx.$$

Then

$$\tilde{y}_n(t) = R^{-1}(t)\beta(t), \tag{23}$$

where

$$R(t) = t - \rho Q + \rho Q e^{-t}. \tag{24}$$

Similar to Lemma 1 we have:

Lemma 2: All roots of $R(t)$ are simple. Moreover, $R(t)$ has a real root only at $t=0$ if $\rho<1$, and has two real roots only, one at $t=0$, the other on the positive real axis if $\rho>1$.

Since $R(t)$ and $I(t) = e^tR(t) = te^t - \rho Qe^t + \rho Q$ have the same roots, we now consider the distribution of zeros of $I(t)=0$ instead. For large modulus of t we find

$$I(t)\sim te^t[1+\epsilon(t)] + \rho Q,$$

where $\lim_{|t|\to\infty}\epsilon(t)=0$. Let $I_c(t) = te^t + \rho Q$. For large $|t|$, the roots of $I(t)=0$ are asymptotic to the roots of the comparison function $I_c(t)=0$. The distribution diagram of the roots of $I_c(t)$ contains the points (0, 0) and (1,1), showing that there is a single chain of roots of retarded type. Hence for large modulus of t the roots of $R(t)$ have the asymptotic form

$$x = \ln(|\rho Q|) - \ln(2k\pi) + O(1),$$
$$y = 2k\pi + \arg(-\rho Q)\mp\pi/2 + O(1),$$

where K is any integer of large magnitude. The upper sign applies to roots for which $y\to+\infty$, the lower sign to roots for which $y\to-\infty$.

We next prove that all roots of $R(t)$ lie in the left half-plane (lhp) except for the root at $t=0$. Let $t = x+iy$ be a root of $R(t)$. Then

$$y^2 + [x+\rho/(1-\rho)]^2 = [\rho/(1-\rho)]^2\,e^{-2x}$$

and $x\leqslant 0$ if $0<\rho<1$. Alternatively, we can consider the perturbation of roots of $I(t)$ by a small positive parameter ϵ. Let

$$I_2(t) = -(\rho+\epsilon)Q\,e^t + \rho Q + te^t$$
$$= -\left[\frac{-\rho-\epsilon}{1-\rho}\,e^t + \frac{\rho}{1-\rho} - t\,e^t\right] = 0.$$

By Hay's theorem[17] all roots of $I_2(t)$ lie in the left half-plane if and only if $\rho<1$. When $|t|\ll 1$, $I_2(t)$ has a root at $t = -\epsilon/(1-\rho)$ in the lhp, which moves toward $t=0$ as $\epsilon\to 0$. Since both $I_2(t)$ and $I(t)$ are entire functions, $I(t) = \lim_{\epsilon\to 0}I_2(t)$, and the roots of $I_2(t)$ depend continuously on ϵ; the roots of $I(t)$ will coincide with the corresponding roots of $I_2(t)$ as $\epsilon\to 0$.

Lemma 3: Except for the root at $t=0$, all roots of $R(t)$ lie in the left half-plane if and only if $\rho<1$.

The inverse Laplace transform of Eq. (23) yields

$$y_n(x) = \frac{1}{2\pi i}\int_\tau R^{-1}(t)\beta(t)\,e^{tx}\,dt$$
$$= \sum_{n=1}^{\infty}\frac{\beta(t_n)}{R'(t_n)}e^{t_nx},\quad x\geqslant 2, \tag{25}$$

where t_n is a root of $R(t)$ and the roots $\{t_n\}$ are arranged in decreasing order of real parts with complex conjugate roots arranged in any prescribed order. In principle the residue at $t=0$ determines the asymptotic behavior of $y_n(x)$ as $x\rightarrow\infty$. Since all roots of $R(t)$ are simple and have nonpositive real parts, it then follows that $y_n(x)$ must be bounded as $x\rightarrow\infty$.[18] But $\lim_{x\rightarrow\infty} G_n(x)=0$. By Eq. (16) we then have $\lim_{x\rightarrow\infty} y_n(x)=0$, i.e., y_n is an asymptotically stable solution. Again, by Theorem (6.10) of Bellman–Cooke, the generalized Fourier series given by Eq. (25) can be shown to be convergent for $x\geqslant 2$ and uniformly convergent over any finite interval for $x\geqslant 2$. By virtue of Lemmas 2 and 3 we can summarize our result as follows:

Theorem 4: Suppose $0<\rho<1$. The solution of Eq. (22) can be expressed as a convergent generalized Fourier series expansion given by Eq. (25), which becomes uniformly convergent over any finite interval for $x\geqslant 2$. Moreover, y_n is asymptotically stable, i.e., $\lim_{x\rightarrow\infty} y_n(x)=0$.

Thus we have completed our discussions of the perturbation solutions.

IV. THE PROBLEM OF PHASE TRANSITION

According to Lemmas 2 and 3, the solution of y_n given by Eq. (25) is no longer asymptotically stable if $\rho>1$. However, when $\rho=1$, the hard rods are in the closest contact and thus $\rho=1$ is the maximum attainable density for the one-dimensional system.[19] By Theorem 4 and the solutions of Eqs. (14)–(16), y_n is continuous and bounded on $(0,\infty]$. On the other hand, by the solutions of Eqs. (11)–(13), Q_n is continuous on $(0, l)$, except possibly a finite discontinuity at $x=0$ and $x=1$. Q_n is also bounded. But Q_n and y_n depend implicitly on n through A_n, $B_n,\ldots,G_n$ in Eqs. (11)–(16) and may increase as n increases. Unfortunately, it is impossible to determine the n dependence. In case of the square-well potential, it is possible to obtain analytical solutions of Q_n and y_n, from which we can examine the convergence of the perturbation series for Q and y. In view of the fact that

$$\left|\sum_{n=0}^{\infty}\frac{(\beta\xi)^n}{n!}Q_n(x)\right|\leqslant\sum_{n=0}^{\infty}\frac{(\beta\xi)^n}{n!}\sup|Q_n|,$$

and

$$\left|\sum_{n=0}^{\infty}\frac{(\beta\xi)^n}{n!}y_n(x)\right|\leqslant\sum_{n=0}^{\infty}\frac{(\beta\xi)^n}{n!}\sup|y_n|,$$

we can at least obtain the upper bound of $\sup|Q_n|$ and $\sup|y_n|$ for the absolute and uniform convergence of the perturbation series, that is, $\sup|Q_n|\leqslant n!$ and $\sup|y_n|\leqslant n!$ for large n.

In order to discuss the physical significance of our result, we next consider the compressibility equation

$$\beta^{-1}\left(\frac{\partial\rho}{\partial P}\right)_T=1+\rho\int_0^{\infty}h(x)\,dx.$$

It is well known that the critical point is characterized by the divergence of the isothermal compressibility $K_T=\rho^{-1}(\partial\rho/\partial P)_T$. The divergence of the integral $\int_0^{\infty}h(x)\,dx$ thus implies the occurrence of a phase transition.

Suppose the perturbation series for y is absolutely and uniformly convergent.[20] Then y is a continuous function of $\beta\xi$. On the other hand, y_n depends continuously on the density ρ (through the expression ρQ) except for the singularity at $\rho=1$. As $\rho\rightarrow 1$, the roots of $H(t)$ and $R(t)$ move toward the imaginary axis given by $e^t=1$. Since

$$h(x)=y(x)-1=y_0(x)-1+\sum_{n=1}^{\infty}\frac{(\beta\xi)^n}{n!}y_n(x),\quad x\geqslant l,$$

the convergence or divergence of the integral $\int_0^{\infty}h(x)\,dx$ is equivalent to the integral $\int_0^{\infty}[y(x)-1]\,dx$. By virtue of Lemma 3 and Theorem 4, the asymptotic behavior of y_n in Eq. (25) is determined by the term corresponding to the closest root t_1 to the origin in the lhp. Thus

$$|y_n(x)|\sim\left|\frac{\beta(t_1)}{R'(t_1)}\right||e^{t_1x}|,\quad x\gg 1.$$

This shows that $y_n\rightarrow 0$ exponentially. Hence $\int_0^{\infty}[y(x)-1]\,dx$ is convergent and the isothermal compressibility is a bounded continuous function of $\beta\xi$ and ρ for $0<\beta\xi<1$ and $0<\rho<1$. This in turn implies that there is no horizontal segment in the p-V diagram and consequently there is no phase transition.

Theorem 5: Suppose $0<\rho<1$ and $0<\beta\xi<1$. Then the pair distribution function $g(x)$ [or equivalently $y(x)$] obtained from the PY equation by the perturbation method is absolutely and uniformly convergent if and only if $\sup|y_n|\leqslant n!$ for large n. Furthermore, $y(x)-1\rightarrow 0$ exponentially so that the isothermal compressibility is finite, which implies that the PY equation cannot exhibit a phase transition.

Before we conclude our discussions, we briefly comment on the case as $l\rightarrow\infty$. It should be noted that Theorems 1 and 2 are strictly valid only for finite l, because the solution of Q_n in $[1, l-1]$ depends on the solution in $[l-1, l]$. But $Q_n(l)=0$. Thus the solution of $Q_n(x)$ on $[l-1, l]$ becomes the asymptotic condition $Q_n(x)\rightarrow 0$ as $l\rightarrow\infty$. The continuation method therefore cannot be applied. Also, the Laplace transform for the differential-difference equation of the advanced type diverges. This is essentially due to the fact that not all zeros of $H(t)$ lie in the lhp. Consequently, the result for $l\rightarrow\infty$ cannot emerge from our solutions.

ACKNOWLEDGMENTS

The author wishes to thank Dr. D. Henderson for an informative correspondence. He also wishes to thank the referee for an interesting comment.

[1] F. Gürsey, Proc. Cambridge Philos. Soc. **46**, 182 (1950).

[2] L. Van Hove, Physica **16**, 137 (1950).

[3] M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. **4**, 216, 229 (1963); **5**, 60 (1964).

[4] (a) J. L. Strecker, J. Math. Phys. **10**, 1541 (1969), (b) F. J. Dyson, Commun. Math. Phys. **12**, 91 (1969); **12**, 212 (1969); **21**, 269 (1971), (c) H. Falk. Phys. Rev. A **9**, 341 (1974); Physica **74**, 591 (1974).

[5] J. K. Percus and G. J. Yevick, Phys. Rev. **110**, 1 (1958).

[6] M. S. Wertheim, J. Math. Phys. **5**, 643 (1964).

[7] M. Chen, J. Math. Phys. **20**, 254 (1979).

[8] L. S. Ornstein and F. Zernike, Proc. Acad. Sci. Amsterdam **17**, 793 (1914).

[9] M. Chen, J. Math. Phys. **16**, 1150 (1975).

[10] R. J. Baxter, Phys. Rev. **154**, 170 (1967); Austral. J. Phys. **21**, 563 (1968).

[11] M. Chen, J. Math. Anal. Appl. **64**, 629 (1978).

[12] R. L. Bellman and K. L. Cooke, *Differential-Difference Equation* (Academic, New York, 1963), p. 408.

[13] See Ref. 12, p. 412 and p. 414.

[14] Suppose $Z = t_1 - \ln(t_1) = t_2 - \ln(t_2)$. Then $t_2/t_1 = e^{t_1}/e^{t_2}$. But this is impossible unless $t_1 = t_2$. Thus $T: t \rightarrow Z$ defined by $Z = t - \ln(t)$ is a one-to-one correspondence.

[15] See Ref. 12, p. 204.

[16] See Ref. 12, p. 59.

[17] See Ref. 12, p. 444.

[18] See Ref. 12, p. 190.

[19] The density ρ is defined as $\rho = N/L$, where N is the total number of the molecules and L is the length of the one-dimensional system, such that in the thermodynamic limit, $\rho = \lim_{\substack{N \to \infty \\ L \to \infty}} N/L < \infty$.

[20] Actually what we need is the uniform convergence rather than the absolute and uniform convergence. However, we have not been able to obtain a uniform bound. This has restricted the validity of $\beta\xi$ only to $(0, 1)$.

Conservation laws for shallow water waves on a sloping beach

Yilmaz Akyildiz[a)]
Department of Mathematical Sciences, University of Petroleum and Minerals, Dhahran, Saudi Arabia

(Received 20 November 1981; accepted for publication 12 March 1982)

Shallow water waves are governed by a pair of nonlinear partial differential equations. We transfer the associated homogeneous and nonhomogeneous systems (corresponding to constant and sloping depth, respectively) to the hodograph plane, where we find all the nonsimple wave solutions and construct infinitely many polynomial conversation laws. We also establish correspondence between conservation laws and hodograph solutions as well as Bäcklund transformations by using the linear nature of the problems on the hodograph plane.

PACS numbers: 92.10.Hm, 92.10.Sx, 02.30.Jr

I. INTRODUCTION

The linearity of a partial differential equation implies that any linear combination of solutions of the equation will also be a solution. This fundamental fact is also the main reason behind the method of separation of variables. In the event that a partial differential equation is nonlinear, this property is lost, and it becomes impossible to employ separation of variable techniques, or any other argument that depends on superpossibility. Another striking difference between linear and nonlinear partial differential equations is that, unlike linear p.d.e.'s, nonlinear equations often do not admit solutions which can be continuously extended wherever the differential equations themselves remain regular.

During the last decade, finding exact solutions to nonlinear differential equations has once more become important for both theoretical and practical purposes (*soliton theory*). It has been observed on some occasions (Korteweg–de Vries, sine–Gordon) that there are close connections between exact solutions, the existence of conservation laws, the inverse scattering method, and Bäcklund transformations. Such cases are called *completely integrable systems*. They come in association with some linear differential equations. In this article we shall obtain similar relations and properties in the case of the shallow water wave theory.

We were introduced to the area of water waves by Nutku's recent paper.[1] Shallow water waves are governed by a system of two nonlinear partial differential equations, which can also be written in the form of two conservation laws. First, we try to find further conservation laws by using the method of Wahlquist and Estabrook.[2] For the homogeneous case (corresponding to constant depth) we are able to construct an infinite family of conservation equations. This leads us to search for the exact solutions. It was at this point that we learned that these results were already known to Whitham.[3] We pass to the hodograph plane where we catch the linear system of equations associated with our nonlinear problem. On this plane we show that conservation laws are easily derivable. On the hodograph plane we obtain all the solutions, except simple waves, by potentials which also satisfy linear equations. These potentials are, in fact, the Legendre transforms of the ones introduced by Nutku. Via these potentials we are also able to construct a correspondence between conservation laws and nonsimple wave solutions of the homogeneous problem.

Finally, we take up the nonhomogenous case corresponding to a sloping beach. By using the polynomial conservation laws of the related homogeneous problem, we construct an infinite family of polynomial conservation laws for the nonhomogeneous case. By using the solutions of the cylindrical wave equation, we also indicate how one can construct auto-Bäcklund and Bäcklund transformations for these homogeneous and nonhomogeneous problems.

II. METHOD OF ESTABROOK AND WAHLQUIST

We consider the following system of two homogeneous first-order quasilinear equations:

$$u_t + uu_x + 2cc_x = 0, \tag{1a}$$

$$c_t + uc_x + \tfrac{1}{2}cu_x 0, \tag{1b}$$

representing shallow water waves, the bottom of the ocean being horizontal.[4] $u(x, t)$ and $c(x, t)$ are the velocities of the fluid and of the disturbance with respect to the fluid, respectively. Subscripts denote partial derivatives.

First we shall apply the techniques of Wahlquist and Estabrook[2] (Sec. III) to the system (1) above to find all the conservation laws, which are used to obtain potentials in their paper.

In the four-dimensional space of all the independent and dependent variables, $\{x, t, u, c\}$, the set of first-order differential equations (1) above can be expressed by the following pair of differential 2-forms[1]:

$$\alpha_1 = du \wedge dx - udu \wedge dt - 2cdc \wedge dt, \tag{2a}$$

$$\alpha_2 = 2cdc \wedge dx - 2cudc \wedge dt - c^2du \wedge dt. \tag{2b}$$

Any regular (differentiable) solution (u, c) of (1) will *annul* this set of forms. Since $d\alpha_i = 0$, $i = 1, 2$, the ideal generated by α_1 and α_2 is closed, and one can, therefore, apply Cartan's theory.

Conservation laws correspond to the existence of *exact* 2-forms contained in the ring of α_i. Let us try to find all the 2-forms

$$\beta = f\alpha_1 + g\alpha_2, \tag{3}$$

satisfying $d\beta = 0$, the condition for exactness. This is the (local) integrability condition for the existence of a 1-form, say ω, such that

$$\beta = d\omega. \tag{4}$$

The following treatment is restricted in that we do not allow

[a)] On leave from the Middle East Technical University, Ankara, Turkey.

f and g to be explicit functions of the independent variables x and t. This seems plausible since the system (1) itself has no explicit (x, t) dependence:

$$\begin{aligned} d\beta &= (f_u du + f_c dc) \wedge \alpha_1 + (g_u du + g_c dc) \wedge \alpha_2 \\ &= (2cg_u - f_c) du \wedge dc \wedge dx \\ &\quad + (c^2 g_c - 2cug_u + uf_c - 2cf_u) du \wedge dc \wedge dt. \end{aligned}$$

Hence, $d\beta = 0$ implies

$$f_c = 2cg_u, \tag{5a}$$

$$2f_u = cg_c, \tag{5b}$$

$$\begin{aligned} \beta = d\omega &= f\alpha_1 + g\alpha_2 \\ &= f du \wedge dx + 2cg dc \wedge dx - (uf + c^2 g) \\ &\qquad \times du \wedge dt - 2c(f + ug) dc \wedge dt, \end{aligned}$$

which, with the help of (5), integrates to

$$\omega = dx \cdot \textstyle\int f\, \partial u - dt \cdot \int (uf + c^2 g)\, \partial u. \tag{6}$$

Since $d\omega$ lies in a closed ideal of differential forms, the "Frobenius theorem" applies: Any local solution which annuls the ideal must also annul ω. This, in turn, gives us the following conservation equation:

$$F_t + G_x = 0, \tag{7}$$

where

$$F = \textstyle\int f\, \partial u \quad \text{with} \quad F_c = 2cg, \tag{8a}$$

and

$$G = \textstyle\int (uf + c^2 g)\, \partial u \quad \text{with} \quad G_c = 2c(f + ug). \tag{8b}$$

When the condition (7) is satisfied, we shall say that the pair (F, G) forms a *conservation law*. If $G = 0$ at $x = 0$ and ∞, we obtain the corresponding conserved quantity $\int_0^\infty F\, dx$.

Since the system of equations (1) is quasilinear (i.e., linear in the derivatives) with polynomial coefficients in u and c, the most interesting conservation equations are polynomial in u and c. They may be obtained consistently from (5) and (8) by taking

$$f = \sum_{i=0}^{n} p_i(u) c^{2i}, \quad g = \sum_{j=0}^{n} q_j(u) c^{2j},$$

from which it follows that

$$p_0' = 0, \quad q_n' = 0,$$

$$mp_m = q_{m-1}', \quad p_m' = mq_m, \quad m = 1, 2, ..., n.$$

(It can easily be checked that the odd powers of c do not survive.) We list the first few of these polynomials:

f	g	F	G
1	0	u	$\frac{1}{2}u^2 + c^2$
0	1	c^2	uc^2
c^2	u	uc^2	$u^2c^2 + \frac{1}{2}c^4$
uc^2	$\frac{1}{2}u^2 + c^2$	$\frac{1}{2}u^2c^2 + \frac{1}{2}c^4$	$\frac{1}{2}u^3c^2 + uc^4$
$u^2c^2 + c^4$	$\frac{1}{3}u^3 + 2uc^2$	$\frac{1}{3}u^3c^2 + uc^4$	$\frac{1}{3}u^4c^2 + \frac{1}{2}3u^2c^4 + \frac{1}{3}c^6$
$\frac{2}{3}u^3c^2 + 2uc^4$	$\frac{1}{6}u^4 + 2u^2c^2 + c^4$	$\frac{1}{6}u^4c^2 + u^2c^4 + \frac{1}{3}c^6$	$\frac{1}{6}u^5c^2 + \frac{4}{3}u^3c^4 + uc^6$

It is interesting to note that F and G are homogeneous in u and c. This observation immediately makes us think of our Russian colleagues who have extracted the algebro–geometric structures of some of the "completely integrable" evolution equations.[5,6] For the boundary conditions $u = 0$ and $c = 0$ at $x = 0$ and ∞, we obtain infinitely many conserved quantities by integrating F's with respect to x from 0 to ∞.

Differentiating (5a) partially with respect to u and (5b) with respect to c and subtracting, we find

$$4g_{uu} = g_{cc} + g_c/c, \tag{9}$$

or

$$4g_{uu} = (1/c)(cg_c)_c. \tag{10}$$

Thus, we have the cylindrical wave equation for $g(u,c)$. This is a linear equation for g which can be solved by standard methods. Similarly, for $f(u, c)$ we have

$$4f_{uu} = f_{cc} - f_c/c, \tag{11}$$

or

$$4f_{uu} = c(f_c/c)_c.$$

On the other hand, upon eliminating f and g from the set of equations (8), we arrive at the following relations:

$$G_u = uF_u + \tfrac{1}{2}cF_c, \tag{12a}$$

$$G_c = 2cF_u + uF_c. \tag{12b}$$

As before, differentiating the first equation in (12) partially with respect to c, the second equation with respect to u, and subtracting, we obtain

$$4F_{uu} = F_{cc} - F_c/c. \tag{13}$$

Unfortunately (maybe fortunately), we don't have a nice equation for G.

We make the following observation: Even though x and t are the independent variables, all our expressions are (linear) partial differential equations in the variables u and c. This is because we have no (x, t) dependence in the system of equations (1) with which we started. This suggests that we should interchange the roles of the dependent and independent variables. This is called the "*hodograph*" *method*, which we will take up in the following section.

III. METHOD OF HODOGRAPH TRANSFORMATION

We consider the system (1) which has no explicit (x, t) dependence. For any region where the Jacobian

$$J = u_x c_t - u_t c_x$$

is nonzero, the system (1) can be transformed into an equivalent linear system by interchanging the roles of dependent and independent variables. If $J \neq 0$ for a solution $u(x,t)$, $c(x,t)$ of (1), we may consider x and t as functions of u and c. From

$$u_x = Jt_c, \quad u_t = -Jx_c, \tag{14a}$$

$$c_x = -Jt_u, \quad c_t = Jx_u, \tag{14b}$$

we see that the highly nonlinear factor J cancels out in (1) and that $x(u,c)$ and $t(u,c)$ satisfy the linear differential equations

$$x_u = ut_u - \tfrac{1}{2}ct_c, \tag{15a}$$

$$x_c = -2ct_u + ut_c. \tag{15b}$$

By eliminating x we obtain the linear equation

$$4t_{uu} = t_{cc} + (3/c)t_c, \tag{16}$$

which can be solved by standard methods. This can be further simplified by introducing the transformation

$$t = s_c/c. \tag{17}$$

We obtain the cylindrical wave equation

$$4s_{uu} = s_{cc} + s_c/c, \tag{18}$$

whose solutions involve Bessel functions. We remark that g and s satisfy the same equation.

The described transformation of the (x,t) plane into the (u,c) plane is called a *hodograph transformation*. Since the possibility of this reduction depends essentially on the assumption $J \neq 0$, solutions for which $J = 0$ cannot be obtained by the hodograph method. These solutions are called *simple waves* and they are important tools for the solutions of flow problems (Courant and Friedrichs,[7] Sec. 29). Wave breaking occurs when $J = 0$ corresponding to the multivaluedness, i.e., shock waves. We notice that the solution

$$u = \tfrac{2}{3}x/t, \quad c = \tfrac{1}{3}x/t,$$

given by Nutku[1] represents a simple wave. So, we could not possibly obtain this solution by the hodograph method.

We would like to mention that in the set of all solutions the simple waves form a set of measure zero. But this is not to say simple waves are unimportant.

Just to show how natural it is to work in the hodograph plane, we shall rederive the conservation equation (12). In the (u,c) plane Eq. (7) becomes

$$\begin{aligned} F_t + G_x &= F_u u_t + F_c c_t + G_u u_x + G_c c_x \\ &= -Jx_c F_u + Jx_u F_c + Jt_c G_u - Jt_u G_c = 0. \end{aligned}$$

Above, we have employed Eq. (14). Again, the nonlinear factor J cancels out, and we arrive at

$$x_u F_c - x_c F_u = t_u G_c - t_c G_u, \tag{19}$$

or

$$dx \wedge dF = dt \wedge dG. \tag{20}$$

Upon using (15), (19) becomes

$$t_c(G_u - uF_u - \tfrac{1}{2}cF_c) = t_u(G_c - 2cF_u - uF_c). \tag{21}$$

Since this is to be an identity, the coefficients of the derivatives must vanish separately:

$$G_u = uF_u + \tfrac{1}{2}cF_c, \quad G_c = 2cF_u + uF_c.$$

These are the same as (12). We note that the computation above is somewhat shorter than the Wahlquist–Estabrook method used in the previous section to establish these equations. Whitham[3] has an even simpler way of deriving them. Even so, we have included the method of Wahlquist and Estabrook because it has provided us with two nice functions—f and g—which we make use of in this paper.

IV. POTENTIALS

We look for potentials in the hodograph plane. The system of equations (15) can be rewritten in the following equivalent form:

$$(2cx - 2cut)_u = -(c^2 t)_c, \tag{22a}$$

$$(x - ut)_c = -(2ct)_u. \tag{22b}$$

These, in return, suggest the existence of potentials $\Psi(u,c)$ and $\Phi(u,c)$, satisfying

$$\Psi_u = -c^2 t, \quad \Psi_c = 2cx - 2cut, \tag{23a}$$

and

$$\Phi_u = x - ut, \quad \Phi_c = -2ct. \tag{23b}$$

Ψ, Φ are, in fact, the Legendre transforms of the potentials introduced by Nutku.[1] Solving them for x and t, we obtain

$$x = \Psi_c/2c - (u/c^2)\Psi_u, \quad t = -\Psi_u/c^2 \tag{24a}$$

and

$$x = \Phi_u - (u/2c)\Phi_c, \quad t = -\Phi_c/2c. \tag{24b}$$

Hence, if we know $\Psi(u,c)$ or $\Phi(u,c)$, by using these formulas we can compute x and t. Combining (24a) with (22b), we obtain

$$4\Psi_{uu} = \Psi_{cc} - \Psi_c/c. \tag{25}$$

Equation (24b) together with (22a) gives

$$4\Phi_{uu} = \Phi_{cc} + \Phi_c/c. \tag{26}$$

Unlike their Legendre transforms, Ψ and Φ satisfy linear equations.

Comparing (9), (11), (13), (18), (25), and (26), our readers realize that we keep encountering the following set of equations:

$$4\chi_{uu} = \chi_{cc} \pm \chi_c/c. \tag{27}$$

In the next section too we shall encounter these equations when we are dealing with a related nonhomogeneous problem. Not only can we derive the conservation laws from the solutions of (27), but we can also construct all the hodograph solutions of the original system of equations with which we started. In this way, we are able to construct a solution of the system of equations (1) from a given conservation law by letting $\Psi = F$ and by using (24a). We can reverse this process for non-simple wave solutions. Now we have an infinite family of solutions associated with the list of polynomial conservation laws listed in Sec. II. Here we list the first few of these special solutions:

F	x	t	u	c^2
u	u/c^2	$1/c^2$	x/t	$1/t$
c^2	1	0	—	—
uc^2	0	1	—	—
$\frac{1}{2}u^2c^2 + \frac{1}{2}c^4$	$-\frac{1}{2}u^2 + c^2$	$-u$	$-t$	$x + \frac{1}{2}t^2$
$\frac{1}{3}u^3c^2 + uc^4$	$\frac{1}{3}2u^3 - uc^2$	$u^2 + c^2$	implicit solution	
$\frac{1}{6}u^4c^2 + u^2c^4 + \frac{1}{3}c^6$	$\frac{1}{2}u^4 - c^4$	$\frac{2}{3}u^3 + 2uc^2$	implicit solution	

V. CASE OF SLOPING BEACH

We consider the following nonhomogeneous system of equations:

$$u_t + uu_x + 2cc_x = g\beta, \tag{28a}$$

$$c_t + uc_x + \frac{1}{2}cu_x = 0, \tag{28b}$$

representing shallow water waves on a sloping beach. The constant term $g\beta$ involves the gravitational constant g and the slope of the bottom β.

In his Tata Institute Notes,[8] Whitham absorbs the nonhomogeneous term $g\beta$ in a conservation form as

$$(u - g\beta t)_t + (\tfrac{1}{2}u^2 + c^2)_x = 0, \tag{29}$$

and adds the following statement: "But this comment does not appear to lead any further." However, by the means of (29) we were fortunate in finding ourselves able to construct conservation laws in the form

$$\Big[F(u,c) - \sum_{i=1}^{m} \frac{1}{i!}(g\beta t)^i P_i(u,c)\Big]_t + \Big[G(u,c) - \sum_{i=1}^{m} \frac{1}{i!}(g\beta t)^i Q_i(u,c)\Big]_x = 0, \tag{30}$$

for the nonhomogeneous system (28) above. We will denote the contents of the two square brackets in (30) as $\tilde{F}$ and $\tilde{G}$, respectively. As one can guess, we shall require (F, G) to form a conservation law for the related homogeneous system (1). Hence, as in (12) of Sec. II, they satisfy the following linear system of equations:

$$G_u = uF_u + \tfrac{1}{2}cF_c, \tag{31a}$$

$$G_c = 2cF_u + uF_c, \tag{31b}$$

whose integrability condition is (13):

$$4F_{uu} = F_{cc} - F_c/c. \tag{32}$$

With the help of (31), (30) simplifies to

$$g\beta\cdot F_u - \sum_{i=1}^{m} \frac{1}{i!}(g\beta t)^i [g\beta\cdot P_{iu} + u_x(Q_{iu} - uP_{iu} - \tfrac{1}{2}cP_{ic}) + c_x(Q_{ic} - 2cP_{iu} - uP_{ic})] - g\beta\cdot \sum_{i=1}^{m} \frac{1}{(i-1)!}(g\beta t)^{i-1} P_i = 0.$$

Imposing the following further conditions,

$$Q_{iu} = uP_{iu} + \tfrac{1}{2}cP_{ic}, \tag{33a}$$

$$Q_{ic} = 2cP_{iu} + uP_{ic}, \tag{33b}$$

forces us to take

$$P_1 = F_u,$$

$$P_2 = -P_{1u} = -F_{uu},$$

$$P_3 = -P_{2u} = +F_{uuu},$$

$$\vdots$$

$$P_m = -P_{m-1,u} = (-1)^{m+1} F_{\underbrace{uu\cdots u}_{m \text{ times}}},$$

and

$$P_{mu} = 0.$$

For convenience, we shall use the notation

$$F_u^{(m)} = F_{\underbrace{uu\cdots u}_{m \text{ times}}}.$$

The last condition requires us to take $F_u^{(m+1)} = 0$, which can automatically be satisfied for a suitable F, if we start with polynomial conservation laws for the related homogeneous problem. What really makes this construction work is the fact that all the P_i turn out to be $\pm F_u^{(i)}$ and the pairs (P_i, Q_i) and (F,G) satisfy the same system of equations. Since the compatibility equation (32) is also satisfied by the u derivatives $F_u^{(i)}$, the integrability condition of the system (33) is automatically guaranteed. We have, therefore, a consistent method, and by using the list in Sec. II, we can construct an infinite family of conservation laws for the nonhomogeneous system (28).

The computations for Q_i become easier once one realizes that the ith u derivative of F on the jth line in the list is proportional to F on the $(j - i)$th line in the same list in Sec. II (excluding the first line).

Here we list the first few of these conservation laws $(\tilde{F}, \tilde{G})$:

$\tilde{F}$	$\tilde{G}$
$u - g\beta t$	$\frac{1}{2}u^2 + c^2$
c^2	uc^2
$uc^2 - g\beta tc^2$	$u^2c^2 + \frac{1}{2}c^4 - g\beta tuc^2$
$\frac{1}{2}u^2c^2 + \frac{1}{2}c^4 - g\beta tuc^2 + \frac{1}{2}(g\beta t)^2c^2$	$\frac{1}{2}u^3c^2 + uc^4 - g\beta t\,(u^2c^2 + \frac{1}{2}c^4) + \frac{1}{2}(g\beta t)^2uc^2$
$\frac{1}{3}u^3c^2 + uc^4 - 2g\beta t\,(\frac{1}{2}u^2c^2 + \frac{1}{2}c^4)$	$\frac{1}{3}u^4c^2 + \frac{1}{2}3u^2c^4 + \frac{1}{3}c^6 - 2g\beta t\,(\frac{1}{2}u^3c^2 + uc^4)$
$\quad + (g\beta t)^2uc^2 - \frac{1}{3}(g\beta t)^3c^2$	$\quad + (g\beta t)^2(u^2c^2 + \frac{1}{2}c^4) - \frac{1}{3}(g\beta t)^3uc^2$
$\frac{1}{6}u^4c^2 + u^2c^4 + \frac{1}{3}c^6 - 2g\beta t\,(\frac{1}{3}u^3c^2 + uc^4)$	$\frac{1}{6}u^5c^2 + \frac{1}{3}4u^3c^4 + uc^6 - 2g\beta t\,(\frac{1}{3}u^4c^2 + \frac{1}{2}3u^2c^4 + \frac{1}{3}c^6)$
$\quad + 2(g\beta t)^2(\frac{1}{2}u^2c^2 + \frac{1}{2}c^4) - \frac{2}{3}(g\beta t)^3uc^2 + \frac{1}{6}(g\beta t)^4c^2$	$\quad + 2(g\beta t)^2(\frac{1}{2}u^3c^2 + uc^4) - \frac{2}{3}(g\beta t)^3(u^2c^2 + \frac{1}{2}c^4) + \frac{1}{6}(g\beta t)^4uc^2$

We note that $\tilde{F}$ and $\tilde{G}$ are homogeneous in u, c, and t. We would like to thank Dr. Mirie for drawing our attention to the fact that the terms containing t can be put into the form $(u - g\beta t)^i$. But, as is clear even from the first line of the above list, we cannot completely eliminate all the u's in $\tilde{G}$'s, although we can write $\tilde{F}$'s in terms of $v = u - g\beta t$ and c only. Nevertheless, v's do not show up separately in $\tilde{G}$'s; they all come multiplied with u's or c's. Hence, for the boundary conditions $u = 0, c = 0$ at $x = 0$ and ∞, we still obtain infinitely many conserved quantities by integrating $\tilde{F}$'s with respect to x from 0 to ∞ (cf. the homogeneous case).

Having constructed an infinite number of conservation laws, one might, therefore, expect to be able to find the solution of the nonhomogeneous system (28) analytically. Indeed, as we have learned from Whitham,[8] Carrier and Greenspan introduced new variables suggested by the characteristic forms of these equations and applied a hodograph transformation to them and obtained

$$g\beta x = -\frac{\phi_\lambda}{2} + \frac{\phi_\sigma^2}{2\sigma^2} + \frac{\sigma^2}{16}, \tag{34a}$$

$$g\beta t = \frac{\lambda}{4} - \frac{\phi_\sigma}{\sigma}, \tag{34b}$$

where $\sigma = 4c, \lambda = -4(u - g\beta t)$, and ϕ satisfies the cylindrical wave equation

$$4\phi_{\lambda\lambda} = \phi_{\sigma\sigma} + \phi_\sigma/\sigma. \tag{35}$$

We observe that ϕ in (35) and Φ in (26) satisfy the same kind of equation. Hence, after the necessary relabelling of the variables, a solution of (35) can be used to generate a solution of either of the problems: homogeneous [via (24b) and nonhomogeneous [via (34)]. In this way, we find a correspondence between the nonsimple wave solutions of the two systems which we have considered in this paper. In a way, this corespondence can be thought of as a *Bäcklund transformation* between the homogeneous and nonhomogeneous problems (1) and (28). By using the linearity of the space of solutions of (35) we can also construct auto-Bäcklund transformations for each of these problems.

We leave it to our interested readers to construct the solutions of the nonhomogeneous problem which corresponds to the solutions listed in Sec. IV of the homogeneous system.

To us, the story of this paper looks similar to the hydrogen atom problem, (the invariance group being the space of solutions of the cylindrical wave equation). We expect to shed more light on this subject by using the orbit theory picture of Krillov, Kostant, and Souriau. This is our forthcoming project.

ACKNOWLEDGMENTS

I would like to apologize to Ayshe/Ruth for the sleepless Arabian Nights and to express my appreciation of the UPM Beach where I meet the real waves and enjoy them. Also, special thanks to R. O. Wells, Jr., and G. Bluman for their encouraging comments.

[1] Y. Nutku, "Potentials for Non-linear Shallow Water Waves," preprint (1981).

[2] H. D. Wahlquist and F. B. Estabrook, "Prolongation structures of nonlinear evolution equations," J. Math. Phys. **16**, 1 (1975).

[3] G. B. Whitham, *Linear and Non-linear Waves* (Wiley, New York, 1974).

[4] J. J. Stoker, *Water Waves* (Interscience, New York, 1957).

[5] V. G. Drinfeld, I. M. Krichever, J. I. Manin, and S. P. Novikov, "Methods of Algebraic Geometry in Contemporary Mathematical Physics," *Soviet Science Reviews, Phys. Reviews, 1978* (Over. Pub. Assoc., Amsterdam, 1980).

[6] I. M. Krichever, "The Method of Algebraic Geometry in the Theory of Non-linear Equations," Usp. Math. Nauk **32**, 183–208 (1977).

[7] R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves* (Springer-Verlag, New York, 1972).

[8] G. B. Whitham, *Lectures on Wave Propogation* (Tata Institute, Bombay, 1979), p. 70.

On approximating the solutions of the Chandrasekhar H-equation

S. R. Vatsya
Center for Research in Experimental Space Science, York University, Downsview, Ontario M3J 1P3, Canada

(Received 18 September 1981; accepted for publication 23 April 1982)

Method of moment and the Newton approximants to the Chandrasekhar H-function are shown to converge to the physical solution uniformly. The latter also converge monotonically faster than the iterative approximations obtained previously.

PACS numbers: 95.30.Jx, 02.60. + y

1. INTRODUCTION

Convergence properties of various methods to approximate the "physical" solution of the nonlinear integral, Chandrasekhar H-equation

$$H(z) = 1 + zH(z)\int_0^1 \frac{d\sigma(x)}{z+x} H(x). \tag{1}$$

have been studied recently by several authors.[1–4] The approximations H_n, K_n obtained by solving

$$H_{n+1}(z) = 1 + zH_n(z)\int_0^1 \frac{d\sigma(x)}{z+x} H_n(x) \tag{2}$$

and

$$K_{n+1}(z) = 1 + zK_{n+1}(z)\int_0^1 \frac{d\sigma(x)}{z+x} K_n(x), \tag{3}$$

$$n \geqslant 0, \quad H_0 = K_0 = 0,$$

have been shown to converge uniformly with respect to z, and monotonically from below to the physical solution H.[1–3] Also $\{K_n\}$ converges faster than does $\{H_n\}$.[3] These results were proved assuming that $d\sigma(x)/dx \geqslant 0$ and that

$$C_0(\sigma) = \int_0^1 d\sigma(x) < \frac{1}{2}.$$

However the assumption of differentiability is unnecessary and it is sufficient to assume that $\sigma(x)$ is nondecreasing. In Ref. (4) the method of moment approximants, denoted here by $\hat{H}_n$, were introduced as the solutions of

$$\hat{H}_n(z) = 1 + z\hat{H}_n(z)\int_0^1 \frac{d\sigma_n(x)}{z+x} \hat{H}_n(x), \tag{4}$$

where $\sigma_n(x)$ is the approximation to $\sigma(x)$ obtained by solving the truncated moment problem of order $(2n-1)$. Assuming further that $\sigma(x)$ is continuous at a dense set of points in $[0,1]$ including the end points, it was shown that $H_n(z) \underset{n\to\infty}{\longrightarrow} H(z)$ uniformly for z in $[0,\infty)$, provided that $C_0(\sigma) < \frac{4}{9}$.

In the present note we improve the result of Ref. (4) to include each $C_0(\sigma) < \frac{1}{2}$. Also, we show that the sequence obtained by the Newton method converges to H monotonically and uniformly with respect to z, faster than does $\{K_n\}$ and hence $\{H_n\}$, under the same assumptions. Since the cases $C_0(\sigma) \geqslant \frac{1}{2}$ are reducible to the case $C_0(\sigma) < \frac{1}{2}$ by some simple transformations,[1] these results are quite satisfactory from the physical view point.

2. PRELIMINARIES

Let $L^1(\mu)$ be the real Banach space of absolutely μ-integrable functions on [0,1] where μ is a nonnegative measure with $\mu\{0\} = 0$. In the sequel μ will be induced by some nondecreasing function $\mu(x)$ and the integration will be assumed to be in the Lebesgue–Stieltjes sense. The norm of the vectors in $L^1(\mu)$ and the operators from $L^1(\mu)$ to $L^1(\mu)$ will be denoted by $\|\cdot\|^\mu$. Consider the nonlinear operator A_μ defined by

$$A_\mu u = 1 + uB_\mu u,$$

where u is the operation of multiplication by $u(x)$ and

$$(B_\mu u)(z) = \int_0^1 d\mu(x) \frac{z}{z+x} u(x).$$

It is clear that $(B_\mu u)(z)$ is well defined for z in $[0,\infty)$. We have

Lemma 1: The Fréchet differential $A'_\mu(u)$ of A_μ at $u \in L^1(\mu)$ exists with $\|A'_\mu(u)\|^\mu \leqslant \|u\|^\mu$.

Proof: It is straightforward to check that

$$(A'_\mu(u)h)(z) = \int_0^1 d\mu(x) \frac{z}{z+x} \alpha(z,x), \quad h \in L^1(\mu),$$

where $\alpha(z,x) = [h(z)u(x) + h(x)u(z)] = \alpha(x,z)$. It follows that

$$\begin{aligned}
&\|A'_\mu(u)h\|^\mu \\
&\quad \leqslant \int_0^1 d\mu(z)\int_0^1 d\mu(x) \frac{z}{z+x} |\alpha(z,x)| \\
&\quad = \frac{1}{2}\left[\int_0^1 d\mu(z)\int_0^1 d\mu(x)|\alpha(z,x)|\right. \\
&\qquad \left. - \int_0^1 d\mu(z)\int_0^1 d\mu(x) \frac{x-z}{x+z} |\alpha(z,x)|\right] \\
&\quad = \frac{1}{2}\int_0^1 d\mu(z)\int_0^1 d\mu(x)|\alpha(z,x)| \\
&\quad \leqslant \|u\|^\mu \|h\|^\mu,
\end{aligned}$$

implying the result. The interchange of the order of integration is justified by Fubini's theorem.

The results of Lemma 2 follow by straightforward substitutions. Therefore we state them without proofs.

Lemma 2: Let $u, h \in L^1(\mu)$.

(i) For $u \geqslant 0$, $B_\mu u \geqslant 0$, $A_\mu u \geqslant 1$ for each $z \geqslant 0$;

(ii) for $u, h \geqslant 0$, $A'_\mu(u)h = hB_\mu u + uB_\mu h \geqslant 0$;

(iii) $\|A_\mu u\|^\mu \leqslant C_0(\mu) + \frac{1}{2}(\|u\|^\mu)^2$;

(iv) $|(B_\mu u)(\mathrm{z})| \leqslant \|u\|^\mu$ for $z \in [0,\infty)$.

It may be remarked that some of the conditions in Lemma 2 may be weakened to be valid a.e.

It is clear from Lemma 1 that if $C_0(\mu) < \frac{1}{2}$ and $\|u\|^\mu \leqslant [1 - \surd 1 - 2C_0(\mu)] = d(\mu)$, then $\|A'_\mu(\mu)\| \leqslant d(\mu) < 1$. Also, from Lemma 2(iii) it follows that if $\|u\|^\mu \leqslant d(\mu)$ then $\|A_\mu u\|^\mu \leqslant d(\mu)$. This means that A_μ is a contraction of the closed ball of radius $d(\mu)$ in $L^1(\mu)$ implying the existence of a unique solution H_μ of $H_\mu = A_\mu H_\mu$ in the ball i.e.,

0022-2488/82/091728-04$02.50

$\|H_\mu\|^\mu \leqslant d(\mu)$. This implies that $1 \leqslant H_\mu(z) \leqslant (1-2C_0(\mu))^{-1/2}$ for all $z \geqslant 0$. Furthermore $\{H_\mu^n\}$ defined by $H_\mu^{n+1} = A_\mu H_\mu^n$, $n \geqslant 0$, with H_μ^0 being an arbitrary vector in the ball, converges to H_μ. Monotonicity and the bound property of the sequence result from a more careful choice of H_μ^0. These results have been obtained previously by a slightly different but equivalent approach.[1] Also the fact that $\|H_\sigma\|^\sigma = \|H_\sigma\| < 1$ was found to be sufficient to establish that $H_\sigma = H$, i.e., the physical solution.[1] The continuity of H_μ is obvious. We state these results for later reference.

Corollary 1: Let H_μ be as above and $C_0(\mu) < \frac{1}{2}$. Then H_μ is continuous with $\|H_\mu\|^\mu \leqslant d(\mu)$ and $1 \leqslant H_\mu(z) \leqslant (1-2C_0(\mu))^{-1/2}$ for $z \in [0,\infty)$.

It is obvious that $H(z)$ is positive for $z \geqslant 0$. In Proposition 1 we characterize $H(z)$ to be the minimal positive fixed point of $A_\sigma = A$ for $z \geqslant 0$.

Proposition 1:

(i) Let $z \in [0,1]$ and $H(z)$ be as above; then $H(z) \leqslant \widetilde{H}(z)$ where $\widetilde{H}(z) \geqslant 0$ is a fixed point of A.

(ii) The statement of (i) is valid with $z \in [0,\infty)$.

Proof:

(i) Let $H_0 = 0$ and $H_{n+1} = AH_n$. Then $0 \leqslant H_n \uparrow H$.[1] Since $\widetilde{H}(z) \geqslant 0$, $H_0 \leqslant \widetilde{H}$. Assume that $H_n \leqslant \widetilde{H}$. It follows that

$$\begin{aligned}\widetilde{H} - H_{n+1} &= A\widetilde{H} - AH_n \\ &= \int_0^1 dt\, A'[H_n + t(\widetilde{H} - H_n)](\widetilde{H} - H_n),\end{aligned}$$

where the integration is understood in the Riemann sense.[5] By assumption, $(\widetilde{H} - H_n) \geqslant 0$, implying that for $0 \leqslant t \leqslant 1$, $[H_n + t(\widetilde{H} - H_n)] \geqslant 0$. Hence, from Lemma 2 (ii), $(H - H_{n+1}) \geqslant 0$. It follows that $H = \lim_{n\to\infty} H_n \leqslant \widetilde{H}$.

(ii) It is clear that

$$\begin{aligned}\widetilde{H} - H &= (\widetilde{H} - H)BH + \widetilde{H}B(\widetilde{H} - H) \\ &= [1 - BH]^{-1}\widetilde{H}B(\widetilde{H} - H).\end{aligned}$$

Since, from (i), $\widetilde{H} - H \geqslant 0$ on $[0,1]$, $B(\widetilde{H} - H) \geqslant 0$ on $[0,\infty)$ (Lemma 2(i)). Also, $0 \leqslant BH \leqslant \|H\| < 1$ on $[0,\infty)$. Thus if $\widetilde{H} \geqslant 0$ on $[0,\infty)$, $\widetilde{H} - H \geqslant 0$ there.

3. CONVERGENCE OF THE MOMENT METHOD APPROXIMANTS

Let $\sigma(x)$ be nondecreasing and continuous at a dense set of points in $[0,1]$ including zero and one; and let $\sigma_n(x)$ be the moment approximation of order $(2n-1)$ to $\sigma(x)$. The nondecreasing, discontinuous function $\sigma_n(x)$ is determined within a constant by

$$\int_0^1 x^m(d\sigma(x) - d\sigma_n(x)) = 0, \quad m = 0,1,\ldots, 2n-1.$$

Continuity of $\sigma(x)$ at zero implies that $\{0\}$ is of measure zero with respect to σ, σ_n. In addition to the abbreviations $A_\sigma = A$, $B_\sigma = B$, $\|\cdot\|^\sigma = \|\cdot\|$ we shall use $A_{\sigma_n} = A_n$, $B_{\sigma_n} = B_n$, $\|\cdot\|^{\sigma_n} = \|\cdot\|^n$ for the sake of convenience of writing. The approximant in question here is the minimal positive fixed point $\widehat{H}_n$ of A_n. It will be assumed that $C_0(\sigma) = C_0 < \frac{1}{2}$. Since $C_0(\sigma_n) = C_0$, it follows from Corollary 1 by setting $\mu = \sigma, \sigma_n$, that $\|H\|, \|\widehat{H}_n\|^n \leqslant [1 - \sqrt{1 - 2C_0}]$, $1 \leqslant H_n, H \leqslant (1-2C_0)^{-1/2}$ for each n and that H, $\widehat{H}_n$ are continuous.

Lemma 3: Let $|w| \leqslant M$, $z \geqslant 0$; then

$$\lim_{z\to 0}(Bw)(z) = \lim_{z\to 0}(B_n w)(z) = 0.$$

Proof: We consider the case of B; the case of B_n follows by the same argument. By definition

$$(Bw)(z) = \int_0^1 d\sigma(x)\,\frac{z}{z+x}\,w(x).$$

Since the integrand is bounded by a σ-integrable function M and converges to zero for $x > 0$, i.e., for a.e. x, the result follows by the Lebesgue dominated convergence theorem.

Lemma 4: Let $|u|$, $|u_n| \leqslant M$, u be continuous on $[0,1]$ and $\|u - u_n\|^n \xrightarrow[n\to\infty]{} 0$.
Then

$$v_n(z) = (B_n u_n)(z) \xrightarrow[n\to\infty]{} (Bu)(z) = v(z).$$

uniformly with respect to z in any compact subset of $[0,\infty)$.

Proof: From Lemma 3, $v(0) = v_n(0) = 0$. Let $z > 0$. We have that

$$\begin{aligned}|v(z) - v_n(z)| \leqslant& \left|\int_0^1 (d\sigma(x) - d\sigma_n(x))\,\frac{z}{z+x}\,u(x)\right| \\ &+ \left|\int_0^1 d\sigma_n(x)\,\frac{z}{z+x}\,(u - u_n)(x)\right|.\end{aligned}$$

Since $z > 0$ and $u(x)$ is continuous, $u(x)/(z+x)$ is continuous on $[0,1]$. Consequently the first term converges to zero with increasing n (see e.g., Theorems 64.1, 64.2, and Sec. 67 of Ref. 6). The second term is majorized by $\|u - u_n\|^n \to_{n\to\infty} 0$. Thus $v_n(z) \to_{n\to\infty} v(z)$ pointwise. Now

$$T_n(z) = |v(z) - v_n(z)| \leqslant M \int_0^1 (d\sigma(x) + d\sigma_n(x))\,\frac{z}{z+x}$$

and

$$\int_0^1 d\sigma_n(x)\,\frac{z}{z+x} \leqslant \int_0^1 d\sigma(x)\,\frac{z}{z+x} \quad \text{for } z \in (0,\infty).^7$$

Hence

$$T_n(z) \leqslant 2M \int_0^1 d\sigma(x)\,\frac{z}{z+x} \xrightarrow[z\to 0]{} 0. \quad \text{(Lemma 3).}$$

Thus for any $\epsilon > 0$ there is a $\delta(\epsilon)$ independent of n such that $z < \delta(\epsilon)$ implies that $T_n(z) < \epsilon$. Now, let z, z' be in a compact subset S_ϵ of $[\delta(\epsilon), \infty)$. We have that

$$\begin{aligned}|T_n(z) - T_n(z')| &= \big|\,|v(z) - v_n(z)| - |v(z') - v_n(z')|\,\big| \\ &\leqslant |(v(z) - v(z')) - (v_n(z) - v_n(z'))| \\ &\leqslant \left|\int_0^1 d\sigma(x)\,\frac{(z-z')xu(x)}{(z+x)(z'+x)}\right| \\ &\quad + \left|\int_0^1 d\sigma_n(x)\,\frac{(z-z')xu_n(x)}{(z+x)(z'+x)}\right| \\ &\leqslant |z - z'|\,\frac{2C_0M}{\delta(\epsilon)} \xrightarrow[z\to z']{} 0.\end{aligned}$$

Therefore $T_n(z)$ is a sequence of uniformly continuous functions converging to zero for each $z = S_\epsilon$. This implies that the convergence is uniform for z in S_ϵ.[8] Thus given $\epsilon > 0$ one can pick a $\delta(\epsilon)$ such that $T_n(z) < \epsilon$ for $z < \delta(\epsilon)$ and then increase n to ensure that $T_n(z) < \epsilon$ on the complement of $[0, \delta(\epsilon))$.

Theorem 1: Let $C_0 < \frac{1}{2}$ and H, $\widehat{H}_n$ be as above. Then

$\hat{H}_n(z) \underset{n\to\infty}{\rightrightarrows} H(z)$ uniformly for z in any compact subset of $[0,\infty)$.

Proof: We have that

$$H(z) - \hat{H}_n(z) = T_n^1(z) + T_n^2(z),$$

where

$$T_n^1(z) = (AH)(z) - (A_nH)(z),$$

and

$$T_n^2(z) = (A_nH)(z) - (A_n\hat{H}_n)(z)$$
$$= \int_0^1 dt\,\{A_n'[\hat{H}_n + t(H-\hat{H}_n)](H-\hat{H}_n)\}(z),$$

Now,

$$|T_n^1(z)| = |H(z)[(B-B_n)H](z)|$$
$$\leqslant (1-2C_0)^{-1/2}|(B-B_n)H(z)|.$$
$$\underset{n\to\infty}{\longrightarrow} 0,$$

uniformly for $z\in[0,\infty)$ from Lemma 4 by setting $u=u_n=H$.

Hence,

$$\|T_n^1\|^n = \int_0^1 d\sigma_n(x)|T_n^1(x)| \underset{n\to\infty}{\longrightarrow} 0.$$

Further,

$$\|T_n^2\|^n \leqslant \operatorname*{Sup}_{t\in[0,1]} \|A_n'[\hat H_n + t(H-\hat H_n)]\|^n\, \|H-\hat H_n\|^n$$

and for each $t\in[0,1]$,

$$\|A_n'[\hat H_n + t(H-\hat H_n)]\|^n,$$
$$\leqslant \|tH + (1-t)\hat H_n\|^n \quad \text{(Lemma 1)}$$
$$\leqslant t\|H\|^n + (1-t)\|\hat H_n\|^n$$
$$\leqslant t\|H\| + (1-t)\|\hat H_n\|^n + t\,|\,\|H\|^n - \|H\|\,|.$$

Also $[t\|H\| + (1-t)\|\hat H_n\|^n] \leqslant (1-\sqrt{1-2C_0})$ from Corollary 1, and $t\,|\,\|H\|^n - \|H\|\,| \leqslant |\int_0^1(d\sigma(x) - d\sigma_n(x))H(x)| \underset{n\to\infty}{\longrightarrow} 0$ because of the continuity of $H(x)$.[6] Consequently, if $C_0<\frac{1}{2}$, one can ensure by increasing n that

$$\|T_n^2\|^n \leqslant (1-\sqrt{1-2C_0}/2)\|H-\hat H_n\|^n.$$

Therefore

$$\|H-\hat H_n\|^n \leqslant \|T_n^1\|^n + (1-\sqrt{1-2C_0}/2)\|H-\hat H_n\|^n$$
$$\leqslant 2(1-2C_0)^{-1/2}\|T_n^1\|^n \underset{n\to\infty}{\longrightarrow} 0.$$

It follows now from Lemma 4 and Corollary 1 that

$$(B_nH_n)(z) \underset{n\to\infty}{\longrightarrow} (BH)(z)$$

uniformly with respect to z. The proof is completed by observing that

$$H(z) = [1-(BH)(z)]^{-1}, \quad \hat H_n(z) = [1-(B\hat H_n)(z)]^{-1}$$

and $|(BH)(z)|, |(B_n\hat H_n)(z)| \leqslant \|H\|, \|\hat H_n\|^n < 1$ [Lemma 2 (iv), Corollary 1].

4. NEWTON'S APPROXIMATIONS TO *H*

In this section we use the techniques of Ref. 9 to deduce the convergence properties of the Newton method to approximate H. Therefore, we establish some parallel results. Since no reference to any measure other than σ will be made, the results are stated for A rather than A_μ. From Lemma 2 we have that A is positive and increasing. In Lemma 5 we establish the convexity of A and the analog of the weak positivity lemma.

Lemma 5: (i) Let $A'(u)$, h be as in Lemma 2 (ii) and let $v\geqslant u$, then $(A'(v) - A'(u))h\geqslant 0$ (ii) Let $u, h\geqslant 0$ and $\|u\|<1$; then $[1-A'(u)]^{-1}h\geqslant 0$

Proof: (i) The result follows by observing that $[A'(v) - A'(u)]h = (v-u)Bh + hB(v-u) = A'(v-u)h$ and Lemma 2(ii).

(ii) Since $\|A'(u)\|\leqslant\|u\|<1$ (Lemma 1), the series expansion of $[1-A'(u)]^{-1}$ converges in $L^1(\mu)$. If $h\in L^1(\mu)$ is nonnegative, then each term in the series for $[1-A'(u)]^{-1}h$ is easily seen to be so, from Lemma 2(ii).

As a consequence of the convexity of A, we have

Corollary 2: Let $A'(u)$ be as in Lemma 5 (i) and $v\geqslant u\geqslant 0$; then

$$\theta(u,v) = Av - Au - A'(u)(v-u)\geqslant 0.$$

Proof: Since

$$\theta(u,v) = \int_0^1 dt\,\{A'[u + t(v-u)] - A'(u)\}(v-u)$$

and $[u + t(v-u)]\geqslant u$ for $0\leqslant t\leqslant 1$, the result follows from Lemma 5 (i).

Let $\chi(v) = [1-A'(v)]^{-1}[A(v) - A'(v)v]$. Newton's approximation ω to a fixed point of A is given by $\omega = \chi(v)$, with v being an initial guess.

Lemma 6: Let $0\leqslant v\leqslant H$ on $[0,1]$. Then

(i) $\|A'(v)\|<1$,

(ii) $\omega = \chi(v) \leqslant H$ on $[0,1]$.

Proof: (i) Since $0\leqslant v\leqslant H, \|v\|\leqslant\|H\|\leqslant(1-\sqrt{1-2C_0})<1$ (Corollary 1). The result now follows from $\|A'(v)\|\leqslant\|v\|$ (Lemma 1).

(ii) We have that

$$[1-A'(v)](H-\omega) = [AH - Av - A'(v)(H-v)]$$
$$= \theta(v,H)$$
$$\geqslant 0$$

for $H\geqslant v\geqslant 0$ (Corollary 2). Since $\|A'(v)\|<1$, from (i), the result follows from Lemma 5(ii).

Let u_0 be arbitrary with $\|u_0\|\leqslant[1-\sqrt{1-2C_0}]$ and $u_{n+1} = \chi(u_n)$, $n\geqslant 0$; $\{u_n\}$ will be called Newton's sequence generated by u_0.

Lemma 7: Let $\{u_n\}$ be Newton's sequence generated by $u_0 = 0$. Then $u_n\leqslant u_{n+1}\leqslant H$ for each n on $[0,1]$.

Proof: It is clear that $u_0\leqslant u_1 = 1\leqslant H$. Now assume that $H\geqslant u_n\geqslant u_{n-1}$. We have that

$$u_{n+1} - u_n = [Au_n + A'(u_n)(u_{n+1} - u_n)]$$
$$- [Au_{n-1} + A'(u_{n-1})(u_n - u_{n-1})]$$
$$= \theta(u_{n-1}, u_n) + A'(u_n)(u_{n+1} - u_n)$$
$$\geqslant A'(u_n)(u_{n+1} - u_n)$$

from Corollary 2.

Since $0\leqslant u_n\leqslant H, \|A'(u_n)\|<1$ [Lemma 6(i)]. Therefore, from Lemma 5(ii), and Lemma 6 (ii) $H\geqslant u_{n+1}\geqslant u_n$. Using the induction principle we have that $0\leqslant u_n\leqslant u_{n+1}\leqslant H$ on $[0,1]$

for all n.

The results obtained so far are sufficient to conclude convergence on $[0,1]$. However, as such it is not even clear if the domain of definition of $\{u_n(z)\}$ extends beyond $[0,1]$. In the following we define $\{u_n(z)\}$ on $[0,\infty)$ and establish the result of Lemma 7 there.

The equation $u_{n+1} = \chi(u_n)$ reduces to

$$u_{n+1}(z) = 1 + u_{n+1}(z)(Bu_n)(z) + u_n(z)[B(u_{n+1} - u_n)](z). \tag{5}$$

Since $|(Bu_n)(z)| \leqslant \|u_n\| \leqslant \|H\| < 1$ from Lemma 2(iv) and Lemma 7, (5) defines a continuous $u_{n+1}(z)$ on $[0,\infty)$ if $u_n(z)$ is defined and continuous there. Since $u_0 = 0$ for $z \geqslant 0$, $\{u_n(z)\}$ is defined by (5) on $[0,\infty)$. Furthermore, we have

Lemma 8: Let $\{u_n(z)\}$ be defined by (5) on $[0,\infty)$ with $u_0 = 0$. Then $0 \leqslant u_n \leqslant u_{n+1} \leqslant H$ for each $z \geqslant 0$, and all n.

Proof: First we show that $0 \leqslant u_n \leqslant u_{n+1}$ for all n on $[0,\infty)$. Since $0 = u_0 \leqslant u_1 = 1$, the result is true for $n = 0$. Assume that $0 \leqslant u_{n-1} \leqslant u_n$. It follows from (5) that

$$\begin{aligned} u_{n+1} - u_n &= (u_{n+1} - u_n)Bu_n + (u_n - u_{n-1}) \\ &\quad \times B(u_n - u_{n-1}) + u_nB(u_{n+1} - u_n) \\ &= [1 - Bu_n]^{-1}[(u_n - u_{n-1})B(u_n \\ &\quad - u_{n-1}) + u_nB(u_{n+1} - u_n)]. \end{aligned}$$

Now, for $z \in [0,1]$, $u_{n+1} \geqslant u_n \geqslant 0$ from Lemma 7, therefore, from Lemma 2(i), $B(u_{n+1} - u_n) \geqslant 0$ for $z \in [0,\infty)$. Also, $0 \leqslant u_n \leqslant H$ on $[0,1]$; hence $0 \leqslant Bu_n \leqslant \|u_n\| < 1$ from Lemma 2(iv) and Lemma 7. These results and the assumption $0 \leqslant u_{n-1} \leqslant u_n$ on $[0,\infty)$ are easily seen to imply that $(u_{n+1} - u_n) \geqslant 0$ on $[0,\infty)$. The result now follows by induction.

The fact that $\{u_n\}$ is bounded by H on $[0,\infty)$ follows by a similar argument. It is clearly true for $n = 0$, and $(H - u_{n+1})$ satisfies

$$\begin{aligned} H - u_{n+1} &= [1 - Bu_n]^{-1}[(H - u_n)B(H - u_n) \\ &\quad + u_nB(H - u_{n+1})]. \end{aligned}$$

The assumption $0 \leqslant u_n \leqslant H$ on $[0,\infty)$ implies that $H - u_{n+1} \geqslant 0$ exactly as above.

After we have established that $0 \leqslant u_n \leqslant u_{n+1} \leqslant H$, a proof of uniform convergence is a routine matter.

Theorem 2: Let $\{u_n\}$ be as in Lemma 8. Then $u_n(z) \uparrow H(z)$ uniformly for z in any compact subset of $[0,\infty)$.

Proof: Let the set under consideration be denoted by S. Since $\{u_n\}$ is nondecreasing sequence bounded by H and S is closed, bounded; $u_n \uparrow u \leqslant H$ on S.

Now from (5)

$$u = \lim_{n\to\infty} u_{n+1} = 1 + \lim_{n\to\infty} [u_{n+1}Bu_n + u_nB(u_{n+1} - u_n)].$$

Since $\{(z/(z+x))(u_{n+1} - u_n)(x)\}$ is bounded by a σ-integrable function H and converges to zero pointwise, the Lebesgue dominated convergence theorem yields that $B(u_{n+1} - u_n) \to_{n\to\infty} 0$. This and the fact that $0 \leqslant u_n \leqslant H$ imply that $u_nB(u_{n+1} - u_n) \to_{n\to\infty} 0$ on S. By a similar argument it follows that

$$\lim_{n\to\infty} u_{n+1}Bu_n = uBu.$$

Thus u satisfies

$$u(z) = [1 - (Bu)(z)]^{-1},$$

with $|(Bu)(z)| < 1$. Since $u(z) \leqslant H(z)$ for $z \in [0,1]$, and H is the minimal solution [Proposition 1 (i)] $u(z) = H$ on $[0,1]$, implying that $u(z) = [1 - (BH)(z)]^{-1}$ on S. Therefore $u(z) = H(z)$ on S. Also, $\{u_n(z)\}$ is a nondecreasing sequence of continuous functions, converging to a continuous function H. Consequently, the convergence is uniform on S, by Dini's theorem.

We have three sequences: $\{H_n\}$,$\{K_n\}$, and $\{u_n\}$, given by (2), (3) and Theorem 2, respectively, which converge monotonically and uniformly to H. The convergence of $\{H_n\}$,$\{K_n\}$ was considered on $[0,1]$[1–3], but, as above, it is sufficient to conclude the uniform convergence on S. The sequence $\{\widehat{H}_n\}$ falls out of this category. Although $\{\widehat{H}_n\}$ converges uniformly, it may not have any bound property. It is known that $H \geqslant K_n \geqslant H_n$ on $[0,1]$ (which implies the same on S) for each n if $K_0 = H_0 = 0$.[3] In the following we show that u_n is even closer to H.

Proposition 2: Let $\{u_n\}$,$\{K_n\}$ be as in Theorem 2 and Eq. (3), respectively; then, for each n, $H \geqslant u_n \geqslant K_n$ on $[0,\infty)$.

Proof: With $u_0 = K_0 = 0$ one has that $1 = u_1 \geqslant K_1 = 1$. Now assume that $u_n \geqslant K_n$. For $n \geqslant 1$, $(u_{n+1} - K_{n+1})$ is given by

$$\begin{aligned} u_{n+1} &- K_{n+1} \\ &= u_{n+1}Bu_n - K_{n+1}BK_n + u_nB(u_{n+1} - u_n) \\ &= (u_{n+1} - K_{n+1})BK_n + u_{n+1}B(u_n - K_n) \\ &\quad + u_nB(u_{n+1} - u_n) \\ &= [1 - BK_n]^{-1}[u_{n+1}B(u_n - K_n) \\ &\quad + u_nB(u_{n+1} - u_n)] \\ &\geqslant 0 \end{aligned}$$

for $u_n \geqslant K_n$ by assumption, $u_{n+1} \geqslant u_n \geqslant 0$ from Lemma 8, and $0 \leqslant BK_n < 1$ follows from $0 \leqslant K_n \leqslant u_n \leqslant H$. Thus $(u_{n+1} - K_{n+1}) \geqslant 0$ on S. The result now follows by induction.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grant No. *A* 3604. The author is grateful to Professor H. O. Pritchard for his hospitality.

[1] R. L. Bowden and P. F. Zweifel, Astrophys. J. **210**, 178 (1976) and references cited therein.

[2] R. L. Bowden, J. Math. Phys. **20**, 608 (1979).

[3] C. T. Kelley, J. Math. Phys. **21**, 408 (1980).

[4] D. Masson, J. Math. Phys. **22**, 462 (1981).

[5] See, e. g., R. H. Moore in *Nonlinear Integral Equations*, edited by P. M. Anselone (The U. Wisconsin P., Madison, WI, 1964), pp. 65–98.

[6] H. S. Wall, *Analytic Theory of Continued Fractions* (Chelsea, New York, 1967).

[7] See, e.g., G. A. Baker, Jr., in the *Padé approximant in Theoretical Physics*, edited by G. A. Baker, Jr., and J. L. Gammel (Academic, New York, 1970).

[8] See, e.g., R. Courant and D. Hilbert, *Method of Mathematical Physics*, Vol. 1 (Interscience, New York, 1961), pp. 57–61.

[9] S. R. Vatsya, J. Math. Phys. **22**, 2977 (1981).

Dynamical importance of vorticity and shear in the universe

J. L. Sanz
Departamento Física Teórica, Facultad de Ciencias, Universidad de Santander, Santander, Spain

(Received 21 August 1981; accepted for publication 20 November 1981)

We study the dynamical importance of vorticity, ω^2/ρ, assuming different upper limits on the relative shear, σ/θ, for a general relativistic model with a content represented by a perfect fluid distribution with a linear equation of state. Adopting a very conservative point of view with respect to the values of the Hubble constant and the density parameters of matter and radiation, we obtain that either $\sigma/\theta > 7\%$ or there was a bounce at some point in the past during the matter era if the present-day relative vorticity is $(\omega/\theta)_0 > 4\%$. Taking into account the latest results on singularities, the possibility of a bounce must be regarded from a local point of view.

PACS numbers: 98.80. – k, 98.80.Dr

INTRODUCTION

Large-scale properties of our real universe are well described by the standard Friedmann–Robertson–Walker (FRW) models,[1-3] which are relativistic models whose geometry possesses homogeneity and isotropy and a content represented by a perfect fluid. Also, these type of models can be characterized by a nonvanishing expansion ($\theta > 0$) and have no rotation, distortion, and acceleration ($\omega = \sigma = \dot{u} = 0$).[4] In spite of the theoretical simplicity and observational evidence supporting this view, there has been a lot of work concerning more general cosmological models.

We shall deal in this paper with models possessing vorticity and, concretely, we shall assume that the present-day relative vorticity $(\omega/\theta)_0 > 4\%$. There are several reasons to relax the rigid assumptions of the standard picture. From a theoretical point of view, the FRW models are highly unstable: vorticity perturbations[5] are amplified when one goes back in time. Also, the possible influence of vorticity on the expansion of the universe has animated many theorists[5-9] as a possibility to escape from the inevitable singularity of FRW models through a bouncing point. Another interesting aspect is that the rotation of galaxies could be explained by the fact that they condensed out of a rotating universe.[10] From an observational point of view: direct observations[11] give very weak limits on the present-day vorticity, $(\omega/\theta)_0 \lesssim 1$. Strong constraints can be inferred from upper limits on anisotropies of the cosmic background radiation if one assumes perturbed FRW models.[12]

Our point of view is that observations do not rule out the possibility of using more general cosmological models that may not differ from the FRW ones from an observational standpoint, but with a quite different metric. In any case, our position makes it possible to understand which statements are geometry-dependent and which are not.

We shall study in this paper the dynamical importance of vorticity in the past of the universe using, essentially, the equations derived from conservation of energy–momentum for a perfect fluid and the law governing the evolution of vorticity.[4,5,9] The possibility of a bounce in a dust-filled universe is carefully examined using Raychaudhuri's equation,[13] and conclusions are drawn concerning the relevant physical quantities σ/θ and ω/θ.

1. THE EVOLUTION OF VORTICITY

A. Basic equations

In general relativity (GR), the conservation of energy and momentum for a perfect fluid is expressed by the well-known equations[4,5,9] (we choose units such that $c = 8\pi G = 1$; $a,b,\ldots = 0,1,2,3$)

$$\dot{\rho} + (\rho + p)\theta = 0\,, \tag{1}$$

$$\dot{u}^a + (\rho + p)^{-1} h^{ab} p_{;b} = 0\,, \quad h_{ab} \equiv g_{ab} + u_a u_b\,, \tag{2}$$

where u^a is the average velocity of matter, ρ is the energy density, p is the isotropic pressure, $\theta \equiv u^a{}_{;a}$ is the expansion scalar, $\dot{t} \equiv t_{;a} u^a$ for any tensor t, $\dot{u}^a$ is the acceleration, and g_{ab} is the metric tensor.

By defining the vorticity vector and the shear as usual,

$$2\omega^a \equiv \eta^{abcd} u_b u_{c;d}\,, \quad \sigma_{ab} \equiv \tfrac{1}{2} h_a{}^c h_b{}^d (u_{c;d} + u_{d;c}) - \tfrac{1}{3}\theta h_{ab}\,, \tag{3}$$

and applying the Ricci identity to the velocity u^a, one arrives at the propagation equation for ω^a along the flow lines of the fluid,[4,5,9]

$$h^a{}_b \dot{\omega}^b + \tfrac{2}{3}\theta\omega^a = \sigma^{ab}\omega_b + \tfrac{1}{2}\eta^{abcd} u_b \dot{u}_{c;d}\,. \tag{4}$$

Let us assume a barotropic equation of state for the fluid $p = p(\rho)$. Thus, substituting $\dot{u}^a$ given by Eq. (2) into Eq. (4), one easily obtains

$$h^a{}_b \dot{\omega}^b + \left[\tfrac{5}{3}\theta + (\rho + p)^{-1}(\rho + p)^{\cdot}\right]\omega^a = \sigma^{ab}\omega_b\,, \tag{5}$$

where we have taken into account Eq. (1). From Eq. (5) one arrives at the equation governing the evolution of the vorticity scalar,

$$\{\ln[(\rho + p)R^5\omega]\}^{\cdot} = \sigma_{ab} n^a n^b\,, \quad \omega \equiv +(\omega^a\omega_a)^{1/2}\,, \quad n^a \equiv \omega^{-1}\omega^a \tag{6}$$

where $R(x^a)$ is defined by $\dot{R}R^{-1} = \tfrac{1}{3}\theta$. Thus angular momentum, $L \propto (\rho + p)R^5\omega$, is conserved if $\sigma_{ab} n^a n^b \equiv 0$, i.e., the component of σ_{ab} along the axis of rotation vanishes.

By using comoving coordinates ($u^a = \delta^a_0$) and the variable $x \equiv R_0 R^{-1}$ (hereafter a subscript 0 will denote a present-day value), the vorticity evolution equation (6) can be integrated in the form

0022-2488/82/091732-05$02.50

$$\omega = \omega_0 x^5(\rho + p)_0(\rho + p)^{-1} \times \exp\left[- 3 \int_1^x dx(x\theta)^{-1}\sigma_{ab} n^a n^b \right] . \tag{7}$$

On the other hand, considering the inequalities

$$-\frac{2}{\sqrt{3}}\,\sigma \leqslant \sigma_{ab} n^a n^b \leqslant \frac{2}{\sqrt{3}}\,\sigma , \quad \sigma \equiv +\left(\tfrac12 \sigma_{ab}\sigma^{ab}\right)^{1/2}$$

that hold for any unit vector n^a, one obtains for $\theta > 0$ and $t < t_0$

$$x^5(\rho + p)_0(\rho + p)^{-1} \exp\left\{ - 2\sqrt{3} \int_1^x \frac{dx}{x}\left(\frac{\sigma}{\theta}\right)\right\} \leqslant \frac{\omega}{\omega_0} \leqslant x^5(\rho + p)_0(\rho + p)^{-1} \exp\left\{2\sqrt{3} \int_1^x \frac{dx}{x}\left(\frac{\sigma}{\theta}\right)\right\} . \tag{8}$$

The particular case of a linear equation of state

$$p = n\rho , \quad n \in [0,1] \Rightarrow \rho = \rho_0\, x^{3(1+n)} ,$$

implies the following bounds on ω and ω^2/ρ, which measures the dynamical importance of vorticity,

$$x^{2-3n} e^{-a} \leqslant \frac{\omega}{\omega_0} \leqslant x^{2-3n} e^{a} , \quad a \equiv 2\sqrt{3} \int_1^x \frac{dx}{x}\left(\frac{\sigma}{\theta}\right), \tag{9}$$

$$x^{1-9n} e^{-2a} \leqslant \left(\frac{\omega^2}{\rho}\right)\left(\frac{\omega^2}{\rho}\right)_0^{-1} \leqslant x^{1-9n} e^{2a} . \tag{10}$$

B. The dynamical importance of vorticity

Let us assume that $\sigma/\theta < (\sqrt{3}/2)|n - \frac23|$, $n \neq \frac23$, $\forall t \in [t_1,t_0]$, $t_1 < t_0$. Then Eq. (9) implies

$$\forall t \in [t_1,t_0) \Rightarrow \begin{cases} \omega > \omega_0 & \text{for } n \in [0,\frac23) \\ \omega < \omega_0 & \text{for } n \in (\frac23,1] \end{cases}, \tag{11}$$

i.e., the rotation was greater (or smaller) in the past.

On the other hand, if we assume $\sigma/\theta < (3\sqrt{3}/4)|n - \frac19|$, $n \neq \frac19$, $\forall t \in [t_1,t_0]$, $t_1 < t_0$, Eq. (10) leads to

$$\forall t \in [t_1,t_0) \Rightarrow \begin{cases} \dfrac{\omega^2}{\rho} \geqslant \left(\dfrac{\omega^2}{\rho}\right)_0 x^{\epsilon} > \left(\dfrac{\omega^2}{\rho}\right)_0 \\ \qquad (\epsilon > 0) \quad \text{for } n \in [0,\frac19) \\ \dfrac{\omega^2}{\rho} \leqslant \left(\dfrac{\omega^2}{\rho}\right)_0 x^{-\epsilon'} < \left(\dfrac{\omega^2}{\rho}\right)_0 \\ \qquad (\epsilon' > 0) \quad \text{for } n \in (\frac19,1] \end{cases} \tag{12}$$

i.e., vorticity has been dynamically important in the past or not, depending on the equation state chosen to represent the content of the universe.

Regarding our real universe, a semirealistic representation can be made by means of the particular equations of state $p = 0$ (matter era) and $p = \frac13$ (radiation era). In these cases we have

1. Case $n = 0$

On choosing $\sigma/\theta < 1/\sqrt{3}$, $\forall t \in [t_1,t_0]$, $t_{\rm eq} < t_1 < t_0$ (hereafter a subscript eq will refer to the equilibrium point, i.e., the point where the density of radiation and matter are the same), Eq. (9) leads to

$$\forall t \in [t_1,t_0) \Rightarrow 1 < \frac{\omega}{\omega_0} < x^4 , \tag{13}$$

i.e., if the relative distortion σ/θ has been smaller than 57% between the equilibrium point and the present time, we can conclude that vorticity has been greater in the past.

The condition $\sigma/\theta < (1/4\sqrt{3})(1 - \epsilon)$, $\epsilon \in (0,1)$, $\forall t \in [t_1,t_0]$, $t_1 < t_0$, substituted in Eq. (10), gives

$$\forall t \in [t_1,t_0) \Rightarrow 1 < x^{\epsilon} \leqslant \left(\frac{\omega^2}{\rho}\right)\left(\frac{\omega^2}{\rho}\right)_0^{-1} \leqslant x^{2-\epsilon} < x^2 , \tag{14}$$

i.e., vorticity was dynamically important in the past if the relative distortion has been smaller than 14%. As a consequence of this analysis an interesting question arises: Can the dynamical effect of vorticity have produced a bounce in the past? We shall give an answer in the next section.

2. Case $n = 1/3$

On choosing $\sigma/\theta < (1/2\sqrt{3})(1 - \epsilon)$, $\epsilon \in (0,1)$, $\forall t \in [t_1,t_{\rm eq}]$, $t_1 < t_{\rm eq}$, Eqs. (9) and (10) lead to the following bounds:

$$\forall t \in [t_1,t_{\rm eq}) \Rightarrow \begin{cases} 1 < w^{\epsilon} \leqslant \dfrac{\omega}{\omega_{\rm eq}} \leqslant w^{2-\epsilon} < w^2 , \quad w \equiv \dfrac{R_{\rm eq}}{R} \\ w^{-4} < w^{-2(2-\epsilon)} \leqslant \left(\dfrac{\omega^2}{\rho}\right)\left(\dfrac{\omega^2}{\rho}\right)_{\rm eq}^{-1} \leqslant w^{-2\epsilon} < 1 , \end{cases} \tag{15}$$

i.e., if the relative distortion has been smaller than 28% before the equilibrium point then vorticity was greater in the past of this point but was not dynamically important.

2. THE POSSIBLE BOUNCE OF A DUST-FILLED UNIVERSE

A. Assumptions

Let us consider the following hypothesis in $(t_1,t_0]$,$t_1 < t_0$: (i) There is expansion, $\theta > 0$, (ii) the content can be represented by "dust," $p = 0$, (iii) $0 < \sigma/\theta \leqslant (\sqrt{3}/12)(1 - \epsilon)$, $\epsilon \in (0,1)$.

Obviously, for dust Eq. (2) gives $\dot{u}^a = 0$, i.e., the fluid lines are geodesics (though $\omega \neq 0$ in general). Assumption (iii) leads to the bounds given by Eq. (14):

$$x \geqslant 1 \Rightarrow x^{\epsilon} \leqslant \left(\frac{\omega^2}{\rho}\right)\left(\frac{\omega^2}{\rho}\right)_0^{-1} \leqslant x^{2-\epsilon} . \tag{16}$$

On the other hand, the equation governing the evolution of the expansion scalar for a dust is

$$\dot\theta + \tfrac13\theta^2 + 2(\sigma^2 - \omega^2) + \tfrac12\rho = 0 , \tag{17}$$

that is, a particular case of Raychaudhuri's equation.[13]

By defining the function $y \equiv \theta^2/\rho$, and using Eqs. (1) and (17), one arrives at the law

$$\frac{\partial y}{\partial x} + x^{-1}\left[1 - 12\left(\frac{\sigma}{\theta}\right)^2\right] y = -12x^{-1}\left(\frac{\omega^2}{\rho} - \frac14\right), \tag{18}$$

where comoving coordinates and the variable $x \equiv R_0 R^{-1}$ have been introduced, where $\dot{R}R^{-1} \equiv \frac13\theta$. Equation (18) can be formally integrated in the form

$$y = x^{-1}e^{A}\left[y_0 - 12\int_1^x dx\left(\frac{\omega^2}{\rho} - \frac{1}{4}\right)e^{-A}\right], \qquad (19)$$

$$A \equiv 12\int_1^x dx\, x^{-1}\left(\frac{\sigma}{\theta}\right)^2 .$$

B. A sufficient condition for a bounce

We mean by a bounce the possibility that there was a time $t_b \in (t_1, t_0)$ such that the "radius of the universe" R reached a minimum value at this time, and the relative distortion was bounded, $0 < (\sigma/\theta)_b < \infty$ (hereafter a subscript b will refer to the bounce point). Obviously, a sufficient condition is

$$\exists t_b \in (t_1, t_0)/\dot{R}_b \equiv 0 , \quad \ddot{R}_b > 0 ; \qquad (20)$$

the second condition, $0 < \sigma/\theta < \infty$, is trivially ensured under assumption (iii). The condition for a bounce can be rewritten, taking into account Eqs. (17) and (19), as

$$\exists x_b \in (1, x_1)/y_0 = 12\int_1^{x_b} dx\left(\frac{\omega^2}{\rho} - \frac{1}{4}\right)e^{-A} ,$$

$$\left(\frac{\omega^2}{\rho}\right)_b > \frac{1}{4} . \qquad (21)$$

Obviously, as $y_0 \equiv (\theta^2/\rho)_0$ must be positive, a sufficient condition for the existence of such a bounce is

$$\exists x_b > 1 \Big/ \int_1^{x_b} dx\left(\frac{\omega^2}{\rho} - \frac{1}{4}\right)$$

$$\times \exp\left\{-12\int_1^x dx\, x^{-1}\left(\frac{\sigma}{\theta}\right)^2\right\} > 0 , \quad \left(\frac{\omega^2}{\rho}\right)_b > \frac{1}{4} . \qquad (22)$$

C. Case $(\omega^2/\rho)_0 \geqslant \frac{1}{4}$

If we assume a lower limit of 25% on the present-day value of the vorticity–density ratio, Eq. (16) implies

$$x \geqslant 1 \Rightarrow \left(\frac{\omega^2}{\rho}\right) \geqslant \frac{1}{4}x^{\epsilon} > \frac{1}{4} ; \qquad (23)$$

thus the two inequalities given by Eq. (22) are trivially satisfied. Therefore, in a dust-filled universe, if a time t_0 exists such that $(\omega^2/\rho)_0 \geqslant \frac{1}{4}$ then $\theta > 0$ and $0 < \sigma/\theta < \sqrt{3}/12$ cannot be satisfied indefinitely in the past, i.e., either $\sigma/\theta > 14\%$ or there is a bounce at some point in the past.

Next, we shall try to obtain upper limits on x_b depending on the value of y_0 under the possibility of a real bounce. On the one hand, assumption (iii) implies

$$0 < \frac{\sigma}{\theta} \leqslant \frac{1}{4\sqrt{3}}(1 - \epsilon) , \quad \epsilon \in (0,1) \Rightarrow x^{-s} \leqslant e^{-A} < 1 ,$$

$$s \equiv \tfrac{1}{4}(1 - \epsilon)^2 . \qquad (24)$$

This, together with inequality (23), can be substituted into Eq. (19):

$$x > 1 \Rightarrow y \leqslant x^{s-1}\left\{y_0 - 12\int_1^x dx\left[\left(\frac{\omega^2}{\rho}\right)_0 x^{\epsilon} - \frac{1}{4}\right]x^{-s}\right\} .$$

After a lengthy but easy calculation, one arrives at the inequality

$$x > 1 \Rightarrow y < 3(1 - s)^{-1} - 3(1 - s + \epsilon)^{-1}x^{\epsilon} + x^{s-1}$$

$$\times [y_0 - 3\epsilon(1 - s)^{-1}(1 - s + \epsilon)^{-1}] . \qquad (25)$$

From this last expression, if we assume $y_0 \leqslant 3\epsilon(1 - s)^{-1}(1 - s + \epsilon)^{-1}$, we obtain the bound

$$x > 1 \Rightarrow y < 3(1 - s)^{-1} - 3(1 - s + \epsilon)^{-1}x^{\epsilon}$$

and as $y \equiv \theta^2/\rho$ must be positive,

$$x^{\epsilon} < 1 + \epsilon(1 - s)^{-1} . \qquad (26)$$

In the opposite case, $y_0 > 3\epsilon(1 - s)^{-1}(1 - s + \epsilon)^{-1}$, Eq. (25) leads to the inequality

$$x > 1 \Rightarrow y < 3(1 - s + \epsilon)^{-1}\left[1 + \tfrac{1}{3}(1 - s + \epsilon)y_0 - x^{\epsilon}\right]$$

and analogously $y > 0$ implies

$$x^{\epsilon} < 1 + \tfrac{1}{3}(1 - s + \epsilon)y_0 . \qquad (27)$$

Summing up: If a time t_0 exists such that $(\omega^2/\rho)_0 \geqslant \frac{1}{4}$ for a dust-filled universe, then the assumptions $\theta > 0$ and $0 < \sigma/\theta < \infty$ can be maintained in the past at most in

$$x \in [1, \{1 + \epsilon(1 - s)^{-1}\}^{\epsilon^{-1}})$$
$$\text{for } y_0 \leqslant 3\epsilon(1 - s)^{-1}(1 - s + \epsilon)^{-1} ,$$
$$x \in [1, \{1 + \tfrac{1}{3}(1 - s + \epsilon)y_0\}^{\epsilon^{-1}})$$
$$\text{for } y_0 > 3\epsilon(1 - s)^{-1}(1 - s + \epsilon)^{-1} . \qquad (28)$$

These are very important upper limits on x that we will use from a physical point of view in Sec. 3.

D. Case $(\omega^2/\rho)_0 < \frac{1}{4}$

In this case, if the relative radius $\bar{x} \equiv [\frac{1}{4}(\omega^2/\rho)_0^{-1}]^{\epsilon^{-1}} > 1$ can be reached in the past, Eq. (16) leads to

$$\left(\frac{\omega^2}{\rho}\right)_{\bar{x}} > \left(\frac{\omega^2}{\rho}\right)_0 \bar{x}^{\epsilon} > \frac{1}{4} , \qquad (29)$$

i.e., we can always choose as an initial condition a lower limit of 25% on the vorticity–density ratio unless the assumptions (i), (ii), or (iii) of Sec. 2A are dropped before. In principle there is a possible behavior such that these conditions can be satisfied, but the relative radius $\bar{x}$ is never reached, in this case the radius R reaches asymptotically a minimum value such that $x_m < \bar{x}$. However, in this case there necessarily exists a point x_2 such that $\dot{\theta}_2 \equiv 0$; but then Raychandhuri's equation gives $(\omega^2/\rho)_2 > \frac{1}{4}$ that can be used as an initial condition.

3. APPLICATION TO OUR REAL UNIVERSE

The results obtained in the last section allow a maximum relative radius $x_{\max}$, given by formulas (28), up to which the conditions (i) $\theta > 0$, (ii) dust era, and (iii) $0 < \sigma/\theta < (\sqrt{3}/12)(1 - \epsilon)$, $\epsilon \in (0,1)$, can be maintained. Of course, it is very interesting to know whether this upper limit $x_{\max}$ lies in the dust era, characterized by the upper bound x_{eq}.

A. Assumptions on observational parameters

We shall adopt a conservative point of view. Current values of the Hubble constant H_0, in km s^{-1} Mpc^{-1}, are in the range $H_0 \in [40,100]$, to which corresponds the normalized value

$$h \in [0.4,1] \, , \quad h \equiv 10^{-2} H_0 \, . \tag{30}$$

A standard value[14] usually taken in the literature is $h = 0.5$.

Observations of visible mass[15,16] give a firm lower limit on the density of matter: $\rho_{m0} > 4 \times 10^{-31} h^2 (\mathrm{g\ cm^{-3}})$, which corresponds to the following bound on the density parameter:

$$\Omega_m > 0.02 \, . \tag{31}$$

Also the cosmological origin of the cosmic background radiation gives a lower-limit form on the density of radiation: $\rho_{r0} > 4.5 \times 10^{-34} (\mathrm{g\ cm^{-3}})$, which corresponds to the density parameter

$$\Omega_r > 2.25 \times 10^{-5} h^{-2} \, . \tag{32}$$

The equilibrium point is given by the theoretical expression

$$x_{\mathrm{eq}} \equiv \Omega_m / \Omega_r \, ; \tag{33}$$

then according to Eqs. (30) and (32)

$$x_{\mathrm{eq}} < 4.5 \times 10^5 \Omega_m \, . \tag{34}$$

On the other hand, $y_0 \equiv (\theta^2/\rho)_0 = 3\Omega_m^{-1}$, so inequality (31) leads to

$$y_0 < 150 \, . \tag{35}$$

B. Results

Taking into account the upper limits given by Eqs. (31) and (35), the last paragraph of Sec. 2B can be rewritten as follows: If a time t_0 exists such that $(\omega/\theta)_0 \geqslant (\Omega_m/12)^{1/2} > 4\%$ for a dust-filled universe, assumptions $\theta > 0$ and $0 < \sigma/\theta < (\sqrt{3}/12)(1-\epsilon)$ can be maintained in the past at most in

$$\begin{cases} x \in [1, \{1 + \epsilon(1-s)^{-1}\}^{\epsilon^{-1}}) \\ \qquad \text{for } \Omega_m \geqslant \epsilon^{-1}(1-s)(1-s+\epsilon) \, , \\ x \in [1, \{1 + 50(1-s+\epsilon)\}^{\epsilon^{-1}}) \\ \qquad \text{for } \Omega_m < \epsilon^{-1}(1-s)(1-s+\epsilon) \, . \end{cases} \tag{36}$$

Let us choose as an indicative value $\epsilon \equiv 0.5$ (i.e., $0 < \sigma/\theta < \sqrt{3}/24 \sim 7\%$). Then the corresponding intervals are

$$\begin{cases} x \in [1,2.4) & \text{for } \Omega \geqslant 2.7 \, , \\ x \in [1,5311) & \text{for } \Omega < 2.7 \, . \end{cases} \tag{37}$$

It is remarkable that the two upper limits, up to which the conditions $\theta > 0$, $0 < \sigma/\theta < 7\%$ can be maintained, are below the upper limit obtained for the equilibrium point in the most unfavorable case because Eq. (34) leads to $x_{\mathrm{eq}} < 9 \times 10^3$ for $\Omega_m = 0.02$. Thus, if the present-day relative rotation $(\omega/\theta)_0 > 4\%$, either $\sigma/\theta > 7\%$ or there was a bounce at some point in the past during the matter-dominated era!

The last strong conclusion has been obtained under the crucial hypothesis $(\omega/\theta)_0 > 4\%$; however, if $(\omega/\theta)_0 < (\Omega m/12)^{1/2}$ we have seen in Sec. 2D that the sign of this inequality can be reversed once the universe reached the relative radius $\bar{x} \equiv [(\Omega_m/12)(\omega/\theta)_0^{-2}]^{\epsilon^{-1}}$. For the indicated value $\epsilon \equiv 0.5$ if one wishes the $\bar{x}$ value to lie below the equilibrium point, the following inequality must be satisfied:

$$\left(\frac{\omega}{\theta}\right)_0 > 2.8\% \, , \tag{38}$$

which is a lower limit to obtain the above conclusion with the simple analysis we have made.

Of course, this kind of analysis does not allow us to draw any physically interesting conclusion for low densities if ϵ is chosen to be close to zero $(0 < \sigma/\theta < \sqrt{3}/12 \sim 14\%)$, because then the upper limit given by Eq. (36) goes to infinity.

4. CONCLUSIONS

We have carefully examined the law governing the evolution of vorticity for general relativistic models with linear equations of state, assuming different upper limits on the relative shear, and we have emphasized the possible dynamical importance of vorticity in the past of the universe. This suggests, the possibility of a bounce at some point in the past with a finite relative distortion.

Next, we have analyzed such a possibility for a dust-filled universe and have found that, if the present-day relative vorticity $(\omega/\theta)_0 > 4\%$, either σ/θ was greater than a 7% at some point in the past or there was a bounce point. The two possibilities lie in the matter-dominated era.

Regarding the well-known theorems on singularities, the latest results[17-19] seem to indicate that, for the equations of state we have assumed, there must be a true curvature singularity. Thus the bounce must be local unless the space-time be very pathological.

It might be, as suggested by some authors,[17,20] that the singularity consisted of a small region of space-time with most of the matter avoiding it. Nevertheless, our opinion is that the occurrence after the equilibrium point of a local bounce, but extended to a very large region of space-time, does not seem plausible. If that is so, our present understanding about primordial element formation, the evolution of galaxies, cosmic background radiation, etc., could be highly affected. For instance, the possibility of such an extended bounce after decoupling (the point where the radiation became transparent to the matter) could imply that stars, clusters, and galaxies were in existence before that bounce point and thus their age could be very much greater than the standard observational values. Regarding the existence of a bounce in the radiation era, let us remark that we cannot extend our analysis to this epoch because pressure gradients must be incorporated into the Raychaudhuri equation. We want to stress that, in general, all the observations explained in the standard picture of the universe must be carefully reexamined and also possible observational consequences explored.

If one rejects the occurrence of a bounce after the equilibrium point, one can conclude that σ/θ has been greater than 7% in the past if $(\omega/\theta)_0 > 4\%$. Of course, if one assumes from the beginning that $\sigma/\theta \ll 1$ during the matter era, this necessarily implies $(\omega/\theta)_0 < 4\%$ from a theoretical standpoint. Let us remark that direct observations[11] give poor upper limits on the relative rotation, $(\omega/\theta)_0 \lesssim 1$, and indirect bounds obtained through measurements of the cosmic background radiation (anisotropies of the observed temperature) are inferred with the use of approximations to Friedmann models.[12] However, our analysis leads one to wonder if the standard picture of the universe, i.e., a perturbed Friedmann model, is the only one allowed by observations. In principle, the use of general cosmological solutions that may not differ from the Friedmann ones from an observational standpoint, the metric being quite different, is an open possibility from a theoretical and observational point of view.

ACKNOWLEDGMENTS

The author is indebted to the referee for useful suggestions and expresses his gratitude to Dr. M. A. H. MacCallum for helpful comments given during a short stay at the Department of Applied Mathematics, Queen Mary College, London, Oct. 1981.

[1] A. Friedmann, Z. Phys. **10**, 377 (1922).
[2] H. P. Robertson, Astrophys. J. **82**, 284 (1935).
[3] A. G. Walker, Proc. London Math. Soc. (2) **42**, 90 (1936).
[4] G. F. R. Ellis, in "General Relativity and Cosmology," *Proceedings of the International School of Physics "Enrico Fermi,"* Course 47, edited by R. K. Sachs (Academic, New York, 1971).
[5] S. W. Hawking, Astrophys. J. **145**, 544 (1966).
[6] K. Gödel, Proc. Int. Cong. Math. (Cambridge, Mass.) **1**, 175 (1952).
[7] O. Heckmann, Astron. J. **66**, 599 (1961).
[8] P. Yodzis, Proc. R. Irish Acad. Sect. A **74**, 61 (1974).
[9] M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U. P., Princeton, N. J., 1975) p. 56.
[10] G. Ganow, Nature **158**, 549 (1946).
[11] J. Kristian and R. K. Sachs, Astrophys. J. **145**, 379 (1966).
[12] C. B. Collins and S. W. Hawking, Mon. Not. R. Astron. Soc. **162**, 307 (1973).
[13] A. Raychaudhuri, Phys. Rev. **98**, 1123 (1955).
[14] A. Sandage, Astrophys. J. **197**, 265 (1975).
[15] P. J. E. Peebles, *Physical Cosmology* (Princeton U. P., Princeton, N. J., 1971).
[16] G. A. Tammann, A. Sandage, and A. Yahil, in *Physical Cosmology*, Les Houches Summer School, edited by R. Balian *et al.* (North-Holland, Amsterdam, 1980), p. 119.
[17] F. J. Tipler, Ann. Phys. (N.Y.) **108**, 1 (1977).
[18] C. J. S. Clarke, Gen. Relativ. Gravit. **10**, 999 (1979).
[19] F. J. Tipler, C. J. S. Clarke, and G. F. R. Ellis, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2.
[20] S. W. Hawking and G. F. R. Ellis, Astrophys. J. **152**, 25 (1968).

Erratum: Interpolation theory and refinement of nested Hilbert spaces [J. Math. Phys. 22, 2489 (1981)]

J. -P. Antoine
Institut de Physique Théorique, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium
W. Karwowski[a)]
Centre de Physique Théorique, CNRS, Luminy, F-13288 Marseille, Cedex 2, France

(Received 7 May 1982; accepted for publication 14 May 1982)

PACS numbers: 02.30.Sa, 02.30.Tb, 99.10. + g

The following references were omitted:

[7] A. Grossmann, "Elementary Properties of Nested Hilbert Spaces," Commun. Math. Phys. **2**, 1–30 (1966).

[8] E. Nelson, "Construction of quantum fields from Markoff fields," J. Funct. Anal. **12**, 97–112 (1973).

[9] E. Nelson, "The free Markoff field," J. Funct. Anal. **12**, 211–227 (1973).

[10] W. Karwowski, "On Borchers class of Markoff fields," Proc. Camb. Phil. Soc. **76**, 457–463 (1974).

[11] E. Nelson, "Analytic vectors," Ann. Math. **70**, 572–615 (1959); R. Goodman, "One parameter groups generated by operators in an enveloping algebra," J. Funct. Anal. **6**, 218–236 (1970); B. Nagel, "Generalized eigenvectors in group representations," in *Studies in Mathematical Physics*, edited by A. O. Barut (Reidel, Dordrecht, 1973), pp. 135–154.

Erratum: Ground state energy bounds for potentials $|x|^{\nu}$ [J. Math. Phys. 23, 64 (1982)]

R. E. Crandall and Mary Hall Reno
Department of Physics, Reed College, Portland, Oregon 97202

(Received 16 March 1982; accepted for publication 16 April 1982)

PACS numbers: 03.65.Ge, 02.30.Hq, 02.60.Lj, 99.10. + g

In Sec. III, the ground state estimate for the quartic potential should read 1.060 362 090 484 1820···. The original text is in error at the sixth decimal digit.

Erratum: Linearization stability of Einstein equations coupled with self-gravitating scalar fields [J. Math. Phys. 22, 343 (1981)]

R. V. Saraykar and N. E. Joshi
Department of Mathematics, Nagpur University, University Campus, Nagpur-440 010, India

(Received 18 May 1982; accepted for publication 28 May 1982)

PACS numbers: 04.40. + c, 99.10. + g

$T_{\mu\nu}$ should be $T_{\mu\nu} = 2\beta(2\phi_{,\mu}\phi_{,\nu} - g_{\mu\nu}(\phi_{,\rho}\phi^{,\rho} + m^2\phi^2))$. The remarks in the introductory paragraph regarding π^0 mesons and C fields are not valid, as the linearization stability theorem is true only for massless scalar fields. Thus Brans–Dicke scalar fields are covered. σ should be $+4\beta\gamma\mu_g$ and $\mathscr{H}_F$ should be $2\beta(\gamma^2 + A(\phi))\mu_g$ (positivity of energy).

In Eq. (4), in the first coordinate, the sign before $\beta N(2\widetilde{\phi} - gA(\phi))\mu_g$ should be negative, whereas in the second coordinate, the signs of both the terms should be reversed. The last coordinate should be $+$. Corresponding changes should be made in Eqs. (7) and (9). In the expression for $D_g\,\mathscr{H}_F \cdot h$ before Eq. (10), the sign of $\beta(2\widetilde{\phi} - gA(\phi)) \cdot h$ should be negative. A corresponding change should be made in the expression before Eq. (10) and Eq. (10) itself; and similarly in Eq. (11), and before it. Signs in the expression for $D_\phi\,\mathscr{H}_F \cdot \psi$ should be reversed. Sign changes as mentioned above should be made in the equations before Eqs. (13) and (14).

These cause further changes in the calculations in the proof of linearization stability: Equations (17) and (18) should be written according to Eq. (4). With corresponding changes in further calculations, Eq. (21) now reads

$$(\Delta N)\mu_g + \beta N(2\gamma^2 - m^2\phi^2)\mu_g - \tfrac{1}{2}\,\mathrm{tr}(L_X\,\pi) = 0.$$

Corresponding changes are in order in later expressions. Thus we can conclude "N is constant" only if $m = 0$. In other words, linearization stability is implied only for massless scalar fields. This conclusion is consistent with the fact that the energy–momentum tensor for the scalar fields satisfies the physically reasonable strong energy condition only when $m = 0$. If $m \neq 0$, as in Ref. 1 of this erratum, $T_{ab}W^aW^b - \tfrac{1}{2}W_aW^aT = (\phi_{,a}W^a)^2 - \tfrac{1}{2}m^2\phi^2$. Thus, $\phi_{,a}W^a$ is exactly γ when $W^a = Z^a_\Sigma$, the forward pointing unit timelike vector normal to the hypersurface Σ [$Z^a_\Sigma = (1/N, X^i/N)$ in terms of lapse and shift]. This can be shown easily by using the evolution equation for ϕ. Thus the right-hand side of the above equation is precisely $\gamma^2 - \tfrac{1}{2}m^2\phi^2$ which is involved in the elliptic equation (21). However, for π mesons ($m \neq 0$), although the energy–momentum tensor may not satisfy the strong energy condition at every point, this does not affect the physically reasonable convergence of timelike geodesics over distances greater than 10^{-12} cm. (For detailed argument see Ref. 1.) Thus, expecting linearization stability for coupled gravitational and massive scalar fields would be physically unreasonable.

The other theorems in the paper are valid for $m \neq 0$ since they are consequences of the evolution equations written in the adjoint form.

[1] S. Hawking and G. Ellis, *Large Scale Structure of Space-Time* (Cambridge U. P., New York, 1973), pp. 95 and 96.