A unified theory of the point groups. V. The general projective corepresentations of the magnetic point groups and their applications to magnetic space groups

Shoon K. Kim
Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122

(Received 11 March 1983; accepted for publication 29 July 1983)

This paper presents the general expressions of the projective irreducible corepresentations of the 20 homologous sets of the magnetic double point groups from which follow all those of the remaining groups of finite order through isomorphisms (the icosahedral group is excluded). These are explicitly given in terms of the irreducible representations of the unitary double point groups. Their special cases provide all the irreducible corepresentations of all magnetic space groups of wave vector through simple gauge transformations. A method of determining the gauge transformations is discussed through typical examples.

PACS numbers: 02.20. + b

1. INTRODUCTION

In a series of papers[1–4] (referred to as I–IV) we have developed a theory of representation for the point groups through their double groups regarding them as subgroups of SU(2). With use of the simple subgroup conditions we have constructed the general expressions of the unirreps (unitary irreducible representations) for the double point groups C_n, D_n, T, and O. These four sets of the proper point groups and their direct product groups with the group of inversion C_i constitute a characteristic set of the finite double point groups; any one of the remaining double point groups is isomorphic to one of them (here and hereafter the icosahedral group is excluded). Through the vector unirreps of their representation groups[5–7] we have constructed all the projective unirreps of all double point groups based on the modified theory of the induced representation in II. Upon introducing a new system of classification for the improper point groups $\overline{G}$ which is best suitable for describing their isomorphisms we have described the basis functions of $\overline{G}$ in III by the angular momentum eigenfunctions in the general manner as in the cases of the proper point groups. Extending the system of classification for the Shubnikov (or magnetic or antiunitary) point groups G^s in IV we have also constructed all the vector counirreps (irreducible unitary corepresentations) of all G^s from those of 10 G^s through isomorphisms.

In the present work we shall construct the general expressions of the projective counirreps of all G^s of finite order. This will be achieved via the vector counirreps of the representation groups of a characteristic set of 20 G^s with full use of the isomorphisms. The projective counirreps of G^s presently available are limited to the crystallographic point groups and are given by the counirreps of the magnetic space groups of wave vector $M(k)$ in terms of the actual matrices.[8] We are seeking the general expressions of the projective counirreps common to all isomorphic point groups. Their special cases then provide all the counirreps of all $M(k)$ through simple gauge transformations. The present result will greatly simplify the existing matrix representations[8] of $M(k)$ and thereby reduce the labor involved in their applications.

In Sec. 2, we shall first discuss the isomorphisms of the representation groups $G^{s\prime}$ based on the new system of classification introduced in III and IV. Then, we shall construct $G^{s\prime}$ for the characteristic set of 20 G^s mentioned above. In Sec. 3, we shall first describe briefly the modified theory of corepresentation introduced in IV and then construct the general expressions of the projective counirreps of G^s via the vector counirreps of $G^{s\prime}$. In the final section, we shall illustrate how to determine the gauge transformations which connect the counirreps of the magnetic space groups of wave vector $M(k)$ and those of the corresponding $G^{s\prime}$ determined in this work.

2. THE REPRESENTATION GROUPS OF THE MAGNETIC POINT GROUPS

Before constructing the representation groups $G^{s\prime}$ for the magnetic point groups G^s, we shall first discuss their isomorphisms in terms of the new system of classification of G^s introduced in IV and thus minimize the number of $G^{s\prime}$ to be constructed. Here and hereafter, we mean by a point group its double group unless otherwise specified.

Let H be the halving subgroup of G^s, then

$$G^s = H + aH, \quad a = \theta z, \tag{2.1}$$

where a is an antiunitary operator defined by the time inversion θ and a unitary operator z, which is compatible with H. Following the new system of notations introduced in III and IV, we denote G^s defined above by H^z. Then, the one-to-one correspondence $\theta\hat{i}\leftrightarrow\theta$ gives the following isomorphisms[4] ($\simeq$):

$$H^i \simeq H^e, \quad H^p \simeq H^q, \quad H^v \simeq H^u \tag{2.2a}$$

while the one-to-one correspondences $\bar{c}_n \leftrightarrow c_n$ and $\bar{c}_2' \leftrightarrow c_2'$ yield

$$C^z_{np} \simeq C^z_{2n}, \quad D^z_{np} \simeq D^z_{2n}, \quad T^z_p \simeq O^z, \quad C^z_{nv} \simeq D^z_n. \tag{2.2b}$$

On account of these isomorphisms, a characteristic set of G^s may be given by H^e, H^q, and H^u, where H is a proper point group P or a direct product group $P_i = P\times C_i$ of P and the group of inversion C_i. It should be noted here[9] that $P_i^{z\prime}$

 0022-2488/84/020189-08$02.50

TABLE I. The representation groups of the unitary and antiunitary point groups (finite order).

1.	C_n':	$x^n = \bar{e}, \quad \bar{e}^2 = e$
2.	$C_n^{e\prime}$	$C_n' \quad xa = ax, \ a^2 = \tau\bar{e} \ \tau^2 = e$ ($\tau = 1$, if n is odd)
3.	$C_n^{q\prime}$:	C_n': $xa = ax, \ a^2 = \tau\bar{e}x; \ \tau^2 = e$ ($\tau = 1$, if n is even)
4.	$C_n^{u\prime}$:	C_n'; $xax = \xi a, \ a^2 = \tau e; \ \xi^2 = \tau^2 = e$ ($\xi = 1$, if n is odd)
5.	C_{ni}':	$x^n = \bar{e}, \ \bar{e}^2 = e, \ x\hat{\imath} = \beta\hat{\imath}x, \ \hat{\imath}^2 = e; \ \beta^2 = e$ ($\beta = 1$, if n is odd)
6.	$C_{ni}^{e\prime}$:	$C_{ni}'(\beta)$: $xa = ax, \ \hat{\imath}a = \zeta a\hat{\imath}, \ a^2 = \tau\bar{e}; \ \zeta^2 = \tau^2 = e$ ($\beta = \tau = 1$, if n is odd)
7.	$C_{ni}^{q\prime}$:	$C_{ni}'(\beta = 1)$; $xa = ax, \ \hat{\imath}a = \zeta a\hat{\imath}, \ a^2 = \tau\bar{e}x; \ \zeta^2 = \tau^2 = e$ ($\tau = 1$, if n is even)
8.	$C_{ni}^{u\prime}$:	$C_{ni}'(\beta)$;$xax = \xi a, \ \hat{\imath}a = \zeta a\hat{\imath}, \ a^2 = \tau e; \ \xi^2 = \zeta^2 = \tau^2 = e$ ($\xi = 1$, if n is odd)
9.	D_n':	$x^n = y^2 = (xy)^2 = \bar{e}, \ \bar{e}^2 = e$
10.	$D_{2r}^{e\prime}$:	D_{2r}'; $xa = \xi ax, \ ya = \eta ay, \ a^2 = \tau\bar{e}; \ \xi^2 = \eta^2 = \tau^2 = e$
11.	$D_{2r+1}^{e\prime}$:	D_{2r+1}'; $xa = ax, \ ya = ay, \ a^2 = \tau\bar{e}; \ \tau^2 = e$
12.	$D_{2r}^{q\prime}$:	D_{2r}'; $xa = ax, \ ya = ayx, \ a^2 = \tau\bar{e}x; \ \tau^2 = e$
13.	$D_{2r+1}^{q\prime}$:	D_{2r+1}': $xa = ax, \ ya = \eta ayx, \ a^2 = \tau\bar{e}x; \ \eta^2 = \tau^2 = e$
14.	D_{ni}':	$x^n = y^2 = (xy)^2 = \bar{e}, \ \bar{e}^2 = e, \ x\hat{\imath} = \beta\hat{\imath}x, \ y\hat{\imath} = \gamma\hat{\imath}y, \ \hat{\imath}^2 = e;$ $\beta^2 = \gamma^2 = e$ ($\beta = 1$, if n is odd)
15.	$D_{2r,i}^{e\prime}$:	$D_{2r,i}'$ (β, γ); $xa = \xi ax, \ ya = \eta ay, \ \hat{\imath}a = \zeta a\hat{\imath}, \ a^2 = \tau\bar{e};$ $\xi^2 = \eta^2 = \zeta^2 = \tau^2 = e$
16.	$D_{2r+1,i}^{e\prime}$:	$D_{2r+1,i}^{q\prime}(\gamma)$; $xa = ax, \ ya = ay, \ \hat{\imath}a = \zeta a\hat{\imath}, \ a^2 = \tau\bar{e};$ $\zeta^2 = \tau^2 = e$
17.	$D_{2r,i}^{q\prime}$:	$D_{2r,i}'(\beta = 1, \gamma)$; $xa = ax, \ ya = ayx, \ \hat{\imath}a = \zeta a\hat{\imath}, \ a^2 = \tau\bar{e}x;$ $\zeta^2 = \tau^2 = e$
18.	$D_{2r+1,i}^{q\prime}$:	$D_{2r+1,i}'(\gamma)$; $xa = ax, \ ya = \eta ayx, \ \hat{\imath}a = \zeta a\hat{\imath}, \ a^2 = \tau\bar{e}x;$ $\eta^2 = \xi^2 = \tau^2 = e$
19.	T':	$x^2 = y^3 = \bar{e}, \ (xy)^3 = e, \ \bar{e}^2 = e$
20.	$T^{e\prime}$:	T'; $xa = ax, \ ya = ay, \ a^2 = \tau\bar{e}; \ \tau^2 = e$
21.	$T^{q\prime}$:	T'; $xa = ax, \ yay = ax, \ a^2 = \tau\bar{e}x; \ \tau^2 = e$
22.	T_i':	$x^2 = \bar{e}, \ \bar{e}^2 = e, \ s^6 = e, \ xs^2 = (sx)^2$
23.	$T_i^{e\prime}$:	T_i'; $xa = ax, \ sa = \eta as, \ a^2 = \tau\bar{e}, \ \bar{e}^2 = e; \ \eta^2 = \tau^2 = e$
24.	$T_i^{q\prime}$:	T_i'; $xa = ax, \ sas = \eta ax, \ a^2 = \tau\bar{e}x; \ \eta^2 = \tau^2 = e$
25.	O':	$x^4 = y^3 = (xy)^2 = \bar{e}, \ \bar{e}^2 = e$
26.	$O^{e\prime}$:	O'; $xa = \xi ax, \ ya = ay, \ a^2 = \tau\bar{e}; \ \xi^2 = \tau^2 = e$
27.	O_i':	$x^4 = y^3 = (xy)^2 = \bar{e}, \ \bar{e}^2 = e, \ x\hat{\imath} = \beta\hat{\imath}x, \ y\hat{\imath} = \hat{\imath}y, \ \hat{\imath}^2 = e;$ $\beta^2 = e$
28.	$O_i^{e\prime}$:	$O_i'(\beta)$; $xa = \xi ax, \ ya = ay, \ \hat{\imath}a = \zeta a\hat{\imath}, \ a^2 = \tau\bar{e};$ $\xi^2 = \zeta^2 = \tau^2 = e$

(i) e = identity element, $\bar{e} = 2\pi$ rotation, $\hat{\imath}$ = inversion.
(ii) The second order elements β, γ, ξ, η, ζ, τ, and $\bar{e}$ are all in the center of the respective $G^{s\prime}$.

$\not\simeq P^{z\prime} \times C_i$ in general even though $P_i^z \simeq P^z \times C_i$. The representation groups $G^{s\prime}$ of a characteristic set of 20 G^s are given in Table I (their construction will be discussed shortly). These are described by the defining relations of the abstract group generators, which are common to all groups isomorphic to each other. Possible realizations of these generators of the respective double group may be made through (2.13) of I. To simplify the tabulation, we have also included in the table the representation groups of the characteristic set of eight unitary groups determined in II; these are then used to describe the representation groups $G^{s\prime}$. Thus, Table I provides all the representation groups of all finite point groups (unitary or antiunitary) through isomorphisms (icosahedral groups are excluded).

Construction of any of these representation groups $G^{s'}$ is straightforward as in the cases[2,7] of those of the unitary point groups except that the operator a is antiunitary. We start from the most general form of the projective counirrep of the defining relations for the generators $\{x_j, a\}$ of a given antiunitary group G^s, where x_j are the generators of H. Then, through gauge transformations and the inner automorphisms (in particular, with respect to the antiunitary operator a) we arrive at the minimum set of the independent coefficients (or projective factors) $\{\alpha_i;\ i=1,2,...,\nu\}$ for each G^s, all of which can be shown to satisfy the same quadratic equation:

$$\alpha_i^2 = 1, \quad i = 1, 2, ..., \nu. \tag{2.3}$$

This is completely analogous to the case of the unitary groups.[2,7] These coefficients are given by the Greek symbols β, γ, ξ, η, ζ, and τ in Table I. Now, we regard these coefficients as the second order generators which commute with each other and with all the elements of G^s. Then, we arrive at the representation group $G^{s'}$ described by a set of the defining relations of the generators $\{x_j, a, \alpha_i\}$ as given in Table I.

By construction, every projective counirrep of G^s follows from the vector counirreps of $G^{s'}$ up to p-equivalence. Each representation of the coefficient set $\{\alpha_i\}$ in $G^{s'}$ defines a class of factor systems for G^s. Thus, the number of p-inequivalent classes (or order of the multiplicator) for a given G^s equals 2^ν, where ν is the number of the coefficients in $G^{s'}$. According to Table I, the maximum number of classes for a given G^s equals 64 ($=2^6$) which occurs for $D^{e'}_{2r,i}$, while the minimum number equals 1 which occurs for $C^{e'}_{2r+1}$ or $C^{q'}_{2r}$. According to the present classification there exists a total of $180\,p$-inequivalent classes for all G^s while there exists a total of $13\,p$-inequivalent classes for the unitary point groups. These are all of finite order and the icosahedral groups are excluded.

3. THE PROJECTIVE COUNIRREPS OF THE MAGNETIC POINT GROUPS

We shall construct all the counirreps of the representation groups $G^{s'}$ listed in Table I based on the modified theory of corepresentations developed in IV. For convenience, we may reproduce here some of the basic results of the theory and introduce small modification for the notations of the counirreps.

Let $H(G^{s'})$ be the halving subgroup of $G^{s'}$ and $\{\Delta^\nu(h')\}$ be a complete set of the unirreps of $h' \in H(G^{s'})$. Then, there exists a unitary matrix such that

$$\Delta^\nu(a^{-1}h'a)^* = N(a)^{-1}\Delta^\mu(h')N(a) \tag{3.1}$$

for all $h' \in (G^{s'})$. Based on this, one can show that there exist three types of counirreps.

Case (*a*): $\nu=\mu$, $N(a)N(a)^* = \Delta^\nu(a^2)$: There exist two equivalent counirreps given by

$$S^{(\nu\pm)}(h') = \Delta^\nu(h'), \quad S^{(\nu\pm)}(a) = \pm N(a). \tag{3.2a}$$

Either one of them provides the required counirrep.

Case (*b*): $\nu=\mu$, $N(a)N(a)^* = -\Delta^\nu(a^2)$: The counirrep is given by

$$S^{(\nu,\nu)}(h') = \begin{bmatrix}\Delta^\nu(h') & 0\\ 0 & \Delta^\nu(h')\end{bmatrix},$$

$$S^{(\nu,\nu)}(a) = \begin{bmatrix}0 & -N(a)\\ N(a) & 0\end{bmatrix}. \tag{3.2b}$$

Case (*c*): $\nu\neq\mu$: The counirrep is given by

$$S^{(\nu,\mu)}(h') = \begin{bmatrix}\Delta^\nu(h') & 0\\ 0 & \Delta^\mu(h')\end{bmatrix},$$

$$S^{(\nu,\mu)}(a) = \begin{bmatrix}0 & \Delta^\nu(a^2)N(a)^{-1*}\\ N(a) & 0\end{bmatrix}. \tag{3.2c}$$

These three types of counirreps are denoted by the following notations:

$$\begin{aligned} &S(\Delta^\nu; N) && \text{for } S^{(\nu+)} \text{ of (3.2a)} \in \text{Case (a)},\\ &S(\Delta^\nu, \Delta^\nu; N) && \text{for } S^{(\nu,\nu)} \text{ of (3.2b)} \in \text{Case (b)},\\ &S(\Delta^\nu, \Delta^\mu; N) && \text{for } S^{(\nu,\mu)} \text{ of (3.2c)} \in \text{Case (c)}.\end{aligned} \tag{3.3}$$

It is noted that the notation S is introduced here in the place of D used previously in IV. This is to avoid the confusion with the notations for the induced unirreps introduced in II,

$$D^{(\nu,\mu)} = D(\Gamma^\nu, \Gamma^\mu; N), \quad D^{(\nu\pm)} = D(\Gamma^\nu; \pm N), \tag{3.4}$$

for a unitary group augmented by a unitary operator A. Here $D^{(\nu,\mu)}$ has a similar structure as that of $S^{(\nu,\mu)}$ except for $N(A)$ in the place of $N(a)^*$ in $S^{(\nu,\mu)}$; and $D^{(\nu,\nu)}$ always reduces to two inequivalent unirreps $D^{(\nu+)}$ and $D^{(\nu-)}$ [see Eqs. (6) and (7) in II]. These notations will also be used later in describing the unirreps Δ^ν of $H(G^{s'})$.

The counirreps of $G^{s'}$ given in (3.2) are completely explicit in terms of the unirreps $\Delta^\nu(h')$ of $h' \in H(G^{s'})$ except for the transformation matrix $N(a)$ and $\Delta^\nu(a^2)$. When Δ^ν is one-dimensional, one can take $N(a)=1$. In such a trivial case we shall delete N from (3.3). For a higher-dimensional case, one determines $N(a)$ from (3.1) for each class of factor systems by using the defining relations of the group generators given in Table I. Let us illustrate the determination of $N(a)$ for a representation group $C^{u'}_{n,i}$ given in Table I(8). It has seven generators $\{x, \hat{i}, a; \beta; \xi, \zeta; \tau\}$, of which $x, \hat{i}, \beta$ are the generators of the representation group $C'_{n,i}$ of $C_{n,i}$ as given in Table I(5). The halving subgroup $H(C^{u'}_{n,i})$ is defined by a direct product group $C'_{n,i}\times C_\xi \times C_\zeta \times C_\tau$, where $C_\xi = \{e, \xi\}$, $C_\zeta = \{e, \zeta\}$, and $C_\tau = \{e, \tau\}$. The transformation matrix $N(a)$ is determined from

$$\begin{aligned}\xi\Delta^\nu(x^{-1})^* &= N(a)^{-1}\Delta^\mu(x)N(a),\\ \zeta\Delta^\nu(\hat{i})^* &= N(a)^{-1}\Delta^\mu(\hat{i})N(a)\end{aligned} \tag{3.5}$$

for a given set of values of the coefficients ξ and ζ. Calculation of $N(a)$ is straightforward. For almost all $G^{s'}$, the transformation matrices are expressed by the unit matrix 1_d with appropriate dimension d or the Pauli spin matrices $(\sigma_x, \sigma_y, \sigma_z)$ or their direct products except for a few cases where the antiunitary operator a involves the operator q ($=c_{2n}$) [see, for example, Table II(12)]. Finally, from $\Delta^\nu(a^2) = \tau\Delta^\nu(\bar{e}z^2)$, we have for $C^{u'}_{n,i}$

TABLE II. The projective unirreps and counirreps of the unitary and antiunitary point groups (finite order).

1. $C_n(K^0)$:
$K^0;\ M_m;\ m=m^0,\ -\frac{1}{2}n<m^0\leqslant\frac{1}{2}n$

2. $C^e_n(K;\ \tau=1$, if n is odd):
$K;\ S(M_0),\ S(M_{n_e/2}),\ S(M_{n_o/2},M_{n_o/2}),\ S(M_m,M_{-m})$;
$m=m^*=\frac{1}{2},1,\dots,\frac{1}{2}(n-1)$

3. $C^q_n(K;\ \tau=1$, if n is even):
$K;\ S(M_0),\ (M_{n_e/2},\ M_{n_e/2}),\ S(M_{n_o/2}),\ S(M_m,M_{-m})$;
$m=m^*$

4. $C^u_n(K_t;\ t=\{\xi\};\ \xi=1$ if n is odd):
$K_1;\ S(M_m);\ m=m^0$
$K_2(n=2r);\ S(M_m,M_{m-r});\ m=m'=\frac{1}{2},1,\dots,r$

5. $C_{ni}(K^0_s;\ s=\{\beta\};\ \beta=1,$ if n is odd):
$K^0_1;\ C_n(K^0)\times C_i;\ M^\pm_m;\ m=m^0$
$K^0_2(n=2r);\ D_m=D(M_m,M_{m-r});\ m=m'$

6. $C^e_{ni}(K_{st};\ s=\{\beta\},\ t=\{\zeta\};\ \beta=\tau=1$, if n is odd):
$K_{11};\ C^e_n(K)\times C_i$
$K_{12};\ S(M^+_m,M^-_{-m});\ m=m^0$
$K_{21}(n=2r);\ S(D_r;1_2),\ S(D_{r_e/2};\sigma_x),\ S(D_{r_o/2},D_{r_o/2};\sigma_x)$,
$S(D_m,D_{r-m};\sigma_x);\ m=m^\dagger=\frac{1}{2},1,\dots,\frac{1}{2}(r-1)$
$K_{22}(n=2r);\ S(D_r;\sigma_z),\ S(D_{r_e/2},D_{r_e/2};\sigma_y),\ S(D_{r_o/2};\sigma_y)$
$S(D_m,D_{r-m};\sigma_y);\ m=m^\dagger$

7. $C^q_{ni}(K_t;\ t=\{\zeta\};\ \tau=1$, if n is even)
$K_1;\ C^q_n(K)\times C_i$
$K_2;\ S(M^+_m,M^-_{-m});\ m=m^0$

8. $C^u_{ni}\ (K_{st};\ s=\{\beta\},\ t=\{\xi,\zeta\};\ \xi=1$, if n is odd):
$K_{11};\ C^u_n(K_1)\times C_i$
$K_{12};\ S(M^+_m,M^-_m);\ m=m^0$
$K_{13}(n=2r);\ C^u_{2r}(K_2)\times C_i$
$K_{14}(n=2r);\ S(M^+_m,M^-_{m-r});\ m=m^0,\ -r<m^0\leqslant r$
$K_{21}(n=2r);\ S(D_m;1_2);\ m=m'=\frac{1}{2},1,\dots,r$
$K_{22}(n=2r);\ S(D_m;\sigma_z);\ m=m'$
$K_{23}(n=2r);\ S(D_m;\sigma_x);\ m=m'$
$K_{24}(n=2r);\ S(D_m,D_m;\sigma_y);\ m=m'$

9. $D_n(K^0)$:
$K^0;\ A_1,A_2,B_1,B_2,E_m;\ m=m^*=\frac{1}{2},1,\dots,\frac{1}{2}(n-1)$

10. $D^e_{2r}(K_t;\ t=\{\xi,\eta\})$
$K_1;\ S(A_1),S(A_2),\ S(B_1),\ S(B_2),\ S(E_m;1_2,\sigma_y)$;
$m=m^*=\frac{1}{2},1,\dots,(r-\frac{1}{2})$
$K_2;\ S(A_1,A_2),\ S(B_1,B_2),\ S(E_m,E_m;\sigma_y,1_2);\ m=m^*$
$K_3;\ S(A_1,B_2),\ S(A_2,B_1),\ S(E_{r_e/2};\sigma_z),\ S(E_{r_o/2},E_{r_o/2};\sigma_x)$
$S(E_m,E_{r-m};\sigma_z,\sigma_x);\ m=m^\dagger=\frac{1}{2},1,\dots,\frac{1}{2}(r-1)$
$K_4;\ S(A_1,B_1),\ S(A_2,B_2),\ S(E_{r_e/2};\sigma_x),\ S(E_{r_o/2},E_{r_o/2};\sigma_z)$
$S(E_m,E_{r-m};\sigma_x,\sigma_z);\ m=m^\dagger$

11. $D^e_{2r+1}(K)$:
$K;\ S(A_1),\ S(A_2),\ S(B_1,B_2),\ S(E_m;1_2,\sigma_y);\ m=m'=\frac{1}{2},1,\dots,r$

12. $D^q_{2r}(K)$:
$K;\ S(A_1),\ S(A_2),\ S(B_1,B_2),\ S(E_m;R_m,\sigma_yR_m);\ m=m^*=\frac{1}{2},1,\dots,(r-\frac{1}{2})$

13. $D^q_{2r+1}(K_t;\ t=\{\eta\})$
$K_1;\ S(A_1),\ S(A_2),\ S(B_1),\ S(B_2),\ S(E_m;R_m,\sigma_yR_m);\ m=m'$
$K_2;\ S(A_1,A_2),\ S(B_1,B_2),\ S(E_m,E_m;\sigma_yR_m,R_m);\ m=m'$

14. $D_{ni}(K^0_s;\ s=\{\beta,\gamma\};\ \beta=1$, if n is odd)
$K^0_1;\ D_n(K^0)\times C_i;\ A^\pm_1,A^\pm_2,B^\pm_1,B^\pm_2,E^\pm_m$;
$m=m^*=\frac{1}{2},1,\dots,\frac{1}{2}(n-1)$
$K^0_2;\ D_A=D(A_1,A_2),\ D_B=D(B_1,B_2),\ D^{\pm y}_m=D(E_m;\ \pm\sigma_y);\ m=m^*$
$K^0_3(n=2r);\ D_{12}=D(A_1,B_2),\ D_{21}=D(A_2,B_1),\ D^{\pm z}_{r/2}=D(E_{r/2};\ \pm\sigma_z)$
$D^z_{m,r-m}=D(E_m,E_{r-m};\sigma_z);\ m=m^\dagger=\frac{1}{2},1,\dots,\frac{1}{2}(r-1)$
$K^0_4(n=2r);\ D_{11}=D(A_1,B_1),\ D_{22}=D(A_2,B_2),\ D^{\pm x}_{r/2}=D(E_{r/2};\ \pm\sigma_x)$;
$D^x_{m,r-m}=D(E_m,E_{r-m};\sigma_x);\ m=m^\dagger$

15. $D^e_{2r,i}(K_{st};\ s=\{\beta,\gamma\},\ t=\{\xi,\eta,\zeta\})$:
$K_{11};\ D^e_{2r}(K_1)\times C_i$
$K_{12};\ S(A^+_1,A^-_1),\ S(A^+_2,A^-_2),\ S(B^+_1,B^-_1),\ S(B^+_2,B^-_2)$,
$S(E^+_m,E^-_m,1_2,\sigma_y),\ m=m^*=\frac{1}{2},1,\dots,r-\frac{1}{2}$
$K_{13};\ D^e_{2r}(K_2)\times C_i$
$K_{14};\ S(A^\pm_1,A^\mp_2),\ S(B^\pm_1,B^\mp_2),\ S(E^+_m,E^-_m;\sigma_y,1_2)$: $m=m^*$
$K_{15};\ D^e_{2r}(K_3)\times C_i$
$K_{16};\ S(A^\pm_1,B^\mp_2),\ S(A^\pm_2,B^\mp_1),\ S(E^+_m,E_{r-m};\sigma_z,\sigma_x);\ m=m^*$
$K_{17};\ D^e_{2r}(K_4)\times C_i$
$K_{18};\ S(A^\pm_1,B^\mp_1),\ S(A^\pm_2,B^\mp_2),\ S(E^+_m,E_{r-m};\sigma_x,\sigma_z);\ m=m^*$

$K_{21};\ S(D_A;1_2),\ S(D_B;1_2),\ S(D^{+y}_m,D^{-y}_m;1_2,\sigma_y);\ m=m^*$
$K_{22};\ S(D_A;\sigma_z),\ S(D_B;\sigma_z),\ S(D^{\pm y}_m;1_2,\sigma_y);\ m=m^*$
$K_{23};\ S(D_A;\sigma_x),\ S(D_B;\sigma_x),\ S(D^{+y}_m,D^{-y}_m;\sigma_y,1_2);\ m=m^*$
$K_{24};\ S(D_A,D_A;\sigma_y),\ S(D_B,D_B;\sigma_y),\ S(D^{\pm y}_m,D^{\pm y}_m;\sigma_y,1_2);\ m=m^*$
$K_{25};\ S(D_A,D_B;\sigma_x),\ S(D^{\pm y}_{r_e/2};\sigma_z),\ S(D^{\pm y}_{r_o/2},D^{\pm y}_{r_o/2};\sigma_x)$,
$S(D^{\pm y}_m,D^{\pm y}_{r-m};\sigma_z,\sigma_x);\ m=m^\dagger=\frac{1}{2},1,\dots,\frac{1}{2}(r-1)$
$K_{26};\ S(D_A,D_B;\sigma_y),\ S(D^{+y}_m,D^{-y}_{r-m};\sigma_z,\sigma_x);\ m=m^*$
$K_{27};\ S(D_A,D_B;1_2),\ S(D^{\pm y}_{r_e/2};\sigma_x),\ S(D^{\pm y}_{r_o/2},D^{\pm y}_{r_o/2};\sigma_z)$
$S(D^{\pm y}_m,D^{\pm y}_{r-m};\sigma_x,\sigma_z);\ m=m^\dagger$
$K_{28};\ S(D_A,D_B;\sigma_z),\ S(D^{+y}_m,D^{-y}_{r-m};\sigma_x,\sigma_z);\ m=m^*$

$K_{31};\ S(D_{12};1_2),\ S(D_{21};1_2),\ S(D^{\pm z}_{r_e/2};1_2),\ S(D^{+z}_{r_o/2},D^{-z}_{r_o/2};\sigma_y)$
$S(D^z_{m,r-m};1_4,\sigma_z\times\sigma_y);\ m=m^\dagger=\frac{1}{2},1,\dots,\frac{1}{2}(r-1)$
$K_{32};\ S(D_{12};1_2),\ S(D_{21};\sigma_z),\ S(D^{+z}_{r_e/2},D^{z}_{r_e/2};1_2),\ S(D^{\pm z}_{r_o/2};\sigma_y)$
$S(D^z_{m,r-m};\sigma_z\times1_2,1_2\times\sigma_y);\ m=m^\dagger$
$K_{33};\ S(D_{12},D_{21};1_2),\ S(D^{+z}_{r_e/2};D^{z}_{r_o/2};\sigma_y),\ S(D^{\pm z}_{r_o/2},D^{\pm z}_{r_o/2};1_2)$
$S(D^z_{m,r-m},D^z_{m,r-m};\sigma_z\times\sigma_y,1_4);\ m=m^\dagger$
$K_{34};\ S(D_{12},D_{21};\sigma_z),\ S(D^{\pm z}_{r_e/2},D^{\pm z}_{r_e/2};\sigma_y),\ S(D^{+z}_{r_o/2},D^{-z}_{r_o/2};1_2)$
$S(D^z_{m,r-m},D^z_{m,r-m};1_2\times\sigma_y,\sigma_z\times1_2);\ m=m^\dagger$
$K_{35};\ S(D_{12};\sigma_x),\ S(D_{21};\sigma_x),\ S(D^{\pm z}_{r_e/2};\sigma_z),\ S(D^{+z}_{r_o/2},D^{-z}_{r_o/2};\sigma_x)$
$S(D_{m,r-m};\sigma_x\times\sigma_z,\sigma_y\times\sigma_x);\ m=m^\dagger$
$K_{36};\ S(D_{12},D_{12};\sigma_y),\ S(D_{21},D_{21};\sigma_y),\ S(D^{+z}_{r_e/2},D^{-z}_{r_e/2};\sigma_z)$
$S(D^{\pm z}_{r_o/2},D^{+z}_{r_o/2};\sigma_x),\ S(D^z_{m,r-m},D^z_{m,r-m};\sigma_y\times\sigma_z,\sigma_x\times\sigma_z)$;
$m=m^\dagger$
$K_{37};\ S(D_{12},D_{21};\sigma_x),\ S(D^{+z}_{r_e/2},D^{-z}_{r_e/2};\sigma_x),\ S(D^{+z}_{r_o/2},D^{\pm z}_{r_o/2};\sigma_z)$
$S(D^z_{m,r-m},D^z_{m,r-m};\sigma_y\times\sigma_x,\sigma_x\times\sigma_z);\ m=m^\dagger$
$K_{38};\ S(D_{12},D_{21};\sigma_y),\ S(D^{\pm z}_{r_e/2};\sigma_x),\ S(D^{+z}_{r_o/2},D^{-z}_{r_o/2};\sigma_z)$
$S(D^z_{m,r-m};\sigma_x\times\sigma_x,\sigma_y\times\sigma_z);\ m=m^\dagger$

$K_{41};\ S(D_{11};1_2),\ S(D_{22};1_2),\ S(D^{\pm x}_{r_e/2};1_2),\ S(D^{+x}_{r_o/2},D^{-x}_{r_o/2};\sigma_y)$
$S(D^x_{m,r-m};1_4,\sigma_z\times\sigma_y);\ m=m^\dagger$
$K_{42};\ S(D_{11};\sigma_z),\ S(D_{22};\sigma_z),\ S(D^{+x}_{r_e/2},D^{-x}_{r_e/2};1_2),\ S(D^{\pm x}_{r_o/2};\sigma_y)$
$S(D^x_{m,r-m};\sigma_z\times1_2,1_2\times\sigma_y);\ m=m^\dagger$
$K_{43};\ S(D_{11},D_{22};1_2),\ S(D^{+x}_{r_e/2},D^{-x}_{r_e/2};\sigma_y),\ S(D^{\pm x}_{r_o/2};1_2)$
$S(D^x_{m,r-m},D^x_{m,r-m};\sigma_z\times\sigma_y,l_4);\ m=m^\dagger$
$K_{44};\ S(D_{11},D_{22};\sigma_z),\ S(D^{\pm x}_{r_e/2},D^{\pm x}_{r_e/2};\sigma_y),\ S(D^{+x}_{r_o/2},D^{x}_{r_o/2};1_2)$
$S(D^x_{m,r-m},D^x_{m,r-m};1_2\times\sigma_y,\sigma_z\times1_2);\ m=m^\dagger$
$K_{45};\ S(D_{11},D_{22};\sigma_x),\ S(D^{+x}_{r_e/2},D^{-x}_{r_e/2};\sigma_z),\ S(D^{\pm x}_{r_o/2},D^{+x}_{r_o/2};\sigma_x)$
$S(D^x_{m,r-m},D^x_{m,r-m};\sigma_y\times\sigma_z,\sigma_x\times\sigma_x);\ m=m^\dagger$
$K_{46};\ S(D_{11},D_{22};\sigma_y),\ S(D^{\pm x}_{r_e/2};\sigma_z),\ S(D^{+x}_{r_o/2},D^{-x}_{r_o/2};\sigma_x)$
$S(D^x_{m,r-m};\sigma_x\times\sigma_z,\sigma_y\times\sigma_x);\ m=m^\dagger$
$K_{47};\ S(D_{11};1_2),\ S(D_{22};\sigma_x),\ S(D^{\pm x}_{r_e/2};\sigma_x),\ S(D^{+x}_{r_o/2},D^{-x}_{r_o/2};\sigma_z)$,
$S(D^x_{m,r-m};\sigma_x\times\sigma_x,\sigma_y\times\sigma_z);\ m=m^\dagger$
$K_{48};\ S(D_{11},D_{11};\sigma_y),\ S(D_{22},D_{22};\sigma_y),\ S(D^{+x}_{r_e/2},D^{-x}_{r_e/2};\sigma_x)$
$S(D^{\pm x}_{r_o/2},D^{\pm x}_{r_o/2};\sigma_z),\ S(D^x_{m,r-m},D^x_{m,r-m};\sigma_y\times\sigma_x,\sigma_x\times\sigma_z)$;
$m=m^\dagger$

16. $D^e_{2r+1,i}(K_{st};\ s=\{\gamma\},\ t=\{\zeta\})$:
$K_{11};\ D^e_{2r+1}(K_1)\times C_i$
$K_{12};\ S(A^+_1,A^-_1),\ S(A^+_2,A^-_2),\ S(B^\pm_1,B^\mp_2),\ S(E^+_m,E^-_m;1_2,\sigma_y)$;
$m=m'=\frac{1}{2},1,\dots,r$
$K_{21};\ S(D_A,1_2),\ S(D_B,D_B;\sigma_x),\ S(D^{+y}_m,D^{-y}_m;1_2,\sigma_y);\ m=m'$
$K_{22};\ S(D_A;\sigma_z),\ S(D_B;\sigma_y),\ S(D^{\pm y}_m;1_2,\sigma_y);\ m=m'$

TABLE II. (*cont'd.*)

17. $D^q_{2r,i}(K_{st}; s=\{\gamma\}, t=\{\zeta\})$:
K_{11}; $D^q_{2r}(K)\times C_i$
K_{12}; $S(A_1^+, A_1^-)$, $S(A_2^+, A_2^-)$, $S(B_1^\pm, B_2^\mp)$, $S(E_m^+, E_m^-; R_m, \sigma_y R_m)$
$m = m^* = \frac12, 1, \ldots, r-\frac12$
K_{21}; $S(D_A; 1_2)$, $S(D_B, D_B; \sigma_x)$, $S(D_m^{+y}, D_m^{-y}; R_m, \sigma_y R_m)$; $m = m^*$
K_{22}; $S(D_A; \sigma_z)$, $S(D_B; \sigma_y)$, $S(D_m^{\pm y}; R_m, \sigma_y R_m)$; $m = m^*$

18. $D^q_{2r+1,i}(K_{st}; s=\{\gamma\}, t=\{\eta, \zeta\})$
K_{11}; $D^q_{2r+1}(K_1)\times C_i$
K_{12}; $S(A_1^+, A_1^-)$, $S(A_2^+, A_2^-)$, $S(B_1^+, B_1^-)$, $S(B_2^+, B_2^-)$,
$S(E_m^+, E_m^-; \sigma_y R_m)$; $m = m' = \frac12, 1, \ldots, r$
K_{13}; $D^q_{2r+1}(K_2)\times C_i$
K_{14}; $S(A_1^\pm, A_2^\mp)$, $S(B_1^\pm, B_2^\mp)$, $S(E_m^+, E_m^-; \sigma_y R_m, R_m)$; $m = m'$
K_{21}; $S(D_A; 1_2)$, $S(D_B; 1_2)$, $S(D_m^{+y}, D_m^{-y}; R_m, \sigma_y R_m)$; $m = m'$
K_{22}; $S(D_A; \sigma_z)$, $S(D_B; \sigma_z)$, $S(D_m^{\pm y}; R_m, \sigma_y R_m)$; $m = m'$
K_{23}; $S(D_A; \sigma_x)$, $S(D_B; \sigma_x)$, $S(D_m^{+y}, D_m^{-y}; \sigma_y R_m, R_m)$; $m = m'$
K_{24}; $S(D_A, D_A; \sigma_y)$, $S(D_B, D_B; \sigma_y)$, $S(D_m^{\pm y}, D_m^{\pm y}; \sigma_y R_m, R_m)$; $m = m'$

19. $T(K^0)$:
K^0; A, B_1, B_2, T, $E_{1/2}$, $E'_{1/2} = E_{1/2}\times B_1$, $E''_{1/2} = E_{1/2}\times B_2$

20. $T^e(K)$:
K; $S(A)$, $S(B_1, B_2)$, $S(T; 1_3)$, $S(E_{1/2}; \sigma_y)$, $S(E'_{1/2}, E''_{1/2}; \sigma_y)$

21. $T^q(K)$:
K; $S(A)$, $S(B_1)$, $S(B_2)$, $S(T; \hat{C}_4^z)$, $S(E_{1/2}; Z)$, $S(E'_{1/2}; Z)$, $S(E''_{1/2}; Z)$

22. $T_i(K^0)$:
K^0; $T(K^0)\times C_i$; $A^\pm, B_1^\pm, B_2^\pm, T^\pm, E_{1/2}^\pm, E'^\pm_{1/2}, E''^\pm_{1/2}$

23. $T_i^e(K_t; t=\{\eta\})$:
K_1; $T^e(K)\times C_i$
K_2; $S(A^+, A^-)$, $S(B_1^\pm, B_2^\mp)$, $S(T^+, T^-; 1_3)$, $S(E_{1/2}^+, E_{1/2}^-; \sigma_y)$
$S(E'^\pm_{1/2}, E''^\mp_{1/2}; \sigma_y)$

24. $T_i^q(K_t; t=\{\eta\})$
K_1; $T^q\times C_i$
K_2; $S(A^+, A^-)$, $S(B_1^+, B_1^-)$, $S(B_2^+, B_2^-)$, $S(T^+, T^-; \hat{C}_4^z)$,
$S(E_{1/2}^+, E_{1/2}^-; Z)$, $S(E'^+_{1/2}, E'^-_{1/2}; Z)$, $S(E''^+_{1/2}, E''^-_{1/2}; Z)$

25. $O(K^0)$:
K^0; A_1, A_2, E, T_1, T_2, $E_{1/2}$, $E'_{1/2} = E_{1/2}\times A_2$, $Q = E_{1/2}\times E$

26. $O^e(K_t; t=\{\xi\})$:
K_1; $S(A_1)$, $S(A_2)$, $S(E; 1_2)$, $S(T_1; 1_3)$, $S(T_2; 1_3)$, $S(E_{1/2}; \sigma_y)$,
$S(E'_{1/2}; \sigma_y)$, $S(Q; \sigma_y\times 1_2)$
K_2; $S(A_1, A_2)$, $S(E, E; \sigma_y)$, $S(T_1, T_2; 1_3)$, $S(E_{1/2}, E'_{1/2}; \sigma_y)$,
$S(Q, Q; \sigma_y\times\sigma_y)$

27. O_i $(K_s^0; s=\{\beta\})$:
K_1^0; $O(K^0)\times C_i$; $A_1^\pm, A_2^\pm, E^\pm, T_1^\pm, T_2^\pm, E_{1/2}^\pm, E'^\pm_{1/2}, Q^\pm$
K_2^0; $D_A = D(A_1, A_2)$, $D_E^{\pm y} = D(E; \pm\sigma_y)$, $D_T = D(T_1, T_2; 1_3)$
$D_{1/2,1/2} = D(E_{1/2}, E'_{1/2}; 1_2)$, $D_Q^{\pm y} = D(Q; \pm 1_2\times\sigma_y)$

28. O_i^e $(K_{st}; s=\{\beta\}, t=\{\xi, \zeta\})$:
K_{11}; $O^e(K_1)\times C_i$
K_{12}; $S(A_1^+, A_1^-)$, $S(A_2^+, A_2^-)$, $S(E^+, E^-; 1_2)$, $S(T_1^+, T_1^-; 1_3)$
$S(T_2^+, T_2^-; 1_3)$, $S(E_{1/2}^+, E_{1/2}^-; \sigma_y)$, $S(E'^+_{1/2}, E'^-_{1/2}; \sigma_y)$
$S(Q^+, Q^-; \sigma_y\times 1_2)$
K_{13}; $O^e(K_2)\times C_i$
K_{14}; $S(A_1^\pm, A_2^\mp)$, $S(E^+, E^-; \sigma_y)$, $S(T_1^\pm, T_2^\mp; 1_3)$
$S(E_{1/2}^\pm, E'^\mp_{1/2}; \sigma_y)$, $S(Q^+, Q^-; \sigma_y\times\sigma_y)$
K_{21}; $S(D_A; 1_2)$, $S(D_E^{+y}, D_E^{-y}; 1_2)$, $S(D_T; 1_6)$, $S(D_{1/2,1/2}; 1_2\times\sigma_y)$,
$S(D_Q^{+y}, D_Q^{-y}; \sigma_y\times 1_2)$
K_{22}; $S(D_A; \sigma_z)$, $S(D_E^{\pm y}; 1_2)$, $S(D_T; \sigma_z\times 1_3)$
$S(D_{1/2,1/2}; \sigma_z\times\sigma_y)$, $S(D_Q^{\pm y}; \sigma_y\times 1_2)$
K_{23}; $S(D_A; \sigma_x)$, $S(D_E^{+y}, D_E^{-y}; \sigma_y)$, $S(D_T; \sigma_x\times 1_3)$
$S(D_{1/2,1/2}; \sigma_x\times\sigma_y)$, $S(D_Q^{+y}, D_Q^{-y}; \sigma_y\times\sigma_y)$
K_{24}; $S(D_A, D_A; \sigma_y)$, $S(D_E^{\pm y}, D_E^{\pm y}; \sigma_y)$, $S(D_T, D_T; \sigma_y\times 1_3)$
$S(D_{1/2,1/2}, D_{1/2,1/2}; \sigma_y\times\sigma_y)$, $S(D_Q^{\pm y}, D_Q^{\pm y}; \sigma_y\times\sigma_y)$

(i) All unirreps given for the ordinary unitary point groups are defined in I and II. All counirreps given in the table are for $\tau = 1$.
(ii) m^0, m^*, m', and $m^\dagger$ are integers or half-integers defined by $-\frac12 n < m^0 \leqslant \frac12 n$, $m^* = \frac12, 1, \ldots, \frac12(n-1)$, $m' = \frac12, 1, \ldots, r$, $m\dagger = \frac12, 1, \ldots, \frac12(r-1)$ for a given integer n or r.
(iii) n_o (n_e) and r_o (r_e) are odd (even) integers.
(iv) The unirreps E_m of D_n given in (8) are defined by 2×2 matrices M^j_m taking the integral part of j even [see Eq. (11) of Ref. 2].
(v) When two transformation matrices are given for a set of counirreps such as in $S(E_m; 1_2, \sigma_y)$ with $m = m^*$ for K_1 of D^e_{2r} in (10), the first one is for every integral m and the second one is for every half-integral m.
(vi) $D_m^{\pm y} = D(E_m; \pm\sigma_y)$ in (14) means $D_m^{+y} = D(E_m; \sigma_y)$ and $D_m^{-y} = D(E_m; -\sigma_y)$. $S(A_1^\pm, A_2^\mp)$ in (15) means $S(A_1^+, A_2^-)$ and $S(A_1^-, A_2^+)$, The ($\pm$) in the remaining notations should be understood similarly.
(vii) The transformation matrix R_m in (12), (13), (17), and (18) is defined by

$$R_m = \begin{bmatrix} \cos(\pi m/n) & -\sin(\pi m/n) \\ \sin(\pi m/n) & \cos(\pi m/n) \end{bmatrix}.$$

(viii) The transformation matrix Z in (21) and (24) is defined by $Z = 2^{-1/2}(\sigma_y - i1_2)$ if the basis of $E_{1/2}$ is $[\phi_+(\frac12, \frac12), \phi_-(\frac12, \frac12)]$ and $Z = i2^{-1/2}(\sigma_x - \sigma_y)$ if the basis is $[\phi(\frac12, \frac12), \phi(\frac12, -\frac12)]$ (see Ref. 12 of the previous work[4]). The matrix $\hat{C}_4^z$ in (21) and (24) is defined by

$$\hat{C}_4^z = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

$$\Delta^v(a^2) = \tau = \pm 1. \tag{3.6}$$

It should be noted here that in general the coefficient τ is the only coefficient which does not affect the transformation matrix $N(a)$. This fact will be utilized below to simplify the tabulation of the projective counirreps.

The projective counirreps of G^s given by the vector counirreps of $G^{s'}$ thus determined for those listed in Table I are given in Table II together with the projective unirreps of the unitary point groups determined previously.[2] These are classified by the classes of factor systems specified by the representation of the coefficient set $\{\alpha_i\}$ in the representation groups. It is noted, however, that only those counirreps belonging to the classes (90 of them) with $\tau = 1$ are tabulated in Table II. Let us call two classes mutually dual if they differ only in the coefficient τ. Then the counirreps belonging to the class K' with $\tau = -1$ follow immediately from those of its dual K with $\tau = 1$. In general a class K and its dual K' are

p-inequivalent except for a few cases [e.g., $C^{e\prime}_{2r+1}$ in Table I(2)]. In such a case, τ is fixed to 1.

Since several coefficients α_i are involved in specifying a class, it is necessary to introduce a convenient system of notation for expressing a class: For example, according to Table I(15), the representation group $D^{e\prime}_{2r,i}$ has a set of six coefficients $\{\beta, \gamma; \xi, \eta, \zeta; \tau\}$, where the subset $s = \{\beta, \gamma\}$ characterizes the unitary representation group $D'_{2r,i}$, the subset $t = \{\xi, \eta, \zeta\}$ characterizes the defining relations linear in a, and finally τ characterizes a^2. As one can see from Table I or II, this is the most complicated case. Usually fewer coefficients are contained in each subset s or t and frequently there exists only one subset s or t or none besides τ. In the extreme cases even τ is fixed. In any case, each representation of the subset s or t may be denoted by a number such that for a one-member subset, $j = \{\alpha_1\}$,

$$1 = \{1\}, \quad 2 = \{-1\}; \tag{3.7a}$$

for a two-member subset, $j = \{\alpha_1, \alpha_2\}$,

$$\begin{aligned} &1 = \{1, 1\}, \quad 2 = \{1, -1\}, 3 = \{-1, 1\}, \\ &4 = \{-1, -1\}; \end{aligned} \tag{3.7b}$$

and for a three-member subset, $j = \{\alpha_1, \alpha_2, \alpha_3\}$,

$$\begin{aligned} &1 = \{1, 1, 1\}, && 2 = \{1, 1, -1\}, \\ &3 = \{1, -1, 1\}, && 4 = \{1, -1, -1\}, \\ &5 = \{-1, 1, 1\}, && 6 = \{-1, 1, -1\}, \\ &7 = \{-1, -1, 1\}, && 8 = \{-1, -1, -1\}. \end{aligned} \tag{3.7c}$$

There exists no subset with more than three members. Now a class specified by a representation of s and t and $\tau = 1$ is denoted by K_{st} and its dual with $\tau = -1$ is denoted by K'_{st}. Analogously, a class involved with one subset t is denoted by K_t and a class with no subset besides τ is denoted by K and their duals are denoted by K'_t and K', respectively. If K and K' are p-equivalent, we denote it by $K \sim K'$. Finally, the classes of the unitary point groups are denoted by K^0_s or K^0. Obviously no dual can exist for these unitary classes. Table II contains a total of 90 classes of factor systems for G^s and a total of 13 classes of factor systems for the unitary point groups.

It is worthwhile to illustrate Table II through an example. According to Table II(10) the group D^e_{2r} has a total of eight classes of factor systems given by K_t and K'_t, where $t = \{\xi, \eta\}$. The counirreps belonging to one of them, say K_2 ($\xi = 1, \eta = -1, \tau = 1$), are given by

$$\begin{aligned} &S(A_1, A_2),\ A(B_1, B_2),\ S(E_m, E_m; \sigma_y, 1_2), \\ &m = m^* = \tfrac{1}{2}, \ldots, r - \tfrac{1}{2}. \end{aligned} \tag{3.8}$$

The class structure (or the type distribution of the class) may conveniently be denoted by

$$K_2 = (c_2^2 b_4^{r-1} | b_4^r), \tag{3.9}$$

where c, b denote the types of the counirreps, their subscripts denote the dimensions of representation, their superscripts denote the numbers of the respective types, and finally, the left half of the bracket contains the integral counirreps for which $\bar{e} = 1$, and the right half contains the half-integral counirreps for which $\bar{e} = -1$. It should also be noted that the last $(2r - 1)$ counirreps given in (3.8) contain two transformation matrices σ_y and 1_2. In such a case, the first one σ_y is for every integral m and the second one 1_2 is for every half-integral m. One can immediately write down the counirreps belonging to the dual class K'_2 from (3.8) as follows:

$$S(A_1, A_2),\ S(B_1, B_2),\ S(E_m; \sigma_y, 1_2), \quad m = m^*, \tag{3.10}$$

keeping in mind that $\Delta^\nu(a^2) = -\Delta^\nu(\bar{e})$. The type distribution is given by $(c_2^2 a_2^{r-1} | a_2^r)$.

It is worthwhile to comment on the dimensions of the projective corepresentations of G^s. According to Table II, one-dimensional projective counirreps of G^s occur only for the vector corepresentations. This can easily be shown. Excluding these trivial cases, the dimensions of the projective counirreps are all even and limited to 2, 4, 6, 8, and 12. The highest dimension 12 occurs for O^e_i. For the projective unirreps of the ordinary unitary groups, the dimensions are limited to 2, 4, and 6 except for the trivial cases of the vector unirreps,[2] for which the dimensions are limited to 1, 2, and 3.

4. APPLICATION TO MAGNETIC SPACE GROUPS

As is well known, the counirreps of a magnetic space group of wave vector $M(k)$ can be regarded as the projective counirreps of the corresponding magnetic point group belonging to a certain factor system. The present results given in Tables I and II specialized for the crystallographic point groups provide all the counirreps of any $M(k)$. It is only necessary to determine the appropriate gauge transformations which connect the generators of $M(k)$ with those of the corresponding representation group $G^{s\prime}$ given in Table I. In order to use the isomorphisms described by (2.2), it is necessary to classify the crystallographic magnetic point groups in terms of the new system of symbols H^z. For convenience we have expressed their international symbols in terms of H^z in Table III. This table is a special case of the more comprehensive one given by Table I of Ref. 4.

To illustrate the procedure of obtaining the counirreps of $M(k)$ from Tables I–III, we shall use some typical examples taken from the tables of irreducible corepresentations of magnetic space groups given by ML (Miller and Love[5]). We shall follow the notations used by them for the magnetic space groups as well as for the special points of the Brillouin zone. In using their tables, caution should be exercised, since the generator sets of $M(k)$ given by ML are in general not in agreement with those given in Table I. It is also noted that ML give the matrix corepresentations explicitly only for the generators while the present general expressions given in Table II provide the corepresentations for all the unitary elements and the antiunitary augmenting operator a through the unirreps of the point groups given in I. These are sufficient to construct all the basis functions belonging to any given counirrep as discussed in IV.

Example 1: Group (76) 9; $P4'_1$: From Table III, the corresponding G^s is identified as C^q_2. According to Table I(3) this is one of the most simple cases where G^s has only one class of factor systems K ($\sim K'$). From Table II(3) the counirreps of C^q_2 are given by

$$S(M_0), \quad S(M_1, M_1), \quad S(M_{1/2}, M_{-1/2}) \tag{4.1}$$

with the type distribution $(a_1 b_2 | c_2)$. Let us consider the wave

TABLE III. The crystallographic magnetic point groups.[a]

No.	Int.	Present	No.	Int.	Present	No.	Int.	Present
1	$\bar{1}'$	C_1^i	21	$4m'm'$	C_4^v	41	$6'2'2$	D_3^q
2	$2'$	$C_1^q(C_1^u)$	22	$\bar{4}'2'm$	C_{2v}^p	42	$62'2'$	C_6^u
3	m'	$C_1^p(C_1^v)$	23	$\bar{4}'2m'$	D_2^p	43	$6'm'm$	C_{3v}^q
4	$2/m'$	C_2^i	24	$\bar{4}2'm'$	C_{2p}^u	44	$6m'm'$	C_6^v
5	$2'/m'$	$C_i^u(C_i^q)$	25	$4/m'mm$	C_{4v}^i	45	$\bar{6}'m'2$	D_3^p
6	$2'/m$	C_{1p}^i	26	$4'/mm'm$	D_{2i}^q	46	$\bar{6}'m2'$	C_{3v}^p
7	$22'2'$	C_2^u	27	$4'/m'm'm$	D_{2p}^i	47	$\bar{6}m'2'$	C_{3p}^u
8	$m'm'2$	C_2^v	28	$4/mm'm'$	C_{4i}^u	48	$6/m'mm$	C_{6v}^i
9	$m'm2'$	C_{1p}^u	29	$4/m'm'm'$	D_4^i	49	$6'/mm'm$	D_{3p}^i
10	$m'mm$	C_{2v}^i	30	$\bar{3}'$	C_3^i	50	$6'/m'm'm$	D_{3i}^q
11	$m'm'm$	C_{2i}^u	31	$32'$	C_3^u	51	$6/mm'm'$	C_{6i}^u
12	$m'm'm'$	D_2^i	32	$3m'$	C_3^v	52	$6/m'm'm'$	D_6^i
13	$4'$	C_2^q	33	$\bar{3}'m$	C_{3v}^i	53	$m'3$	T^i
14	$\bar{4}'$	C_2^p	34	$\bar{3}'m'$	D_3^i	54	$4'32'$	T^q
15	$4'/m$	C_{2i}^q	35	$\bar{3}m'$	C_{3i}^u	55	$\bar{4}'3m'$	T^p
16	$4/m'$	C_4^i	36	$6'$	C_3^q	56	$m'3m$	T_p^i
17	$4'/m'$	C_{2p}^i	37	$\bar{6}'$	C_3^p	57	$m3m'$	T_i^q
18	$4'22'$	D_2^q	38	$6'/m$	C_{3p}^i	58	$m'3m'$	O^i
19	$42'2'$	C_4^u	39	$6/m'$	C_6^i			
20	$4'm'm$	C_{2v}^q	40	$6'/m'$	C_{3i}^q			

[a] The grey groups are not listed.

vector groups $M(k)$ at high symmetry points $GM=(0,0,0)$, $M=(\frac{1}{2},\frac{1}{2},0)$, $A=(\frac{1}{2},\frac{1}{2},\frac{1}{2})$, and $Z=(0,0,\frac{1}{2})$ of the Brillouin zone. From the symmetry elements $(c_2|00\frac{1}{2})$, $(\theta c_4|00\frac{1}{4})=a$ of $M(k)$ and the defining relations for the generators of $C_2^{q\prime}$ given in Table I(3), we have the following one-to-one correspondences,

$$(c_2|00\tfrac{1}{2})=x, \quad \overline{E}=\bar{e} \quad \text{at } GM \text{ or } M, \tag{4.2}$$

$$(c_2|00\tfrac{1}{2})=-x, \quad \overline{E}=-\bar{e} \quad \text{at } A \text{ or } Z, \tag{4.3}$$

where $\overline{E}$ is the 2π rotation for $M(k)$ while $\bar{e}$ is the 2π rotation for C_2^q. All the counirreps of $M(k)$'s are given by (4.1) with appropriate gauge factors given by (4.2) and (4.3). The type distributions for $M(k=GM$ or $M)$ are given by $(a_1b_2|c_2)$ while those for $M(k=A$ or $Z)$ are given by $(c_2|a_1b_2)$. These results are equivalent to those given by ML. On account of the isomorphism a similar treatment may be given for $M(k)$ belonging to C_2^p.

Example 2: Group (222) 99; Pn3'm: The corresponding magnetic point group is a grey group O_i^e. Let us consider the wave vector group $M(k)$ at $R=(\frac{1}{2},\frac{1}{2},\frac{1}{2})$. In the previous work[2] we have considered the unirrep of the space group 222 at R. From the symmetry elements of $M(k)$ and the defining relations of $O_i^{e\prime}$ given in Table I(28), we have the following one-to-one correspondence:

$$(c_4^z|000)=x, \quad (c_3^{xyz}|000)=y, \quad (I|\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2})=\hat{i}, \quad \overline{E}=\bar{e}, \quad \theta=a \tag{4.4}$$

with the coefficient set

$$\beta=-1, \quad (\xi,\zeta)=(1,1), \quad \tau=1. \tag{4.5}$$

Thus, the counirreps belonging to K_{21} of O_i^e given in Table II(28) provides all the counirreps of $M(k)$ with the correspondence (4.4) without any gauge factors. The type distribution is given by $(a_2c_4a_6|a_4c_8)$ which is in agreement with that given by ML.

Example 3: Group (194) 266; $P6_3'/mm'c$: The magnetic point group is D_{3p}^i which is isomorphic to D_6^e. Let $M(k)$ be at $A=(0,0,\frac{1}{2})$. From the symmetry elements of $M(k)$ and the defining relations for $D_{3p}^{i\prime}$ $(\simeq D_6^{e\prime})$ given in Table I(10), we have the following one-to-one correspondence:

$$(\bar{c}_6|00\tfrac{1}{2})=ix, \quad (c_2^x|000)=iy, \quad \overline{E}=-\bar{e}, \quad a=\theta I \tag{4.6}$$

with the coefficient set

$$(\xi,\eta)=(1,-1), \quad \tau=-1, \tag{4.7}$$

which identifies K_2' of $D_{3p}^{i\prime}$. Thus, from K_2 given in Table II(10) we obtain for K_2' of $D_{3p}^{i\prime}$

$$S(A_1,A_2), \quad S(B_1,B_2), \quad S(E_1;1_2), \quad S(E_2,1_2), \quad S(E_{1/2};\sigma_y), \quad S(E_{3/2};\sigma_y), \quad S(E_{5/2};\sigma_y) \tag{4.8}$$

with the type distribution $(c_2^2a_2^2|a_2^3)$. From (4.6) and (4.8) follow all the counirreps of $M(k)$ with the type distribution $(a_2^3|c_2^2a_2^2)$ since $\overline{E}=-\bar{e}$. Analogous treatments may be given for $M(k)$'s belonging to D_{3p}^e, D_6^i, C_{6v}^i, and C_{6v}^e, all of which are isomorphic to each other.

5. CONCLUDING REMARKS

This work demonstrates once again the effectiveness of the new system of classification introduced previously[3,4] for improper as well as for antiunitary point groups which is best suitable for describing their isomorphisms. We are able to describe the representation groups of all finite point groups by those of eight unitary and 20 antiunitary characteristic sets of the point groups in Table I (the icosahedral group is excluded). These are given by the defining relations of the abstract group generators which are common to all point groups isomorphic to each other. Then, by means of the modified theory of induced representations introduced in II and IV, we are able to present in Table II the general expressions of all p-inequivalent projective unirreps or counirreps of all those 28 point groups in terms of the unirreps of the proper point groups previously determined in II.

The present results are more than sufficient to find all the unirreps or counirreps of any space group (unitary or antiunitary) of wave vector through simple gauge transformations. Here it is essential to identify the point groups corresponding to respective space groups in terms of the new system of classifications. Table III provides the identification of the international notations of the crystallographic magnetic point groups in terms of the new system of notations H^z. As one can see from Table III, it is hardly possible to recognize their isomorphisms from the international symbols alone.

The present work can easily be extended to calculate the projective counirreps of the magnetic point groups of infinite order. This problem will, however, be discussed in a forthcoming paper, since they are mixed continuous groups and thus construction of their representation groups requires somewhat different algebraic manipulations from those used in this work.

Note added in proof: In constructing Table II, the required matrix representtions of the halving subgroups are determined from those of the point groups given in Ref. 1 through the following realizations of the abstract group generators $(x, y, \text{and } s)$: $x \leftrightarrow$ the highest axis of rotation c_n^z in the z direction for every point group, $y \leftrightarrow c_2^x$ for D_n or $y \leftrightarrow c_3^{xyz}$ for T and 0, $s \leftrightarrow \hat{i}c_3^{xyz}$ for T_i. Accordingly, their respective matrix representatives thus determined should be assigned to any group generators represented by x, y, and s, when we use Table II. Such a realization seems essential for the definite identification for the coefficient elements α_i of the representation groups, whose values classify the classes of the factor systems given in Table II. This note applies also for the projective unirreps given in Ref. 2.

ACKNOWLEDGMENT

The author wishes to thank J. N. Viesulas for giving valuable comments on the manuscript.

[1] S. K. Kim, J. Math. Phys. **22**, 2101 (1981).
[2] S. K. Kim, J. Math. Phys. **24**, 411 (1983).
[3] S. K. Kim, J. Math. Phys. **24**, 414 (1983).
[4] S. K. Kim, J. Math. Phys. **24**, 419 (1983).
[5] I. Schur, J. Reine, Angew. Math. **20**, 127 (1904); **85**, 132 (1907); **155**, 139 (1911).
[6] J. S. Lomont, *Applications of Finite Groups* (Academic, New York, 1959).
[7] G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors*, translated from Russian by P. Shelnitz (Wiley, New York, 1974).
[8] S. C. Miller and W. H. Love, *Tables of Irreducible Representations of Space Groups and Co-representations of Magnetic Space Groups* (Bruett, Boulder, CO, 1967).
[9] It is possible, however, that some corepresentations of $P_i^{z\prime}$ may satisfy the direct product relation for certain classes of the factor systems where the representation of $\hat{i} \in C_i$ commutes with the corepresentations of all the elements of $P^{z\prime}$ [e.g., see Table II(6), where $C_{ni}^e(K_{11}) = C_n^e(k) \times C_i$].

General irreducibility condition for vector and projective corepresentations of antiunitary groups

Shoon K. Kim
Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122

(Received 11 March 1983; accepted for publication 29 July 1983)

A general condition for irreducibility of vector and projective corepresentations of an antiunitary group is presented. It depends only on the characters of the unitary halving subgroup of the covering group. It reduces to the well known type criterion of corepresentations when it is specialized to the three types of corepresentations.

PACS numbers: 02.20. + b

1. INTRODUCTION

The theory of corepresentation of an antiunitary group was formulated by Wigner.[1] He has shown that there exist three types of irreducible unitary corepresentations (counirreps), which can be constructed from the irreducible unitary representations (unirreps) of the halving subgroup of the antiunitary group. Almost of all work on corepresentation has been carried out using this approach. In particular, a great deal of work[2–10] has been carried out up until recently on the "type criterion" for determining which of the three types is realized for a given unirrep of the halving subgroup. A more general approach to the corepresentation theory based on an extended version of Schur's lemma was formulated by Dimmock.[6] He has obtained the general orthogonality relations for counirreps. He has concluded, however, that "the orthogonality relations have an inconvenient form and further development of the representation theory of nonunitary groups along these lines has so far proven untenable." Since then such a general approach has unfortunately been discontinued.

In this work, we shall resume the general approach initiated by Dimmock and deduce a general condition for a corepresentation to be irreducible. It will be shown that the condition depends only on the characters of the elements of the halving subgroup which is unitary. It is a simple generalization of the corresponding condition for an ordinary representation of a unitary group. Lack of such a simple condition must have been a big pain in the theory of corepresentation. For example, Bradley and Davies,[7] who first rigorously proved the irreducibility of the three types of corepresentations had to show that it is impossible to construct the transformation matrices which reduce them. It is also very satisfying to see that the type criterion follows naturally from the irreducibility condition when it is specialized for the three types of corepresentations. Previously, the type criterion has been introduced heuristically.[5,6]

In Sec. II we shall first extend the orthogonality relations of the counirreps to the projective ones through a covering group extended by an abelian unitary group.[11,12] Then, from the orthogonality relations of the characters of the covering group we shall deduce the irreducibility condition for a projective corepresentation in general. As an application, we shall give in Sec. III an easy proof for the irreducibility of the three types of corepresentations. Here, we base our arguments on the modified forms of the corepresentations recently introduced by the author.[13] In the process of the proof, the type criterion comes in as an essential part of the irreducibility condition. The criterion thus obtained is applicable to any projective corepresentation.

2. THE IRREDUCIBILITY CONDITION FOR PROJECTIVE COREPRESENTATIONS

It is well known[11] that for any group G a covering group G' can be constructed such that all the irreducible projective representations of G can be found from the vector representations of G'. This theorem can be extended to the projective counirreps of an antiunitary group[14] M. Thus we may discuss the irreducibility condition for the projective corepresentations of an antiunitary group M via the vector counirreps of a covering group M' without introducing cumbersome projective factors altogether. In the following we shall first describe a covering group extended by a unitary abelian group in some detail as a preparation.

Let an antiunitary group $M = \{m\}$ be defined through its halving subgroup $H = \{h\}$, which is unitary, as follows;

$$M = H + a^0 H, \tag{2.1}$$

where a^0 is an antiunitary operator. Let $M' = \{m'\}$ be another antiunitary group defined by

$$M' = H' + aH', \tag{2.2}$$

where $H' = \{h'\}$ is the halving subgroup of M' and a is an antiunitary operator. Let T be a unitary abelian group which is contained in the center of M', then it is also contained in the center of H'. If there exists an isomorphism such that

$$M'/T \simeq M, \tag{2.3}$$

then M is called a covering group of M extended by the abelian group T. Analogously,

$$H'/T \simeq H, \tag{2.4}$$

hence H' is also a covering group of H extended by T. Now let $\overline{M} = \{\overline{m}\}$ and $\overline{H} = \{\overline{h}\}$ be the sets of the respective coset representatives of M' and H' with respect to T. Let these be chosen such that

$$\overline{M} = \overline{H} + a\overline{H}. \tag{2.5}$$

In general, $\overline{M}$ and $\overline{H}$ may not close. On account of (2.3) and (2.4) however, there exists one to one correspondence,

$$m \leftrightarrow \overline{m},$$

i.e.,

 0022-2488/84/020197-03$02.50

$$h \leftrightarrow \bar{h} \tag{2.6}$$

and

$$a^0 h \leftrightarrow a\bar{h},$$

such that if $D(m')$ is a vector counirrep of $m' \in M'$ belonging to a definite unirrep of T, then $m \to D(\bar{m})$ provides a projective counirrep of M belonging to a factor system defined by the unirrep of T. Analogously, let $\Delta(h')$ be a vector unirrep of $h' \in H'$, then $h \to \Delta(\bar{h})$ provides a projective unirrep of H specified by the unirrep of T.

With the preparation given above we shall now discuss the orthogonality relations of the projective counirreps. Let $D^{(i)}(m')$ and $D^{(j)}(m')$ be two inequivalent counirreps of $m' \in M'$. Then, from their orthogonality relations given by Dimmock[6] we can easily derive the following orthogonality relations for the projective counirreps $D^{(i)}(\bar{m})$ and $D^{(j)}(\bar{m})$ of $m \in M$:

$$\sum_{\bar{h} \in \bar{H}} D^{(i)}(\bar{h})_{\alpha\mu} D^{(j)}(\bar{h})^*_{\beta\nu} = 0, \tag{2.7}$$

$$\sum_{\bar{h} \in \bar{H}} D^{(i)}(a\bar{h})_{\alpha\mu} D^{(j)}(a\bar{h})^*_{\beta\nu} = 0, \tag{2.8}$$

$$\sum_{\bar{h} \in \bar{H}} \{ D^{(i)}(\bar{h})_{\alpha\mu} D^{(i)}(\bar{h})^*_{\beta\nu} + D^{(i)}(a\bar{h})_{\alpha\nu} D^{(i)}(a\bar{h})^*_{\beta\mu} \} = d_i^{-1} |M| \delta_{\alpha\beta} \delta_{\mu\nu}, \tag{2.9}$$

where d_i is the dimension of $D^{(i)}$ and $|M|$ is the order of the group M. The original orthogonality relations due to Dimmock is recovered if we replace $\bar{h}$ and a by h and a^0, respectively. Taking the traces of (2.7) or (2.9), we obtain

$$\sum_{\bar{h} \in \bar{H}} \kappa^i(\bar{h}) \kappa^j(\bar{h})^* = 0, \tag{2.10}$$

$$\sum_{\bar{h} \in \bar{H}} \kappa^i(a\bar{h}) \kappa^j(a\bar{h})^* = 0, \tag{2.11}$$

$$\sum_{\bar{h} \in \bar{H}} \{ |\kappa^i(\bar{h})|^2 + \kappa^i((a\bar{h})^2) \} = |M|, \tag{2.12}$$

where κ^i and κ^j are traces of $D^{(i)}$ and $D^{(j)}$, respectively. Since $(a\bar{h})^2$ belongs to H', Eq. (2.12) depends only on the elements belonging to the unitary halving subgroup H' of the covering group M'. This fact makes the theory tenable for deducing the irreducibility conditions, since Eq. (2.12) serves as the necessary and sufficient condition for $D^{(i)}$ to be irreducible as it will be shown below.

Let D be a unitary corepresentation of M' corresponding to a definite unirrep of the abelian group T and κ be its character. Let n_i be the number of times a counirrep $D^{(i)}$ appears in the reduced form of D. Then,

$$\kappa(h') = \sum_i n_i \kappa^i(h') \quad \text{for all} \quad h' \in H', \tag{2.13}$$

$$n_i = \sum_{\bar{h} \in \bar{H}} \kappa(\bar{h}) \kappa^i(\bar{h})^* \Big/ \sum_{\bar{h} \in \bar{H}} |\kappa^i(\bar{h})|^2. \tag{2.14}$$

In view of (2.12), let us define a sum $S(D)$ by

$$S(D) = \left(\frac{1}{|M|} \right) \sum_{\bar{h} \in \bar{H}} \{ |\kappa(\bar{h})|^2 + \kappa((a\bar{h})^2) \}. \tag{2.15}$$

Then, from (2.13) and (2.15) we have the following inequality:

$$S(D) \geqslant \sum_i n_i, \tag{2.16}$$

where the equality occurs if and only if $n_i^2 = n_i$ for all i. Thus, the corepresentation D of M' (or the projective corepresentations of M) is irreducible if and only if the sum takes its minimum value

$$S(D) = 1. \tag{2.17}$$

It is a simple extension of the ordinary irreducibility condition for a unitary representation. It should be noted that only the second term in the bracket on the rhs of (2.15) may depend on the factor system. If necessary, one may express the second term in terms of the character $\kappa(\bar{m})$ of the projective counirrep $D(\bar{m})$ of $m \in M$ as follows:

$$\kappa((a\bar{h})^2) = [a^0 h, a^0 h] \kappa(\overline{(ah)^2}), \tag{2.18}$$

where $[a^0 h, a^0 h]$ is the projective factor for the product $\mathbf{D}(a\bar{h})\mathbf{D}(a\bar{h})^*$. As it will be seen in the next section, this second part is related to the type criterion of the counirreps.

3. APPLICATION TO THE THREE TYPES OF COUNIRREPS

By the irreducibility condition (2.17) we shall first give an easy proof for the irreducibility of the three types of the counirreps of M' and at the same time deduce the well known type criterion first introduced by Dimmock and Wheeler[5] in its most general form. We shall base our argument on the modified form of the three types of counirreps recently introduced by the author.[13]

Let $\{\Delta^\nu(h')\}$ be a complete set of the unirreps of the halving subgroup H' of M' belonging to a definite unirrep of the abelian group T. Then, $\Delta^\nu(a^{-1}h'a)^*$ is also a unirrep equivalent to one of the given unirreps, say Δ^μ, of H'. Thus, there exists a unitary matrix $N(a)$ such that

$$\Delta^\nu(a^{-1}h'a)^* = N(a)^{-1} \Delta^\mu(h') N(a) \tag{3.1}$$

for all $h' \in H'$. Since two unirreps Δ^ν and Δ^μ are connected by $N(a)$, we arrive at a corepresentation of M' given by

$$D^{(\nu,\mu)}(h') \begin{bmatrix} \Delta^\nu(h') & 0 \\ 0 & \Delta^\mu(h') \end{bmatrix}, \tag{3.2a}$$

$$D^{(\nu,\mu)}(a) = \begin{bmatrix} 0 & \Delta^\nu(a^2) N(a)^{-1*} \\ N(a) & 0 \end{bmatrix}. \tag{3.2b}$$

There exist three cases: For case (a), $\nu = \mu$, $N(a) N(a)^* = \Delta^\nu(a^2)$, we have two counirreps

$$D^{(\nu\pm)}(h') = \Delta^\nu(h'); \quad D^{(\nu\pm)}(a) = \pm N(a) \tag{3.3}$$

which are mutually equivalent. For case (b), $\nu = \mu$, $N(a) N(a)^* = -\Delta^\nu(a^2)$, we have $D^{(\nu,\nu)}$ which is a special case of (3.2). Finally for case (c), $\nu \neq \mu$, we have $D^{(\nu,\mu)}$ as given by (3.2). The irreducibility of type (a) counirrep is obvious, while the irreducibilities of those of types (b) and (c) are not quite obvious.[7]

In order to reestablish their irreducibilities via the irreducibility condition (2.17) we may rewrite the sum $S(D)$ for the three types of counirreps as follows, using the orthogonality theorem for $\chi^\nu(h')$ of H',

$$S(D^{\nu\pm}) = \frac{1}{2} + \left(\frac{1}{2|H|}\right) \sum_{\bar{h}\in\bar{H}} \chi^\nu((a\bar{h})^2), \tag{3.4a}$$

$$S(D^{(\nu,\nu)}) = 2 + \left(\frac{1}{|H|}\right) \sum_{\bar{h}\in\bar{H}} \chi^\nu((a\bar{h})^2), \tag{3.4b}$$

$$S(D^{(\nu,\mu)}) = 1 + \left(\frac{1}{2|H|}\right) \sum_{\bar{h}\in\bar{H}} \{\chi^\nu((a\bar{h})^2) + \chi^\mu((a\bar{h})^2)\}. \tag{3.4c}$$

Then, the irreducibility condition (2.17) reduces to

$$\left(\frac{1}{|H|}\right) \sum_{\bar{h}\in\bar{H}} \chi^\nu((a\bar{h})^2) = \begin{cases} 1 \ , & \text{case (a)}, \\ -1 \ , & \text{case (b)}, \\ 0 \ , & \text{case (c)}, \end{cases} \tag{3.5}$$

which is the well known type criterion in its most genral form and can be proven by a direct calculation in the usual manner.[7] If necessary the above criterion can be rewritten in terms of the projective factor as in (2.18).

Thus, we have established that the type criterion serves also as the irreducibility condition for the three types of the counirreps. It holds for any projective counirreps of any antiunitary group. Thus, it is applicable to the single or double magnetic point groups as well as to the magnetic space groups. The previous results on the type criterion obtained by Dimmock and Wheeler[5] and by Karavaev, Kudryavtseva, and Chaldyshev[8] and by others[9,10] are easily reproduced from the above result.

[1] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959), Chap. 26.
[2] C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972), and references therein.
[3] G. Frobenius and I. Schur, S. B. Dent Akad. Wiss. **49**, 186 (1906).
[4] C. Herring, Phys. Rev. **52**, 361 (1937).
[5] J. O. Dimmock and R. G. Wheeler, J. Phys. Chem. Solids **23**, 729 (1962); see also Phys. Rev. **127**, 39 (1962).
[6] J. O. Dimmock, J. Math. Phys. **4**, 1307 (1963).
[7] C. J. Bradley and B. L. Davies, Rev. Mod. Phys. **40**, 359 (1968).
[8] G. F. Karavaev, N. V. Kudryavtseva, and V. A. Chaldyshev, Sov. Phys. Solid State **4**, 2540 (1963).
[9] P. M. Van den Broek, Lett. Math. Phys. **3**, 151 (1979).
[10] R. Dirl, J. Math. Phys. **22**, 1139 (1981).
[11] L. S. Lomont, *Application of Finite Groups* (Academic, New York, 1959), p. 230; see also G. L. Bir and G. E. Pikus, *Symmetry and Strain-induced Effects in Semiconductors*, translated from Russian by P. Shelnitz (Wiley, New York, 1974), p. 88.
[12] S. K. Kim, J. Math. Phys. **24**, 411 (1983).
[13] S. K. Kim, J. Math. Phys. **24**, 419 (1983).

[14] As is well known, such a covering group of minimal order is called a representation group of the original group. Recently, the author has constructed all the representation groups of all magnetic point groups (except the Icosahedral group) and their counirreps; see "The unified theory of the point group. V," J. Math. Phys. **25**, 189 (1984).

Semiregular induction of group representations[a]

D. J. Klein
Department of Marine Sciences, Texas A&M University at Galveston, Galveston, Texas 77553

(Received 14 June 1983; accepted for publication 16 September 1983)

The method of inducing an irreducible representation of a group from that of a subgroup is extended. This generalized induction process is illustrated to occur in applications and to account for some occurrences of "intermediate" or "hidden" symmetry. Some general results are proved, including a reciprocity theorem relating general induction and subduction processes.

PACS numbers: 02.20. + b, 02.30. + g

1. BACKGROUND

Frequently in physics and chemistry one encounters spaces invariant to a group $\mathscr{G}$ of operators. Such a $\mathscr{G}$-invariant space is sometimes generated through the action of elements of $\mathscr{G}$ on a subspace of vectors (or kets) already symmetry adapted to an irreducible representation Γ^{α} of a subgroup $\mathscr{A} \subseteq \mathscr{G}$. In the case that this subspace has "no symmetry" other than this $\mathscr{A}$ symmetry one expects the generated $\mathscr{G}$-invariant space to carry the so-called (regular) induced representation $\Gamma^{\alpha\uparrow}$ of $\mathscr{G}$. Now in practice it is sometimes found that the resulting representation is not the group-theoretically definable $\Gamma^{\alpha\uparrow}$ but rather a proper subrepresentation. Such occurrences have been termed[1] "nonregular induction" processes and have been said[2] to be due to an "intermediate symmetry." One might ascribe such results to a "hidden symmetry" of the function or kets involved.

One situation where nonregular induction occurs involves operators V for the crystal field potential around a central metal ion. Griffith[2] takes V as a sum of contributions from each of the ligand species surrounding the metal. For instance, the ligand species might be two fluoride ions and two cyanide ions arranged at the corners of a tetrahedron with the metal at the center. In this case V is of totally symmetric a_1 symmetry with regard to the $\mathscr{C}_{2v}$ point group for the chemical complex; also V may be resolved into irreducible tensorial components for the larger tetrahedral group $\mathscr{T} \supset \mathscr{C}_{2v}$. Now a_1 of $\mathscr{C}_{2v}$ induces to $a_1\uparrow = a_1 + e + t_2$ of $\mathscr{T}$; but taking the individual contribution to V for the ligands to be as for isolated ligands (with cyclic symmetry around the ligand-metal axis), one finds that the e-component does not occur. Griffith[2] and later Clark[3] found many further instances of such nonregular induction for various geometries and ligand arrangements.

A second situation where nonregular induction frequently arises involves the construction of chirality functions f, such as are used to describe the optical activity of molecules. In one approach Ruch and Schönhofer[4] take f to be expressed in terms of parameters l_i associated with the ith ligand occurring on a molecular skeleton; then they seek f which have a chiral symmetry with respect to (point-group-correspondent) permutations of the l_i. For instance five ligands might be attached to the corners of a regular pentagonal molecular skeleton, which has a permutation group $\mathscr{C}_{5v}$ corresponding to the skeletons point-group symmetry. In this case f is to have a (chiral) a_2 symmetry with regard to $\mathscr{C}_{5v}$, and f may be resolved into irreducible components for the larger symmetric group $\mathscr{S}_5 \supset \mathscr{C}_{5v}$. Now $a_2\uparrow = 2[3,1,1]$; but taking f to be the "simplest" function of the l_i (namely the a_2-projection of $l_1\, l_2^2$), one finds that only one [3,1,1] representation is generated from f. Many examples of such nonregular induction have been found[1,4,5]; indeed such occurrences motivate Ruch and Schönhofer's "qualitative completeness" concept.

These two situations serve to illustrate the practical occurrence of nonregular induction, and some additional examples from physics and mathematics will be mentioned in Sec. 6. Our primary interest here is directed to the question: Is the process of nonregular induction something which must be addressed on an individual trial-and-error sort of approach or is there a general purely group-theoretic rationalization of such occurrences? Although the former possibility must be true in the most general case, we here point out a semiregular induction process that accounts for all practical cases of nonregular induction which we have examined.

2. SEMIREGULAR INDUCTION AND SUBDUCTION

The processes of interest are characterized in terms of a subgroup lattice

$$\begin{array}{ccc} & \mathscr{G} & \\ \mathscr{A} & & \mathscr{B} \\ & \mathscr{D} & \end{array} \tag{2.1}$$

where $\mathscr{A}$ and $\mathscr{B}$ are two general subgroups of $\mathscr{G}$, and $\mathscr{D} = \mathscr{A}\cap\mathscr{B}$. Irreducible representations of $\mathscr{G}$, $\mathscr{A}$, $\mathscr{B}$, and $\mathscr{D}$ will be labeled by corresponding Greek-letter labels γ, α, β, and δ; row and column labels for these irreducible representations will have corresponding Latin-letter labels g, a, b, and d.

Suppose that an irreducible space for Γ^{β} of $\mathscr{B}$ is acted upon by elements of $\mathscr{A}$ to project a space of a particular symmetry α of $\mathscr{A}$ and subsequently this $\mathscr{A}$-invariant space is acted on by elements of $\mathscr{G}$ to obtain a $\mathscr{G}$-invariant space. The (maximum possible) representation carried by this $\mathscr{G}$-invariant space we define to be the *semiregular induced* representation denoted by $\Gamma^{\beta\searrow\alpha\uparrow}(\mathscr{G})$.

Next suppose an irreducible space for Γ^{γ} of $\mathscr{G}$ is subduced (i.e., restricted) to a particular symmetry α of $\mathscr{A}$, and subsequently this $\mathscr{A}$-invariant space is acted on by elements of $\mathscr{B}$ to obtain a $\mathscr{B}$-invariant space. The (maximum possible) representation carried by this $\mathscr{B}$-invariant space we de-

[a] Research supported by the Robert A. Welch Foundation of Houston, Texas.

fine to be the *semiregular subduced* representation denoted by $\Gamma^{\gamma\downarrow\alpha\hookrightarrow}(\mathscr{B})$.

In the semiregular induction process the "hidden symmetry" is that associated with Γ^{β} of $\mathscr{B}$. Often $\Gamma^{\beta\hookrightarrow\alpha\uparrow}(\mathscr{G})$ can very readily be seen to be a proper subrepresentation of $\Gamma^{\alpha\uparrow}(\mathscr{G})$ since $\Gamma^{\beta\hookrightarrow\alpha\uparrow}(\mathscr{G})$, being carried by a space initially generated from functions of β symmetry, will not contain any $\Gamma^{\gamma}(\mathscr{G})$ unless it also occurs in $\Gamma^{\beta\uparrow}(\mathscr{G})$. That is, one might expect that $\Gamma^{\beta\hookrightarrow\alpha\uparrow}(\mathscr{G})$ is some sort of "intersection" between $\Gamma^{\alpha\uparrow}(\mathscr{G})$ and $\Gamma^{\beta\uparrow}(\mathscr{G})$. In the semiregular subduction process there is a "hidden symmetry" associated with Γ^{γ} of $\mathscr{G}$.

3. GROUP-ALGEBRAIC FORMULATION

These definitions can be made manifestly independent of reference to external carrier space through group-algebraic considerations. The group algebras for $\mathscr{G}$, $\mathscr{A}$, $\mathscr{B}$, $\mathscr{D}$ are denoted by $\mathscr{A}(\mathscr{G})$, $\mathscr{A}(\mathscr{A})$, $\mathscr{A}(\mathscr{B})$, $\mathscr{A}(\mathscr{D})$, and each has a so-called *matric* basis of elements $e^{\gamma}_{gg'}$, $e^{\gamma}_{aa'}$, $e^{\beta}_{bb'}$, $e^{\delta}_{dd'}$, respectively. Here for instance

$$e^{\gamma}_{gg'} = \frac{|\gamma|}{|\mathscr{G}|} \sum_{G\in\mathscr{G}} \Gamma^{\gamma}_{g'g}(G^{-1})G, \tag{3.1}$$

where $|\mathscr{G}|$ is the order of $\mathscr{G}$ and $|\gamma|$ is the degree of Γ^{γ}. Note that the space $\mathscr{A}^{\gamma}_{g}$ with basis $\{e^{\gamma}_{g'g}; g' = 1 \text{ to } |\gamma|\}$ is left invariant to $\mathscr{G}$,

$$G\, e^{\gamma}_{g'g} = \sum_{g''} \Gamma^{\gamma}_{g''g'}(G)\, e^{\gamma}_{g''g}, \tag{3.2}$$

so that $\mathscr{A}^{\gamma}_{g}$ carries the irreducible representation $\Gamma^{\gamma}(\mathscr{G})$. The direct sum space $\mathscr{A}^{\gamma} \equiv \dot{\Sigma}_g\, \mathscr{A}^{\gamma}_{g}$; with basis $\{e^{\gamma}_{g'g}; g,g' = + \text{ to } |\gamma|\}$ may be characterized as a minimal two-sided subalgebra, which carries $\Gamma^{\gamma}(\mathscr{G})$ $|\gamma|$ times. Of course there are analogous subalgebras $\mathscr{A}^{\alpha}_{a}$, $\mathscr{A}^{\beta}_{b}$, $\mathscr{A}^{\delta}_{d}$ and $\mathscr{A}^{\alpha}$, $\mathscr{A}^{\beta}$, $\mathscr{A}^{\delta}$ for $\mathscr{A}$, $\mathscr{B}$, $\mathscr{D}$, respectively.

The standard induced and subduced representation theory may be cast in terms of these quantities. The *regular induced* representation $\Gamma^{\alpha\uparrow}(\mathscr{G})$ is that carried by $\mathscr{A}(\mathscr{G})\,\mathscr{A}^{\alpha}_{a}$, and the *regular subduced* representation $\Gamma^{\alpha\downarrow}(\mathscr{A})$ is that carried by $\mathscr{A}(\mathscr{A})\,\mathscr{A}^{\gamma}_{g}$. These representations are generally reducible;

$$\Gamma^{\alpha\uparrow}(\mathscr{G}) = \sum_{\gamma} |\alpha\uparrow\gamma|\,\Gamma^{\gamma}(\mathscr{G}),$$
$$\Gamma^{\gamma\downarrow}(\mathscr{A}) = \sum_{\alpha} |\gamma\downarrow\alpha|\,\Gamma^{\alpha}(\mathscr{A}), \tag{3.3}$$

where $|\alpha\uparrow\gamma|$ and $|\gamma\downarrow\alpha|$ are frequencies for the occurrence of the appropriate irreducible representation. As originally shown by Frobenius[6] $|\alpha\uparrow\gamma| = |\gamma\downarrow\alpha|$.

Semiregular induced and subduced representations may be characterized in a manner analogous to that for the corresponding regular representations. The semiregular induced representation $\Gamma^{\beta\hookrightarrow\alpha\uparrow}(\mathscr{G})$ is identified as that carried by $\mathscr{A}(\mathscr{G})\mathscr{A}^{\alpha}\mathscr{A}^{\beta}_{b}$, while the semiregular subduced representation $\Gamma^{\gamma\downarrow\alpha\hookrightarrow}(\mathscr{B})$ is that carried by $\mathscr{A}(\mathscr{B})\mathscr{A}^{\alpha}\mathscr{A}^{\gamma}_{g}$. These representations also are generally reducible;

$$\Gamma^{\beta\hookrightarrow\alpha\uparrow}(\mathscr{G}) = \sum_{\gamma} |\beta\hookrightarrow\alpha\uparrow\gamma|\,\Gamma^{\gamma}(\mathscr{G}),$$
$$\Gamma^{\gamma\downarrow\alpha\hookrightarrow}(\mathscr{B}) = \sum_{\beta} |\gamma\downarrow\alpha\hookrightarrow\beta|\,\Gamma^{\beta}(\beta), \tag{3.4}$$

where $|\beta\hookrightarrow\alpha\uparrow\gamma|$ and $|\gamma\downarrow\alpha\hookrightarrow\beta|$ are frequencies of occurrence of the appropriate irreducible representation.

4. SEQUENCE ADAPTATION AND RECOUPLING COEFFICIENTS

Since the semiregular processes involve the lattice of subgroups in (2.1), one may anticipate that sequence adaptation,[7] say for the chains $\mathscr{D} \subsetneq \mathscr{A} \subsetneq \mathscr{G}$ or $\mathscr{D} \subsetneq \mathscr{A} \subsetneq \mathscr{G}$, will play an important role. For *sequence adaptation* to the simple chain $\mathscr{A} \subsetneq \mathscr{G}$ the row or column labels g can be presented as $\mathscr{a}\alpha a$, where $\mathscr{a}$ is a degeneracy label for the occurrence of α in $\gamma\downarrow$, and

$$\Gamma^{\gamma}_{(\mathscr{a}\alpha a)(\mathscr{a}'\alpha' a')}(A) = \delta(\mathscr{a},\mathscr{a}')\delta(\alpha,\alpha')\Gamma^{\alpha}_{aa'}(A),$$
$$A \in \mathscr{A}. \tag{4.1}$$

Similar comments apply for the $\mathscr{B} \subsetneq \mathscr{G}$ chain where the row or column labels are presented as $\mathscr{b}\beta b$. The *recoupling coefficients* transforming between these two sequence adaptation schemes are of relevance also, and are identified as $(\mathscr{a}\alpha a|\gamma|\mathscr{b}\beta b)$ and $(\mathscr{b}\beta b|\gamma|\mathscr{a}\alpha a)$. They inter-relate irreducible representations for the two sequence adaptation schemes

$$\sum_{\mathscr{b}\beta b} \sum_{\mathscr{b}'\beta' b'} (\mathscr{a}\alpha a|\gamma|\mathscr{b}\beta b)\Gamma^{\gamma}_{(\mathscr{b}\beta b)(\mathscr{b}'\beta' b')}(G)(\mathscr{b}'\beta' b'|\gamma|\mathscr{a}'\alpha' a')$$
$$= \Gamma^{\gamma}_{(\mathscr{a}\alpha a)(\mathscr{a}'\alpha' a')}(G). \tag{4.2}$$

The recoupling coefficients may be defined to be *unitary*

$$(\mathscr{a}\alpha a|\gamma|\mathscr{b}\beta b) = (\mathscr{b}\beta b|\gamma|\mathscr{a}\alpha a)^{*}. \tag{4.3}$$

Sequence-adapted matric basis elements arise and may be transformed between the two sequence adaptation schemes by the recoupling coefficients.

$$\sum_{\mathscr{b}\beta b} \sum_{\mathscr{b}'\beta' b'} (\mathscr{b}\beta b|\gamma|\mathscr{a}\alpha a)e^{\gamma}_{(\mathscr{b}\beta b)(\mathscr{b}'\beta' b')}(\mathscr{a}'\alpha' a'|\gamma|\mathscr{b}'\beta' b')$$
$$= e^{\gamma}_{(\mathscr{a}\alpha a)(\mathscr{a}'\alpha' a')}. \tag{4.4}$$

More generally we define *skew* sequence-adapted matric basis elements

$$e^{\gamma}_{(\mathscr{b}\beta b)(\mathscr{a}\alpha a)} \equiv \sum_{\mathscr{b}'\beta' b'} e^{\gamma}_{(\mathscr{b}\beta' b)(\mathscr{b}'\beta b')}\,(\mathscr{a}\alpha a|\gamma|\mathscr{b}'\beta' b') \tag{4.5}$$

which are sequence adapted to different chains on the left and right. They still form a basis to $\mathscr{A}(\mathscr{G})$. Further it may be verified that

$$e^{\gamma}_{(\mathscr{b}\beta b)(\mathscr{a}\alpha a)}\, e^{\alpha'}_{a'a''} = \delta(\alpha,\alpha')\delta(a,a')e^{\gamma}_{(\mathscr{b}\beta b)(\mathscr{a}\alpha a'')}. \tag{4.6}$$

Other properties[7] will not be needed here.

Sequence-adaptation ideas extend to longer chains, such as $\mathscr{D} \subsetneq \mathscr{A} \subsetneq \mathscr{G}$ and $\mathscr{D} \subsetneq \mathscr{B} \subsetneq \mathscr{G}$. The row and column labels then are $\mathscr{a}\alpha\hat{\mathscr{a}}\delta d$ and $\mathscr{b}\beta\hat{\mathscr{b}}\delta d$, where $\hat{\mathscr{a}}$ and $\hat{\mathscr{b}}$ are degeneracy labels for the occurrence of δ in $\alpha\downarrow$ and $\beta\downarrow$, respectively. If we choose the recoupling coefficients to be such that a matrix element of any $D \in \mathscr{D}$ is left unaffected in the transformation from one sequence-adaptation scheme to another, then

$$(\alpha\alpha\hat{a}\delta d\,|\gamma|\,\ell\beta\hat{\ell}\delta' d') = \delta(\delta,\delta')\delta(d,d')(\alpha\alpha\hat{a}\delta|\gamma|\ell\beta\hat{\ell}\delta), \tag{4.7}$$

where we note the last abbreviated recoupling coefficient is independent of d and d'.

5. THEOREMATIC RESULTS

Since the space $\mathscr{A}(\mathscr{G})\,\mathscr{A}^{\alpha}\mathscr{A}_{b}^{\beta}$ carries the semiregular induced representation labeled by $\beta\hookrightarrow\alpha\uparrow\mathscr{G}$, it is of interest to characterize a (symmetry-adapted) basis for this space. In a step toward this goal note that a spanning set will be obtained if we take any triple of bases for $\mathscr{A}(\mathscr{G})$, $\mathscr{A}^{\alpha}$, $\mathscr{A}_{b}^{\beta}$ and then form all products of ordered triples of elements from these three bases. Consider such a triple of skew-matric bases: first for $\mathscr{A}(\mathscr{G})$ symmetry adapted on the right to $\mathscr{A}$, second for $\mathscr{A}^{\alpha}$ symmetry adapted on the right to $\mathscr{D}$, and third for $\mathscr{A}_{b}^{\beta}$ symmetry adapted on the left to $\mathscr{D}$. Then a typical three-fold product of such matric basis elements will be

$$e^{\gamma}_{g(\alpha,\alpha',a')}\; e^{\alpha}_{a(\hat{a}\delta d)}\; e^{\beta}_{(\hat{\ell}\delta' d')b}$$
$$= \delta(\alpha,\alpha')\delta(a,a')\delta(\delta,\delta')\delta(d,d')e^{\gamma}_{g(\alpha\alpha\hat{a}\delta d)}\; e^{\beta}_{(\hat{\ell}\delta d)b}, \tag{5.1}$$

where a relation as in (4.6) has been used. Now if one expands the skew-matric basis element $e^{\gamma}_{g(\alpha\alpha\hat{a}\delta d)}$ as in (4.5), then a typical nonzero three-fold product will be

$$e^{\gamma}_{g(\alpha\alpha a)}\; e^{\alpha}_{a(\hat{a}\delta d)}\; e^{\beta}_{(\hat{\ell}\delta d)b}$$
$$= \sum_{\ell\beta'\hat{\ell}'\delta' d'} (\alpha\alpha\hat{a}\delta d\,|\gamma|\,\ell\beta'\hat{\ell}'\delta' d')\, e^{\gamma}_{g(\ell\beta'\hat{\ell}'\delta' d')}\, e^{\beta}_{(\hat{\ell}\delta d)b}$$
$$= \sum_{\ell} (\alpha\alpha\hat{a}\delta|\gamma|\ell\beta\hat{\ell}\delta)\, e^{\gamma}_{g(\ell\beta b)}, \tag{5.2}$$

where also relations as in (4.6) and (4.7) have been used. Thus we have our first result:

Lemma 1: The space $\mathscr{A}(\mathscr{G})\mathscr{A}^{\alpha}\mathscr{A}_{b}^{\beta}$ is spanned by the group algebraic elements

$$\sum_{\ell} (\alpha\alpha\hat{a}\delta|\gamma|\ell\beta\hat{\ell}\delta)e^{\gamma}_{g(\ell\beta b)}, \quad \gamma g, \alpha\hat{a}\delta\hat{\ell} \text{ ranging.}$$

Similar considerations for the group-algebraic space carrying the semiregular subduced representation lead us to our second result:

Lemma 2: The space $\mathscr{A}(\mathscr{B})\mathscr{A}^{\alpha}\mathscr{A}_{g}^{\gamma}$ is spanned by the group-algebraic elements

$$\sum_{\ell} (\ell\beta\hat{\ell}\delta|\gamma|\alpha\alpha\hat{a}\delta)e^{\gamma}_{g(\ell\beta b)}, \quad \beta b, \alpha\hat{a}\delta\hat{\ell} \text{ ranging.}$$

As a consequence of Lemma 1 it is seen that $|\beta\hookrightarrow\alpha\uparrow\gamma|$ is the number of linearly independent elements appearing there for a given γ. But this is just the rank of the matrix $M^{\gamma\alpha\beta}$ with the element in the $(\alpha\hat{a}\delta\hat{\ell})$th row and the ℓth column being $(\alpha\alpha\hat{a}\delta\,|\gamma|\ell\beta\hat{\ell}\delta)$. Similarly Lemma 2 implies that $|\gamma\downarrow\alpha\hookrightarrow\beta|$ is just the rank of the transpose of this same $M^{\gamma\alpha\beta}$. Thus we have a reciprocitiy result:

Theorem: $|\beta\hookrightarrow\alpha\uparrow\gamma| = (\text{rank of } M^{\gamma\alpha\beta}) = |\gamma\downarrow\alpha\hookrightarrow\beta|$.

A character formula identifying whether the semiregular induction and subduction frequencies are zero or not is available.

Lemma 3. $|\beta\hookrightarrow\alpha\uparrow\gamma| \neq 0 \Leftrightarrow \chi^{\gamma}(e^{\alpha}e^{\beta}) \neq 0$.

Here e^{α} and e^{β} are central idempotents.

$$e^{\alpha} \equiv \frac{|\alpha|}{|\mathscr{A}|}\sum_{A\in\mathscr{A}}\chi^{\alpha}(A^{-1})A,$$
$$e^{\beta} \equiv \frac{|\beta|}{|\mathscr{B}|}\sum_{B\in\mathscr{B}}\chi^{\beta}(B^{-1})B \tag{5.3}$$

with χ^{α}, χ^{β}, and χ^{β} being characters, so that

$$\chi^{\gamma}(e^{\alpha}e^{\beta}) = \frac{|\alpha|}{|\mathscr{A}|}\frac{|\beta|}{|\mathscr{B}|}\sum_{A\in\mathscr{A}}\sum_{B\in\mathscr{B}}\chi^{\alpha}(A^{-1})\chi^{\beta}(B^{-1})\,\chi^{\gamma}(AB). \tag{5.4}$$

To prove Lemma 3 first rewrite this in terms of irreducible representation matrix elements

$$\chi^{\gamma}(e^{\alpha}e^{\beta}) = \frac{|\alpha|}{|\mathscr{A}|}\frac{|\beta|}{|\mathscr{B}|}\sum_{A\in\mathscr{A}}\sum_{B\in\mathscr{B}}\sum_{ab}\sum_{gg'}\Gamma^{\alpha}_{aa}(A^{-1})$$
$$\times\Gamma^{\beta}_{bb}(B^{-1})\Gamma^{\gamma}_{gg'}(A)\Gamma^{\gamma}_{g'g}(B). \tag{5.5}$$

Next choose Γ^{γ} to be sequence adapted for the chain $\mathscr{A}\subseteq\mathscr{G}$, so that $g = \alpha\alpha' a'$ and $g' = \alpha'\alpha'' a''$. Now from (4.1) it is seen that Kronecker deltas $\delta(\alpha',\alpha'')$ and $\delta(\alpha,\alpha')$ arise with the nonzero $\Gamma^{\gamma}_{gg'}(A)$ terms remaining as $\Gamma^{\alpha'}_{a'a''}(A)$. Application of the (well-known) orthogonality theorem for the $\Gamma^{\alpha}_{aa}(A^{-1})$ and $\Gamma^{\alpha'}_{a'a''}(A)$ irreducible representation matrix elements of $\mathscr{A}$, then leads to

$$\chi^{\gamma}(e^{\alpha}e^{\beta}) = \frac{|\beta|}{|\mathscr{B}|}\sum_{B\in\mathscr{B}}\sum_{ab}\sum_{\alpha}\Gamma^{\beta}_{bb}(B^{-1})\Gamma^{\gamma}_{(\alpha\alpha a)(\alpha\alpha a)}(B). \tag{5.6}$$

Next utilize recoupling coefficients to transform to a Γ^{γ} representation sequence adapted to $\mathscr{B}\subseteq\mathscr{G}$ and reduce the resulting Γ^{γ} representation matrix elements for $B\in\mathscr{B}$ to $\Gamma^{\beta'}$ representation matrix elements; thus

$$\Gamma^{\gamma}_{(\alpha\alpha a)(\alpha\alpha a)}(B) = \sum_{\ell\beta'}\sum_{b'b''}(\alpha\alpha a|\gamma|\ell\beta' b')$$
$$\times\Gamma^{\beta'}_{b'b''}(B)(\ell\beta' b''|\gamma|\alpha\alpha a). \tag{5.7}$$

Now substitute this into (5.6) and use the orthogonality theorem for irreducible representations of $\mathscr{B}$ to find

$$\chi^{\gamma}(e^{\alpha}e^{\beta}) = \sum_{ab}\sum_{\ell\alpha}|(\alpha\alpha a|\gamma|\ell\beta b)|^{2}. \tag{5.8}$$

Finally choosing the a and b labels to reflect sequence adaptation to $\mathscr{D}\subseteq\mathscr{A}$ and $\mathscr{D}\subseteq\mathscr{B}$, respectively, we obtain

$$\chi^{\gamma}(e^{\alpha}e^{\beta}) = \sum_{\ell\alpha}\sum_{\hat{a}\delta\hat{\ell}}|\delta|\,|(\alpha\alpha\hat{a}\delta\,|\gamma|\ell\beta\hat{\ell}\delta)|^{2}. \tag{5.9}$$

This equation just involves a weighted sum over the absolute-value squares of the matrix $M^{\gamma\alpha\beta}$; moveover the weights $|\delta|$ in this sum are positive. Thus $\chi^{\gamma}(e^{\alpha}e^{\beta})$ is nonzero if the rank of $M^{\gamma\alpha\beta}$ is nonzero, and recalling the reciprocity theorem we complete the proof.

6. FURTHER COMMENTS

There are some other symmetry adaptation processes which, although differing from regular and semiregular processes, can be simply expressed in terms of regular induction and subduction ideas. One such process involves symmetry adaptation to $\mathscr{A}$ of a basis of kets already symmetry adapted to β of $\mathscr{B}$; the (maximal) representation so obtained might be labeled $\beta \hookrightarrow \mathscr{A}$. A second process involves subduction to $\mathscr{A}$ of a space previously induced to $\mathscr{G}$ from β of $\mathscr{B}$; the (maximal) representation so obtained might be labeled $\beta \curvearrowright \mathscr{A}$. These representations can readily be verified to be expressible in terms of combinations of regular inductions and subductions

$$\beta \hookrightarrow \mathscr{A} = \dot{\sum_{\delta}} |\beta{\downarrow}\delta|\, \delta{\uparrow}\mathscr{A},$$
$$\beta \curvearrowright \mathscr{A} = \dot{\sum_{\gamma}} |\beta{\uparrow}\gamma|\, \gamma{\downarrow}\mathscr{A}. \tag{6.1}$$

Thus it seems that semiregular induction and subduction are the simplest nontrivial mathematical extensions of the classical (regular) induction and subduction ideas. (Also in terms of the notation introduced here it may be seen that the minimum of $|\beta{\uparrow}\gamma|$ and $|\beta \hookrightarrow \alpha||\alpha{\uparrow}\gamma|$ is an upper bound to $|\beta \hookrightarrow \alpha{\uparrow}\gamma|$.)

In addition to the examples of the Introduction there are other situations which may be interpreted in terms of semiregular induction. One example from nuclear physics involves[8] resonating group functions for the $1s^4 1p^2$ configuration; there the relevant groups are again permutation groups $\mathscr{G} = \mathscr{S}_6$, $\mathscr{A} = \mathscr{S}_3 \times \mathscr{S}_3$, $\mathscr{B} = \mathscr{S}_2 \times \mathscr{S}_2 \times \mathscr{S}_1 \times \mathscr{S}_1$ with $\alpha = [3]\times[3]$, $\beta = [2]\times[2]\times[1]\times[1]$, and $\beta \hookrightarrow \alpha{\uparrow} = [2,2,2] + [2,2,1,1]$. A second example involves the symmetric top, where the relevant groups involve rotations of space-fixed (SF) and body-fixed (BF) coordinates; there $\mathscr{G} = \mathscr{O}^+_{\rm SF}(3) \times \mathscr{O}^+_{\rm BF}(3)$, $\mathscr{A} = \mathscr{O}^+_{\rm SF}(3) \times 1$, and $\mathscr{B} = (\mathscr{O}^+_{\rm SF}(3) \times \mathscr{O}^+_{\rm BF}(3))_D$ with $\alpha = L$, $\beta = 0$, and $\beta \hookrightarrow \alpha{\uparrow} = L \times L$. A third example from mathematics involves Young's theory of the symmetric group; there $\mathscr{G} = \mathscr{S}_N$ while α and β are the symmetric and antisymmetric representations of the row and column groups for two Young diagrams $YD[\lambda]$ and $YD[\mu]$; if $[\lambda] = [\mu]$ then $\beta \hookrightarrow \alpha{\uparrow} = [\lambda]$, while for the more general case, $[\lambda] \neq [\mu]$, ideas involving the (Young) diagram lattice are[9] relevant.

Semiregular induction (and perhaps also subduction) seems to occur widely and frequently. Hence a general unifying theory of semiregular processes seems indicated. The presentation here is an attempt in this direction—to provide a common framework for discussion of various diverse examples and to give some general theoretical results.

Note added in proof. An approach to obtain extra group-theoretic labels for (subduced) symmetry adapted states, as described by Newman,[10] seems to be interpretable in terms of semiregular subduction.

ACKNOWLEDGMENTS

Interesting discussions on semiregular induction are acknowledged with Dr. W. Hässelbarth, Mr. A. Ludwig, Dr. R. Roel, Professor E. Ruch, and Professor T. H. Seligman.

[1] E. Ruch and A. Schönhofer, Theoret. Chim. Acta **3**, 291–304 (1965).
[2] J. S. Griffith, Mol. Phys. **8**, 217–224 (1964).
[3] M. G. Clark, Chem. Phys. Lett. **6**, 558–560 (1970); M. G. Clark, Mol. Phys. **20**, 257–269 (1971).
[4] E. Ruch and A. Schönhofer, Theoret. Chim. Acta **19**, 225–287 (1970).
[5] C. A. Mead, E. Ruch, and A. Schönhofer, Theoret. Chim. Acta **29**, 269–304 (1973); C. A. Mead, *Symmetry and Chirality*, Vol. 49 in *Topics in Current Chemistry* (Springer-Verlag, Berlin, 1974).
[6] G. F. Frobenius, Sitzber. Preuss. Akad. Wiss., Berlin 501–515 (1898).
[7] The nomenclature and results used here are as in D. J. Klein, in *Group Theory and Its Applications*, Vol. III, edited by E. M. Loebl (Academic, New York, 1975).
[8] K. T. Hecht, E. J. Reske, T. H. Seligman, and W. Zahn, Nuclear Phys. A **356**, 146–222 (1981).
[9] See, e.g., E. Ruch, Theoret. Chim. Acta **38**, 167–183 (1975); G. James and A. Kerber, *The Representation Theory of the Symmetric Group* (Addison-Wesley, Reading, MA, 1981).
[10] D. J. Newman, Phys. Lett. A **97**, 153 (1983).

Identities satisfied by the generators of the Dirac algebra

Patrick L. Nash
M. S. 474, NASA Langley Research Center, Hampton, Virginia 23665

(Received 26 July 1983; accepted for publication 23 September 1983)

The geometry of real four-dimensional spinor space and its symmetry groups are reviewed from the perspective of $\overline{\mathrm{SO}(3,3)}$. Two identities that concern the matrix generators of $\overline{\mathrm{SO}(3,3)}$, and which were first proved by Dirac, are generalized.

PACS numbers: 02.20. + b, 02.40. + m

1. INTRODUCTION

This paper contains several new results relating to $\overline{\mathrm{SO}(3,1)}$ spinor algebra that may be of general use. The main result, the lemma of Sec. 3, is a straightforward generalization of an identity discovered by Dirac, which is satisfied by the 4×4 matrix generators of the Dirac algebra. Section 2 is expository, $\overline{\mathrm{SO}(3,1)}$ spinor algebra is discussed in detail from the perspective of $\overline{\mathrm{SO}(3,3)}$. Section 4, provided in the interest of completeness, records the transformation properties of various geometric objects under $\overline{\mathrm{SO}(3,1)}$.

Notations and conventions used in this paper are as follows: upper case Latin indices run from 1 to 6, while both Greek and early lower case Latin indices run from 1 to 4. If M is a matrix, then $\widetilde{M}$ denotes the transpose of M. We work in a coordinate system such that the metric tensor $g_{\alpha\beta}$ on M_4 has components $g_{\alpha\beta} = \mathrm{diag}(1,1,1,-1)$.

2. THE γ^{AB} MATRICES AND $\overline{\mathrm{SO}(3,3)}$

Let $\gamma^{AB} = -\gamma^{BA} A,B,... = 1,...,6$ denote 15 elements, which are defined by[1-3]

$$\gamma^{AB}\gamma^{CD} = \gamma_0(g^{AD}g^{BC} - g^{AC}g^{BD}) - g^{AC}\gamma^{BD} + g^{AD}\gamma^{BC} + g^{BC}\gamma^{AD} - g^{BD}\gamma^{AC} - \tfrac{1}{2}\epsilon^{ABCDEF}g_{EG}g_{FH}\gamma^{GH}, \tag{1}$$

where

$$g_{AB} = g^{AB} = \mathrm{diag}(1,1,1,-1,-1,-1), \tag{2}$$

γ_0 is the identity element, and ϵ^{ABCDEF} is the totally antisymmetric Levi-Cività tensor-density of weight $+1$ in six dimensions, $\epsilon^{123456} = +1$. In virtue of Eq. (1), the set of elements $\{\pm\gamma_0, \pm\gamma^{AB}\}$ forms a finite group of order 32. We shall consider only real irreducible representations of this group in which the $\{\gamma_0,\gamma^{AB}\}$ are linearly independent. By Burnside's theorem,[4] a representation of a finite group of degree f is irreducible if and only if there occur f^2 linearly independent matrices in it; hence, the degree of this representation is four. Thus, each of the γ^{AB} is a real 4×4 matrix, and γ_0 is the 4×4 identity matrix. We shall denote the real four-dimensional vector space that carries this irreducible representation as D_4, and refer to D_4 as (real four-dimensional) Dirac space. The vectors of D_4 will be called (real) contravariant spinors, for reasons that will become apparent below. The elements of D_4^*, the vector space dual to D_4, will be called (real) covariant spinors.

On account of the defining relations of Eq. (1), one finds that

$$\gamma^{AB}\gamma^{CD} - \gamma^{CD}\gamma^{AB} = [\gamma^{AB},\gamma^{CD}] = -2(g^{AC}\gamma^{BD} - g^{AD}\gamma^{BC} - g^{BC}\gamma^{AD} + g^{BD}\gamma^{AC}), \tag{3}$$

so that the $-\tfrac{1}{2}\gamma^{AB}$ comprise a real 4×4 irreducible representation of a linearly independent basis of the $\overline{\mathrm{SO}(3,3)}$ Lie algebra, so(3,3). Moreover, Eq. (1) implies that each of the γ^{AB} matrices has square equal to $\pm\gamma_0$, and either commutes or anticommutes with any other γ^{RS} matrix. Given a particular γ^{AB}, there exists another γ matrix, say τ, which anticommutes with it. Thus trace $(\gamma^{AB}) = \mathrm{tr}(\tau^{-1}\tau\gamma^{AB}) = \mathrm{tr}(\tau\gamma^{AB}\tau^{-1}) = -\mathrm{tr}(\gamma^{AB}) = 0$. Since the γ^{AB} are trace-free and linearly independent, one deduces the well-known real Lie algebra isomorphism $\mathrm{so}(3,3)\cong\mathrm{sl}(4,\mathbb{R})$. Hence D_4 carries an irreducible representation of $\overline{\mathrm{SL}(4,\mathbb{R})}\cong\overline{\mathrm{SO}(3,3)}$; the vectors of D_4 are reduced $\overline{\mathrm{SO}(3,3)}$ spinors.

Under the involutive automorphism $\gamma^{AB}\to -\widetilde{\gamma}^{AB}$ of so(3,3), the Lie algebra decomposes into the eigenvalue (-1) and eigenvalue $(+1)$ subspaces corresponding to, respectively, the nine linearly independent real traceless symmetric 4×4 matrices, and the six linearly independent real skew-symmetric 4×4 matrices. The eigenvalue $(+1)$ subspace is the subalgebra so(4), which is the Lie algebra of $\overline{\mathrm{SO}(4)}$, the maximal compact subgroup of $\overline{\mathrm{SL}(4,\mathbb{R})}$. The subalgebra $\mathrm{so}(4)\cong\mathrm{su}(2)\dotplus\mathrm{su}(2)$ may be further decomposed into the even (eigenvalue $+1$) and odd (eigenvalue -1) subspaces of the linear transformation of so(4) whereby $\tau\in\mathrm{so}(4)$ is mapped into its dual, ${}^*\tau$. The even subspace under * of so(4) corresponds to self-dual tensors, and, say, the first su(2) in the direct sum; the odd subspace corresponds to anti-self-dual tensors, and the second su(2) in the direct sum. A basis for so(4) may be chosen as follows. Each of the six skew-symmetric γ matrices has the property that the square of the matrix is equal to $-\gamma_0$. By Eq. (1), these six matrices are given by $(h = 1,2,3)$,

$$2s^h = (\gamma^{23},\gamma^{31},\gamma^{21}), \tag{4}$$

and

$$2t^h = (\gamma^{45},\gamma^{64},\gamma^{65}). \tag{5}$$

From Eq. (3), these matrices verify $(h,k,m = 1,2,3)$

$$[s^h,t^k] = 0, \tag{6}$$

$$[s^h,s^k] = \epsilon^{hkm}s^m, \tag{7}$$

$$[t^h,t^k] = \epsilon^{hkm}t^m. \tag{8}$$

The s^h (resp. t^h) are anti-Hermitian generators of a real reducible unitary representation of SU(2). We shall assume that the s^h are self-dual, and the t^h are anti-self-dual. The six matrices s^h, t^h, comprise a linearly independent basis for the six-dimensional subalgebra so(4) of so(3,3).

The nine symmetric trace-free γ matrices may be denoted as $\gamma^{h'k}$, $h' = 1',2',3'$, where $1' = 6$, $2' = 5$, $3' = 4$. The $\gamma^{h'k}$ comprise a linearly independent basis for the nine-dimensional symmetric subspace of so(3,3). They may be expressed in terms of s^h, t^h as follows: contracting Eq. (1) with ϵ_{RSABCD} yields:

$$\gamma_{RS} = -(1/4!)\epsilon_{RSABCD}\gamma^{AB}\gamma^{CD}, \tag{9}$$

where $\gamma_{RS} = g_{RA}g_{SB}\gamma^{AB}$. Evaluating the left-hand side of Eq. (9) for $\gamma^{h'k}$ [after repeated use of Eq. (9)] gives

$$\gamma^{h'k} = -4g^{hn}t^n s^m g^{mk}, \tag{10}$$

where

$$g^{hk} = \mathrm{diag}(1,1,-1), \tag{11}$$

and, as we have heretofore implicitly assumed, the summation convention is operative for repeated indices; here m and n assume the values 1, 2, and 3.

Let $\gamma^{ABa}{}_b$ denote the ath row and bth column of γ^{AB}, where $a,b = 1,2,3,4$. A concrete representation of the γ^{AB} is $(h,k,m,n = 1,2,3)$,

$$2(s^h)^a{}_b = -\epsilon_{hab4} - \delta_{ah}\delta_{b4} + \delta_{a4}\delta_{bh} \quad \text{(self-dual)}, \tag{12}$$

$$2(t^h)^a{}_b = -\epsilon_{hab4} + \delta_{ah}\delta_{b4} - \delta_{a4}\delta_{bh} \quad \text{(anti-self-dual)}, \tag{13}$$

and

$$(\gamma^{h'k})^a{}_b = g^{hm}g^{kn}(\delta_{ab}\delta_{mn} - \delta_{am}\delta_{bn} - \delta_{an}\delta_{bm} - 2\delta_{a4}\delta_{b4}\delta_{mn} + \delta_{a4}\epsilon_{mnb4} + \delta_{b4}\epsilon_{mna4}), \tag{14}$$

where ϵ_{abcd} is the totally antisymmetric Levi-Cività tensor density of weight (-1) on D_4; $\epsilon_{1234} = +1$, g^{hm} as defined in Eq. (11); and we have substituted Eqs. (12) and (13) into Eq. (10) to obtain Eq. (14). Denoting the right-hand side of Eq. (12) by s^h_{ab}, by self-dual we mean that $s^h_{ab} = \frac{1}{2}\epsilon_{abcd}s^h_{cd}$.

There does not exist a $\overline{\mathrm{SO}(3,3)}$ invariant bilinear form (inner product) on D_4. The $\overline{\mathrm{SO}(3,3)}$ symmetry must be broken down to, say, $\overline{\mathrm{SO}(4)}$, or $\overline{\mathrm{SO}(3,2)}$ or $\overline{\mathrm{SO}(3,1)}$ in order to define an invariant bilinear form on D_4. To see this, suppose that $\tilde{\lambda}\epsilon\lambda'$ is a $\overline{\mathrm{SO}(3,3)}$ invariant bilinear form, where $\lambda,\lambda' \in D_4$, $\tilde{\lambda}$ denotes the transpose of λ, and ϵ is the "metric" spinor of covariant-rank two. Under $S = \exp(-\frac{1}{4}\omega_{AB}\gamma^{AB}) \in \overline{\mathrm{SO}(3,3)}$ (the $\omega_{AB} = -\omega_{BA}$ are 15 real parameters), $\lambda' \to S\lambda$ and $\tilde{\lambda} \to \tilde{\lambda}\tilde{S}$; in order for $\tilde{\lambda}\epsilon\lambda'$ to be an invariant under $\overline{\mathrm{SO}(3,3)}$ ϵ must be invariant under automorphism by S: $\epsilon \to \tilde{S}\epsilon S = \epsilon$. This is equivalent to $\tilde{\gamma}^{AB}\epsilon = -\epsilon\gamma^{AB}$ ($\tilde{\gamma}^{AB}$ denotes the transpose of γ^{AB}).

Let $\gamma^{AB} = s^h$ or t^h; then ϵ must commute with each of these matrices, since each is skew-symmetric. Hence ϵ commutes also with the products, as defined in Eq. (10), and thus ϵ commutes with every matrix in the irreducible representation. Therefore, by the second part of Schur's lemma,[4] ϵ is a numerical multiple of the unit matrix. However, each $\gamma^{h'k}$ is symmetric, and must therefore anticommute with ϵ: $\tilde{\gamma}^{h'k}\epsilon = \gamma^{h'k}\epsilon = -\epsilon\gamma^{h'k}$. Hence ϵ must be zero; there is no $\overline{\mathrm{SO}(3,3)}$ invariant bilinear form on the real vector space D_4. Another way to show this is to note that γ^{12}, γ^{34}, and γ^{56} commute and satisfy $\tilde{\gamma}^{12} = -\gamma^{12}$, $\tilde{\gamma}^{34} = \gamma^{34}$, $\tilde{\gamma}^{56} = -\gamma^{56}$, and $\gamma^{12}\gamma^{56} = \gamma^{34}$. However, $(-\tilde{\gamma}^{12})(-\tilde{\gamma}^{56}) = \gamma^{34} \neq -\tilde{\gamma}^{34}$, so that $-\tilde{\gamma}^{AB}$ is not equivalent to γ^{AB}: $-\tilde{\gamma}^{AB}\epsilon = \epsilon\gamma^{AB} \Rightarrow \epsilon = 0$.

There are a number of bilinear forms on D_4 that are defined by a nonsingular covariant rank-two spinor ϵ, which are invariant under a subgroup of $\overline{\mathrm{SO}(3,3)}$. If ϵ is symmetric, $\tilde{\epsilon} = \epsilon$, then $\frac{1}{2}\epsilon\gamma^{AB}\omega_{AB}$ is skew-symmetric:

$$\tfrac{1}{2}\omega_{AB}\tilde{\gamma}^{AB}\epsilon = \tfrac{1}{2}\omega_{AB}\tilde{\gamma}^{AB}\tilde{\epsilon} = \widetilde{\tfrac{1}{2}\epsilon\gamma^{AB}\omega_{AB}} = -\tfrac{1}{2}\epsilon\gamma^{AB}\omega_{AB}.$$

Since there are six linearly-independent skew-symmetric real 4×4 matrices, the maximal subgroup of $\overline{\mathrm{SO}(3,3)}$ that leaves ϵ invariant corresponds to the six-parameter subgroup of $\overline{\mathrm{SO}(3,3)}$ generated by $\{s^h,t^h\}$, namely, a $\overline{\mathrm{SO}(4)}$ subgroup of $\overline{\mathrm{SO}(3,3)}$. A $\overline{\mathrm{SO}(4)}$ invariant inner product may be defined on D_4 utilizing a symmetric ϵ.

If ϵ is skew-symmetric, $\tilde{\epsilon} = -\epsilon$; then $\frac{1}{2}\epsilon\gamma^{AB}\omega_{AB}$ is symmetric

$$\tfrac{1}{2}\omega_{AB}\tilde{\gamma}^{AB}\epsilon = -\tfrac{1}{2}\tilde{\gamma}^{AB}\tilde{\epsilon} = -\widetilde{\tfrac{1}{2}\epsilon\gamma^{AB}\omega_{AB}} = -\tfrac{1}{2}\epsilon\gamma^{AB}\omega_{AB}.$$

Since there are ten linearly-independent real symmetric 4×4 matrices, ϵ defines a nonsingular skew-symmetric bilinear form on D_4 whose maximal invariance group is one of the six possible ten-parameter subgroups $\overline{\mathrm{SO}(3,2)}$ and $\overline{\mathrm{SO}(2,3)}$ of $\overline{\mathrm{SO}(3,3)}$ that are generated by ten of the fifteen $-\frac{1}{2}\gamma^{AB}$. (Which particular subgroup, of course, depends upon the choice of ϵ.) Since ϵ defines a symplectic form on D_4, one deduces the real Lie algebra isomorphisms $\mathrm{so}(3,2) \cong \mathrm{sp}(2,\mathbb{R}) \cong \mathrm{so}(2,3)$, where $\mathrm{sp}(n,\mathbb{R})$ is the real symplectic Lie algebra whose defining representation is of degree $2n$.

$\overline{\mathrm{SO}(3,1)}$ is a subgroup of $\overline{\mathrm{SO}(3,2)}$, but not of $\overline{\mathrm{SO}(2,3)}$, so that most interest lies with $\overline{\mathrm{SO}(3,2)}$ invariant-symplectic forms ϵ. There are essentially three distinct choices for ϵ, namely, γ^{45}, γ^{56}, or γ^{64}. From

$$\tfrac{1}{2}\omega_{AB}\tilde{\gamma}^{AB}\epsilon = -\tfrac{1}{2}\epsilon\gamma^{AB}\omega_{AB} \tag{15}$$

and Eq. (1), one concludes the following:

(i) If $\epsilon = \gamma^{45}$, then one must set $\omega_{A6} = 0$ in order to satisfy Eq. (15); the generators of this $\overline{\mathrm{SO}(3,2)}$ are therefore $\{-\frac{1}{2}\gamma^{\alpha\beta}, -\frac{1}{2}\gamma^{\alpha 5}\}$, where $\alpha,\beta = 1,2,3,4$.

(ii) If $\epsilon = \gamma^{56}$, then one must set $\omega_{A4} = 0$; the generators are $\{-\frac{1}{2}\gamma^{hk}, -\frac{1}{2}\gamma^{h5}, -\frac{1}{2}\gamma^{h6}, -\frac{1}{2}\gamma^{56}; h,k = 1,2,3\}$.

(iii) If $\epsilon = \gamma^{64}$, then one must set $\omega_{A5} = 0$; the generators are $\{-\frac{1}{2}\gamma^{\alpha\beta}, -\frac{1}{2}\gamma^{\alpha 6}\}$.

$\overline{\mathrm{SO}(3,3)}$ transformations on D_4 may be associated with SO(3,3) transformations on a flat six-dimensional (three space, three time) Minkowski space-time M_6, whose metric tensor is given by Eq. (2). By restriction to an appropriate four-dimensional affine subspace of M_6, we can realize M_4. For the sake of simplicity, we shall assume that the x^4 axis of M_4 coincides with the x^4 axis of M_6 in every coordinate system. It is customary to exclude choice (ii), $\epsilon = \gamma^{56}$, as an interesting symplectic form on D_4. γ^{56} is invariant under those automorphisms of D_4 that correspond with the automorphisms of M_6 that leave the x^4 axis of M_6 invariant.

Which of the candidates, γ^{45} or γ^{64}, that is adopted for ϵ depends upon the association defined between $\gamma^{\alpha 5}$ and $\gamma^{\alpha 6}$, and Dirac's γ^α matrices, and is also based on the fact that one must restrict the $\overline{\mathrm{SO}(3,2)}$ symmetry to a $\overline{\mathrm{SO}(3,1)}$ subgroup in order to be in accordance with relativity. As things stand, case (i) $\epsilon = \gamma^{45}$, implies that $\epsilon\gamma^{\alpha 6}$ is antisymmetric; $\gamma^{\alpha 5}$ mixes with $\gamma^{\alpha\beta}$ under SO(3,2), while $\{\gamma^{\alpha 6},\gamma^{56}\}$ is a SO(3,2) vector ($\omega_{A6} = 0$) (the transformation properties of the γ matrices are discussed in Sec. 4). Case (iii), $\epsilon = \gamma^{64}$, implies that $\epsilon\gamma^{\alpha 6}$ is symmetric, while $\epsilon\gamma^{\alpha 5}$ is skew-symmetric; $\gamma^{\alpha 6}$ mixes with $\gamma^{\alpha\beta}$ under SO(3,2), while $\{\gamma^{\alpha 5},\gamma^{65}\}$ is a SO(3,2) vector ($\omega_{A5} = 0$).

Equivalent formalisms are: case (i), $\epsilon = \gamma^{45}$; define $\gamma^\alpha = \gamma^{\alpha 5}$, and append to the constraint $\omega_{A6} = 0$, the restriction $\omega_{A5} = 0$, so that $\gamma^{\alpha 5}$ transforms as a vector under SO(3,1); case (iii), $\epsilon = \gamma^{64}$; define $\gamma^\alpha = \gamma^{\alpha 6}$, and append to the constraint $\omega_{A5} = 0$ the restriction $\omega_{A6} = 0$, so that $\gamma^{\alpha 6}$ transforms as a vector under SO(3,1). In both cases, the $\overline{\mathrm{SO}(3,2)}$ symmetry is reduced to $\overline{\mathrm{SO}(3,1)}$.

Without loss of generality, we shall utilize $\epsilon = \gamma^{64}$ as the symplectic form. In order to make contact with the usual conventions found in the literature, it is convenient to make the following definitions.

Let γ^α (Greek indices run from 1 to 4) denote four real 4×4 matrices (Dirac's γ matrices) that generate an irreducible representation of the pseudo-Clifford algebra C_4 (also known as the Dirac algebra). The γ^α are defined by

$$\gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\alpha = 2\gamma_0 g^{\alpha\beta}, \tag{16}$$

where

$$g^{\alpha\beta} = g_{\alpha\beta} = \mathrm{diag}(1,1,1,-1) \tag{17}$$

is the metric tensor on M_4, in a Cartesian coordinate system. Let

$$\begin{aligned}\gamma^5 &= -(1/4!)\epsilon_{\alpha\beta\mu\nu}\gamma^\alpha\gamma^\beta\gamma^\mu\gamma^\nu \\ &= -\gamma^1\gamma^2\gamma^3\gamma^4,\end{aligned} \tag{18}$$

where $\epsilon_{\alpha\beta\mu\nu}$ is the totally antisymmetric Levi-Cività tensor density of weight (-1) in four dimensions, $\epsilon_{1234} = +1$. A representation of a linearly independent basis for C_4 is

$$\gamma^\alpha = \gamma^{\alpha 6}, \tag{19}$$

$$\gamma^5 = \gamma^{56}, \tag{20}$$

$$\gamma^\alpha\gamma^5 = \gamma^{\alpha 5}, \tag{21}$$

and defining

$$S^{\alpha\beta} = -\tfrac{1}{4}[\gamma^\alpha,\gamma^\beta], \tag{22}$$

$$S^{\alpha\beta} = -\tfrac{1}{2}\gamma^{\alpha\beta}. \tag{23}$$

The symplectic form ϵ on D_4 is defined to be

$$\epsilon = \gamma^{64}. \tag{24}$$

As a consequence of Eq. (1), and the definitions of Eqs. (16)–(24), are the identities

$$\tilde{\gamma}^\alpha\epsilon = -\epsilon\gamma^\alpha, \tag{25}$$

$$\tilde{S}^{\alpha\beta}\epsilon = -\epsilon S^{\alpha\beta}, \tag{26}$$

$$[S^{\alpha\beta},\gamma_\mu] = \delta^\alpha_\mu\gamma^\beta - \delta^\beta_\mu\gamma^\alpha, \tag{27}$$

$$[S^{\alpha\beta},S^{\mu\nu}] = g^{\alpha\mu}S^{\beta\nu} - g^{\alpha\nu}S^{\beta\mu} - g^{\beta\mu}S^{\alpha\nu} + g^{\beta\nu}S^{\alpha\mu}, \tag{28}$$

and

$$\gamma^5 S^{\alpha\beta} = \tfrac{1}{2}g^{\alpha\mu}g^{\beta\nu}\epsilon_{\mu\nu\lambda\sigma}S^{\lambda\sigma}. \tag{29}$$

We introduce a $\overline{\mathrm{SO}(3,1)}$ index notation to compliment the matrix notation which we have been using. Associate $\overline{\mathrm{SO}(3,1)}$ indices as follows: $D_4 \ni \lambda \leftrightarrow \lambda^a$; $D_4^* \ni \xi \leftrightarrow \xi_a$ (note that in matrix notation, $\xi\lambda$ denotes $\xi_a\lambda^a$, while $\lambda\xi$ denotes the 4×4 matrix with elements $\lambda^a\xi_b$; one has $\mathrm{tr}\,\lambda\xi = \xi\lambda$); $\epsilon \leftrightarrow \epsilon_{ab} = -\epsilon_{ba}$; $\tilde{\lambda}\epsilon \leftrightarrow \lambda_b = \lambda^a\epsilon_{ab}$, where the tilde denotes the transpose of a matrix (mnemonic $b\leftrightarrow$ below); raise $\overline{\mathrm{SO}(3,1)}$ indices with ϵ^{ab} according as $\xi^a = \epsilon^{ab}\xi_b$ (mnemonic: $a\leftrightarrow$ above). According to this convention $\epsilon^{ab} = \epsilon^{ac}\epsilon^{bd}\epsilon_{cd} = \epsilon^{ac}(\epsilon^{bd}\epsilon_{cd}) = \epsilon^{ac}\delta^b_c$; therefore,

$$(\epsilon^{-1})^{ab} = \epsilon^{ba} = -\epsilon^{ab}, \tag{30}$$

$$\epsilon^{ac}\epsilon_{cb} = -\delta^a_b, \tag{31}$$

and we find the correspondence $\xi^a \leftrightarrow -\epsilon^{-1}\tilde{\xi}$. $\gamma^{AB} \leftrightarrow \gamma^{ABa}{}_b$. In index notation, Eq. (25) is $\gamma^{\alpha c}{}_a\epsilon_{cb} = -\epsilon_{ac}\gamma^{\alpha c}{}_b = \epsilon_{ca}\gamma^{\alpha c}{}_b$, i.e.,

$$\gamma^\alpha{}_{ba} = \gamma^\alpha{}_{ab}. \tag{32}$$

Equation (26) is $S^{\alpha\beta c}{}_a\epsilon_{cb} = -\epsilon_{ac}S^{\alpha\beta c}{}_b$,

$$S^{\alpha\beta}_{ba} = S^{\alpha\beta}_{ab}. \tag{33}$$

$\tilde{S}\epsilon S = \epsilon, S \in \overline{\mathrm{SO}(3,1)}$, reads

$$S^c{}_a\epsilon_{cd}S^d{}_b = \epsilon_{ab}. \tag{34}$$

Since $\tilde{\gamma}^{\alpha 5}\epsilon = \epsilon\gamma^{\alpha 5}$,

$$\gamma^{\alpha 5}_{ba} = -\gamma^{\alpha 5}_{ab}; \tag{35}$$

similarly, $\tilde{\gamma}^5\epsilon = \epsilon\gamma^5$, so that

$$\gamma^5_{ba} = -\gamma^5_{ab}. \tag{36}$$

The determinant of ϵ is given by $\det\epsilon = \epsilon^{abcd}\epsilon_{a1}\epsilon_{b2}\epsilon_{c3}\epsilon_{d4}$, or equivalently, $\epsilon_{a'b'c'd'}\det\epsilon = \epsilon^{abcd}\epsilon_{aa'}\epsilon_{bb'}\epsilon_{cc'}\epsilon_{dd'}$; since $\epsilon^2 = -\gamma_0$, ϵ has eigenvalues $\pm i$; since $\mathrm{tr}\,\epsilon = 0$, the eigenvalues occur with equal multiplicity. Hence $\det\epsilon = 1$, and thus

$$\epsilon_{a'b'c'd'} = \epsilon_{aa'}\epsilon_{bb'}\epsilon_{cc'}\epsilon_{dd'}\epsilon^{abcd}. \tag{37}$$

The fact that the s^h are self-dual, and the t^h are anti-self-dual, may be expressed covariantly in both matrix and index notation. Since $\epsilon^{-1} = -\epsilon$, ${}^*\epsilon = -\epsilon$ may be written as

$$\epsilon = {}^*\epsilon^{-1}; \tag{38}$$

in index notation, Eq. (38) is $\epsilon_{ab} = \frac{1}{2}\epsilon_{abcd}(\epsilon^{-1})^{cd} = \frac{1}{2}\epsilon_{abcd}\epsilon^{dc}$, or

$$\epsilon_{ab} = -\tfrac{1}{2}\epsilon_{abcd}\epsilon^{cd}. \tag{39}$$

Using Eq. (1), one finds that $-\epsilon\gamma^\alpha\gamma^5 = \tilde{\gamma}^\alpha\gamma^5\epsilon^{-1} = 2\delta^\alpha_h g^{hk} s^k - \delta^\alpha_4\gamma^5$. Therefore,

$$-\epsilon\gamma^\alpha\gamma^5 = {}^*\gamma^\alpha\gamma^5\epsilon^{-1} \tag{40}$$

expresses the fact that the s^h are self-dual for $\alpha = 1,2,3$, and that γ^5 is anti-self-dual when $\alpha = 4$. Noting that

$$(-\epsilon\gamma^{AB})_{ab} = \gamma^{AB}_{ab}(= -\epsilon_{ac}\gamma^{ABc}{}_b = \epsilon_{ca}\gamma^{ABc}{}_b = \gamma^{AB}_{ab}),$$

and

$$(\gamma^{AB}\epsilon^{-1})^{ab} = \gamma^{ABab}(= \gamma^{ABa}{}_c\epsilon^{-1cb} = \gamma^{ABa}{}_c\epsilon^{bc} = \gamma^{ABab}),$$

Eq. (40) may be expressed as $(\gamma^\alpha\gamma^5)_{ab} = \frac{1}{2}\epsilon_{abcd}(\gamma^\alpha\gamma^5)^{cd}$, which, using Eq. (21) yields

$$\gamma^{\alpha 5}_{ab} = \tfrac{1}{2}\epsilon_{abcd}\gamma^{\alpha 5cd}. \tag{41}$$

Lastly, from Eq. (1), $-\epsilon\gamma^5 = \gamma^{45} = -\gamma^5\epsilon^{-1}$, which combined with ${}^*\gamma^{45} = -\gamma^{45}$ gives

$$-\epsilon\gamma^5 = {}^*\gamma^5\epsilon^{-1}, \tag{42}$$

i.e.,

$$\gamma^5_{ab} = \tfrac{1}{2}\epsilon_{abcd}\gamma^{5cd}. \tag{43}$$

As an application of Eq. (39), we evaluate

$$\begin{aligned}\epsilon^{da}\epsilon^{bc} + \epsilon^{db}\epsilon^{ca} + \epsilon^{dc}\epsilon^{ab} &= \tfrac{1}{2}\epsilon^{da'}\epsilon^{b'c'}\delta^{abc}_{a'b'c'}\\ &= \tfrac{1}{2}\epsilon^{da'}\epsilon^{b'c'}\delta^{abce}_{a'b'c'e}\\ &= \tfrac{1}{2}\epsilon^{abce}\epsilon_{a'b'c'e}\epsilon^{da'}\epsilon^{b'c'}\\ &= -\epsilon^{abce}\epsilon^{da'}\epsilon_{a'e}\ \text{[using Eq. (39)]}\\ &= \epsilon^{abcd}\ \text{[using Eq. (31)]}.\end{aligned}$$

Thus

$$\epsilon^{abcd} = \epsilon^{da}\epsilon^{bc} + \epsilon^{db}\epsilon^{ca} + \epsilon^{dc}\epsilon^{ab}. \tag{44}$$

(See Ref. 5 for a clear exposition of the properties of the generalized Kronecker delta, ϵ^{abcd}, and ϵ_{abcd}.)

3. A BASIC LEMMA

Lemma: Let X be an arbitrary 4×4 matrix; then

$$\begin{aligned}&\gamma^{56}X\gamma^{56} + \gamma^{64}\tilde{X}\gamma^{64} + \gamma^{45}\tilde{X}\gamma^{45}\\ &\quad = X - \gamma_0 \operatorname{tr} X + \gamma^{56}\operatorname{tr}\gamma^{56}X,\end{aligned} \tag{45}$$

where $\tilde{X}$ denotes the transpose of X, and tr X is the trace of X. This identity is valid for any cyclic permutation of $(\gamma^{56},\gamma^{64},\gamma^{45})$, and under the replacement $\gamma^{56}\to\gamma^{12}$, $\gamma^{64}\to\gamma^{31}$, and $\gamma^{45}\to\gamma^{23}$.

Proof: Eq. (45) is linear in X; we verify that this equation is true for $X = \gamma_0$, γ^{56}, $\gamma^{\alpha 5}$, $\gamma^{\alpha 6}$, and $\gamma^{\alpha\beta}$. Note that only for $X = \gamma_0$ (resp γ^{56}) is tr X (resp tr $\gamma^{56}X$) nonvanishing.

(i) $X = \gamma_0$; since $(\gamma^{56})^2 = -\gamma_0 = (\gamma^{64})^2 = (\gamma^{45})^2$, Eq. (45) yields

$$\begin{aligned}&(\gamma^{56})^2 + (\gamma^{64})^2 + (\gamma^{45})^2 = -3\gamma_0\\ &\quad = \gamma_0 - \gamma_0\operatorname{tr}\gamma_0 + \gamma^{56}\operatorname{tr}\gamma^{56} = \gamma_0 - 4\gamma_0;\end{aligned}$$

(ii) $X = \gamma^{56}$; since $\tilde{\gamma}^{56} = -\gamma^{56}$, and γ^{56} anticommutes with both γ^{64} and γ^{45}; Eq. (45) gives

$$\begin{aligned}&-\gamma^{56} - \gamma^{64}\gamma^{56}\gamma^{64} - \gamma^{45}\gamma^{56}\gamma^{45} = -3\gamma^{56}\\ &\quad = \gamma^{56} - \gamma_0\operatorname{tr}\gamma^{56} + \gamma^{56}\operatorname{tr}(-\gamma_0) = \gamma^{56} - 4\gamma^{56};\end{aligned}$$

(iii) $X = \gamma^{\alpha 5}$; from Eq. (1),

$\gamma^{\alpha 5}\gamma^{56} = -\gamma^{56}\gamma^{\alpha 5}$, $\tilde{\gamma}^{\alpha 5}\gamma^{64} = \gamma^{64}\gamma^{\alpha 5}$, and $\tilde{\gamma}^{\alpha 5}\gamma^{45} = -\gamma^{45}\gamma^{\alpha 5}$; by Eq. (45),

$$\begin{aligned}&\gamma^{56}\gamma^{\alpha 5}\gamma^{56} + \gamma^{64}\tilde{\gamma}^{\alpha 5}\gamma^{64} + \gamma^{45}\tilde{\gamma}^{\alpha 5}\gamma^{45}\\ &\quad = (\gamma^{56})^2(-\gamma^{\alpha 5}) + (\gamma^{64})^2\gamma^{\alpha 5} + (\gamma^{45})^2(-\gamma^{\alpha 5}) = \gamma^{\alpha 5};\end{aligned}$$

(iv) $X = \gamma^{\alpha 6}$; one deduces from Eq. (1) that

$\gamma^{\alpha 6}\gamma^{56} = -\gamma^{56}\gamma^{\alpha 6}$, $\tilde{\gamma}^{\alpha 6}\gamma^{64} = -\gamma^{64}\gamma^{\alpha 6}$, and $\tilde{\gamma}^{\alpha 6}\gamma^{45} = \gamma^{45}\gamma^{\alpha 6}$; Eq. (45) becomes

$$\begin{aligned}&\gamma^{56}\gamma^{\alpha 6}\gamma^{56} + \gamma^{64}\tilde{\gamma}^{\alpha 6}\gamma^{64} + \gamma^{45}\tilde{\gamma}^{\alpha 6}\gamma^{45}\\ &\quad = (\gamma^{56})^2(-\gamma^{\alpha 6}) + (\gamma^{64})^2(-\gamma^{\alpha 6}) + (\gamma^{45})^2\gamma^{\alpha 6} = \gamma^{\alpha 6};\end{aligned}$$

(v) $X = \gamma^{\alpha\beta}$; from Eq. (1),

$\gamma^{\alpha\beta}\gamma^{56} = \gamma^{56}\gamma^{\alpha\beta}$, $\tilde{\gamma}^{\alpha\beta}\gamma^{64} = -\gamma^{64}\gamma^{\alpha\beta}$, and $\tilde{\gamma}^{\alpha\beta}\gamma^{45} = -\gamma^{45}\gamma^{\alpha\beta}$; Eq. (45) gives

$$\begin{aligned}&\gamma^{56}\gamma^{\alpha\beta}\gamma^{56} + \gamma^{64}\tilde{\gamma}^{\alpha\beta}\gamma^{64} + \gamma^{45}\tilde{\gamma}^{\alpha\beta}\gamma^{45}\\ &\quad = (\gamma^{56})^2\gamma^{\alpha\beta} + (\gamma^{64})^2(-\gamma^{\alpha\beta}) + (\gamma^{45})^2(-\gamma^{\alpha\beta}) = \gamma^{\alpha\beta}.\end{aligned}$$

■

Since $\gamma^{56} = \gamma^5$, $\gamma^{64} = \epsilon = -\epsilon^{-1}$, and $\gamma^{45} = \gamma^{46}\gamma^{56} = -\epsilon\gamma^5 = -\gamma^5\epsilon^{-1}$, Eq. (45) may be written covariantly as

$$\gamma^5 X\gamma^5 - \epsilon^{-1}\tilde{X}\epsilon + \gamma^5\epsilon^{-1}\tilde{X}\epsilon\gamma^5 = X - \gamma_0\operatorname{tr}X + \gamma^5\operatorname{tr}\gamma^5 X.$$

Bringing X to the left-hand side of this equation, and then multiplying by $-\gamma^5$ gives

$$[X + \epsilon^{-1}\tilde{X}\epsilon,\gamma^5]_+ = \gamma^5\operatorname{tr}X + \gamma_0\operatorname{tr}\gamma^5 X, \tag{46}$$

where $[A,B]_+ = AB + BA$ denotes the anticommutator of A and B. In index notation, $X + \epsilon^{-1}\tilde{X}\epsilon$ is $X^a{}_b + \epsilon^{-1ac}X^d{}_c\epsilon_{db} = X^a{}_b + \epsilon^{ca}X^d{}_c\epsilon_{db} = X^a{}_b - \epsilon_{db}X^d{}_c\epsilon^{ac} = X^a{}_b - X_b{}^a$; Eq. (46) can be written as

$$\begin{aligned}&(X^a{}_c - X_c{}^a)\gamma^{5c}{}_b + \gamma^{5a}{}_c(X^c{}_b - X_b{}^c)\\ &\quad = \gamma^{5a}{}_b X^c{}_c + \delta^a_b\gamma^{5c}{}_d X^d{}_c.\end{aligned} \tag{47}$$

Eq. (45) is a simple but useful identity, and is a generalization of an identity first proved by Dirac[2] in 1963. The assertion that Eq. (45) is valid under permutation of γ^{56}, γ^{64}, γ^{45} is true because, as far as so(3,3) is concerned, no su(2) generator is to be preferred over the remaining two. Eq. (45) remains valid under the replacement $\gamma^{h'k'}\to\gamma^{hk}$ because of the symmetric roles played by the two su(2) subalgebras in the direct sum of so(4) [self-dual and anti-self-dual, required in order that the six skew-symmetric matrices be linearly independent, plays no role in Eq. (45)].

As an application of this lemma, we prove that

$$\gamma_\alpha\lambda\xi\gamma^\alpha = \gamma_0\xi\lambda + \gamma^5\xi\gamma^5\lambda + \epsilon^{-1}\tilde{\xi}\tilde{\lambda}\epsilon + \gamma^5\epsilon^{-1}\tilde{\xi}\tilde{\lambda}\epsilon\gamma^5. \tag{48}$$

The starting point of this evaluation is to replace γ^{56}, γ^{64}, and γ^{45} in Eq. (45) with, respectively, γ^{12}, γ^{23}, γ^{31}. This yields

$$\gamma^{12}X\gamma^{12} + \gamma^{23}\tilde{X}\gamma^{23} + \gamma^{31}\tilde{X}\gamma^{31} = X - \gamma_0\operatorname{tr}X + \gamma^{12}\operatorname{tr}\gamma^{12}X. \tag{49}$$

Let X be an arbitrary symmetric matrix, $\tilde{X} = X$; then $\operatorname{tr}\gamma^{12}X = 0$, because X may be expanded in terms of γ_0 and $\gamma^{h'k}$; each of these matrices, when multiplied by γ^{12}, has vanishing trace [see Eq. (10)]. Consider

$$\gamma^{34}(\gamma^{12}X\gamma^{12}+\gamma^{23}X\gamma^{23}+\gamma^{31}X\gamma^{31})\gamma^{34}$$
$$=\gamma^5X\gamma^5-\gamma^{24}X\gamma^{24}-\gamma^{14}X\gamma^{14}$$
[using Eqs. (1) and (9)]
$$=\gamma^{34}(X-\gamma_0\,\mathrm{tr}\,X)\gamma^{34}=\gamma^{34}X\gamma^{34}-\gamma_0\,\mathrm{tr}\,X.$$

Hence $\gamma^{h4}X\gamma^{h4}=\gamma_0\,\mathrm{tr}\,X+\gamma^5X\gamma^5$; since X is an arbitrary symmetric matrix, this implies that (for convenience we write $\gamma^{ABa}{}_b$ as $\gamma^{AB}{}_{ab}$ in this paragraph)

$$\gamma^{h4}_{ab}\gamma^{h4}_{cd}+\gamma^{h4}_{ac}\gamma^{h4}_{bd}=2\delta_{ad}\delta_{bc}+\gamma^5_{ab}\gamma^5_{cd}+\gamma^5_{ac}\gamma^5_{bd}.$$

Holding d fixed, one may obtain two similar equations by cyclically permuting (a,b,c). Upon adding two of these equations and subtracting the third, one finds that

$$\gamma^{h4}_{ab}\gamma^{h4}_{cd}=-\delta_{ab}\delta_{cd}+\delta_{ac}\delta_{bd}+\delta_{ad}\delta_{bc}+\gamma^5_{ac}\gamma^5_{bd}+\gamma^5_{ad}\gamma^5_{bc}.$$

Contracting this result with $\lambda_b\xi_c$ yields

$$\gamma^{h4}\lambda\xi\gamma^{h4}=-\lambda\xi+\tilde{\xi}\tilde{\lambda}+\gamma_0\xi\lambda+\gamma^5\tilde{\xi}\tilde{\lambda}\gamma^5-\gamma^5\xi\gamma^5\lambda.$$

Therefore,

$$\gamma_4(\gamma^{h4}\lambda\xi\gamma^{h4}+\lambda\xi)\gamma^4=\gamma^{h4}\gamma^{46}\lambda\xi\gamma^{h4}\gamma^{46}+\gamma_4\lambda\xi\gamma^4,$$
$$\gamma_\alpha\lambda\xi\gamma^\alpha=\epsilon^{-1}(\tilde{\xi}\tilde{\lambda}+\gamma_0\xi\lambda+\gamma^5\tilde{\xi}\tilde{\lambda}\gamma^5-\gamma^5\xi\gamma^5\lambda)\epsilon$$
$$=\gamma_0\xi\lambda+\gamma^5\xi\gamma^5\lambda+\epsilon^{-1}\tilde{\xi}\tilde{\lambda}\epsilon$$
$$+\gamma^5\epsilon^{-1}\tilde{\xi}\tilde{\lambda}\epsilon\gamma^5.$$

An orthogonality relationship satisfied by the γ^{AB} is

$$-\tfrac{1}{2}\gamma^{ABa}{}_b\gamma_{AB}{}^c{}_d=4\delta^a_d\delta^c_b-\delta^a_b\delta^c_d. \tag{50}$$

To prove this, construct a matrix $Y(X)=-\frac{1}{2}\gamma^{AB}X\gamma_{AB}+X$, where X is an arbitrary 4×4 matrix. Since $(\gamma^{AB})^{-1}=-\gamma_{AB}$, this may be written as $Y=X+\Sigma_{i=1}^{15}\gamma^iX(\gamma^i)^{-1}$, where $\gamma^{AB}\leftrightarrow\gamma^i, i=1,\dots,15$. If τ is any γ^{AB} matrix, then $\tau Y=\tau X\tau^{-1}\tau+\Sigma_{i=1}^{15}\tau\gamma^iX(\tau\gamma^i)^{-1}\tau=Y\tau$; hence $\gamma^{AB}Y=Y\gamma^{AB}$, and since the γ^{AB} comprise an irreducible set, Y is a multiple $T^c{}_bX^b{}_c$ of γ_0: $\delta^a_dT^c_bX^b_c=X^b_c(-\frac{1}{2}\gamma^{ABa}{}_b\gamma_{AB}{}^c{}_d+\delta^a_b\delta^c_d)$. Thus $\delta^a_dT^c{}_b=-\frac{1}{2}\gamma^{ABa}{}_b\gamma_{AB}{}^c{}_d+\delta^a_b\delta^c_d$, which implies $T^a{}_b=-\frac{1}{2}\gamma^{ABc}{}_b\gamma_{AB}{}^a{}_c+\delta^a_b=4\delta^a_b$, which in turn implies Eq. (50).

4. TRANSFORMATION PROPERTIES OF γ^{AB}

The $\gamma^{ABa}{}_b$ are numerically invariant under combined SO(3,3) transformations of M_6 indices $\{A,B\}$, and $\overline{\mathrm{SO}(3,3)}$ transformations of spinor indices $\{a,b\}$. To see this, suppose $L^A{}_B\in\mathrm{SO}(3,3)$; then the metric on M_6, g_{AB} of Eq. (2), is invariant under automorphism by L:

$$g_{AB}\to L^C{}_Ag_{CD}L^D{}_B=g_{AB}. \tag{51}$$

The matrices $L^A{}_CL^B{}_D\gamma^{CD}$ satisfy Eq. (1) on account of Eq. (51), and the fact that $(L^A{}_B$ $\det(L^A{}_B)=1$: $\epsilon^{A'B'C'D'E'F'}L^A{}_{A'}L^B{}_{B'}L^C{}_{C'}L^D{}_{D'}L^E{}_{E'}L^F{}_{F'}=\det L\cdot\epsilon^{ABCDEF}$. Hence the $L^A{}_CL^B{}_D\gamma^{CD}$ provide a real 4×4 irreducible representation of the group. The sum of the squares of the degrees of the irreducible representations of the group equals the order of the group, $32=1^2+4^2+\cdots$ (the degree one irrep is the trivial representation), so that there can be only one irreducible representation of degree four. Therefore, $L^A{}_CL^B{}_D\gamma^{CD}$ is equivalent to γ^{AB}, there exists a real nonsingular 4×4 matrix $S=S(L)$ such that

$$\gamma^{AB}=L^A{}_CL^B{}_DS\gamma^{CD}S^{-1}. \tag{52}$$

S may be assumed to have determinant equal to $+1$, and is determined up to a factor of ±1. The set of all such matrices S provides an irreducible representation $\overline{\mathrm{SL}(4,\mathbb{R})}\cong\overline{\mathrm{SO}(3,3)}$.

A special Lorentz transformation $x\to x'=Lx$ on M_6 is accompanied by a $\overline{\mathrm{SO}(3,3)}$ transformation on D_4: $\lambda\to\lambda'=S\lambda$. By Eq. (3), S is generated by $-\frac{1}{2}\gamma^{AB}$; for if $L^A{}_B=\delta^A{}_B-\omega^A{}_B+\cdots=(e^{-\omega})^A{}_B$, where $\omega_{AB}=-\omega_{BA}$ are 15 real parameters, then

$$S=\gamma_0-\tfrac{1}{4}\omega_{AB}\gamma^{AB}+\cdots=\exp\{-\tfrac{1}{4}\omega_{AB}\gamma^{AB}\} \tag{53}$$

satisfies Eq. (52).

One can construct a 2–1 representation of $\overline{\mathrm{SO}(3,3)}$ onto SO(3,3) as follows. Let Γ^A denote six matrices defined by

$$\Gamma^h=2g^{hk}\begin{pmatrix}0 & s^k\\ -s^k & 0\end{pmatrix} \tag{54}$$

and

$$\Gamma^{h'}=2g^{hk}\begin{pmatrix}0 & t^k\\ t^k & 0\end{pmatrix}, \tag{55}$$

where $h,k=1,2,3$; $h'=1',2',3'$ and $1'=6$, $2'=5$, $3'=4$; and $g^{hk}=\mathrm{diag}(1,1,-1)$. One may easily verify that the Γ^A satisfy

$$\Gamma^A\Gamma^B+\Gamma^B\Gamma^A=2g^{AB}I, \tag{56}$$

where I denotes the 8×8 unit matrix

$$-\tfrac{1}{4}[\Gamma^A,\Gamma^B]=-\tfrac{1}{2}\begin{pmatrix}\gamma^{AB} & 0\\ 0 & -\tilde{\gamma}^{AB}\end{pmatrix}; \tag{57}$$

and

$$\Gamma^A=L^A{}_BM\Gamma^BM^{-1}, \tag{58}$$

where

$$M=\begin{pmatrix}S & 0\\ 0 & \tilde{S}^{-1}\end{pmatrix}, \tag{59}$$

S being defined in Eq. (53). Therefore, given $S\in\overline{\mathrm{SO}(3,3)}$, the map $\overline{\mathrm{SO}(3,3)}\to\mathrm{SO}(3,3)$ defined by

$$L^A{}_B=\tfrac{1}{8}\,\mathrm{tr}\,(M^{-1}\Gamma^AM\Gamma_B) \tag{60}$$

is a 2–1 representation of $\overline{\mathrm{SO}(3,3)}$ onto SO(3,3).

Concomitant with the identification of γ^{64} as a $\overline{\mathrm{SO}(3,1)}$ invariant symplectic form on D_4 is the reduction of $\overline{\mathrm{SO}(3,3)}$ symmetry to $\overline{\mathrm{SO}(3,1)}$ defined by setting $\omega_{A5}=0=\omega_{A6}$. According to this restriction, we have $L^A{}_6=\delta^A_6$, $L^A{}_5=\delta^A_5$, and $L^\alpha{}_\beta=\delta^\alpha_\beta-\omega^A{}_\beta+\cdots=(e^{-\omega})^\alpha{}_\beta$, where $\omega_{\alpha\beta}=-\omega_{\beta\alpha}$ are six real parameters. $S\in\overline{\mathrm{SO}(3,1)}$ is given by

$$S(\omega)=\exp\{\tfrac{1}{2}\omega_{\alpha\beta}S^{\alpha\beta}\}, \tag{61}$$

where $S^{\alpha\beta}$ is defined in Eqs. (22) and (23). Under the restriction to a $\overline{\mathrm{SO}(3,1)}$ subgroup of $\overline{\mathrm{SO}(3,3)}$, the γ^{AB} decompose into sets transforming as tensors under $\overline{\mathrm{SO}(3,1)}$:

$$\epsilon\to\tilde{S}\epsilon S=\epsilon, \tag{62}$$
$$\gamma^\alpha\to L^\alpha{}_\beta S\gamma^\beta S^{-1}=\gamma^\alpha, \tag{63}$$

$$\gamma^5 \to S\gamma^5 S^{-1} = \gamma^5, \tag{64}$$

and

$$S^{\alpha\beta} \to L^{\alpha}{}_{\mu} L^{\beta}{}_{\nu} S S^{\mu\nu} S^{-1} = S^{\alpha\beta}. \tag{65}$$

Let $\lambda \in D_4$ and $\xi \in D_4^*$; under $\overline{\mathrm{SO}(3,1)}$,

$$\lambda \to S\lambda \tag{66}$$

and

$$\xi \to \xi S^{-1}. \tag{67}$$

ACKNOWLEDGMENT

Most of this work was conducted in the Department of Mathematics at the University of Arizona at Tucson. I would like to thank the Department for its hospitality during my most enjoyable stay in Tucson.

[1] Sir A. S. Eddington, *Fundamental Theory* (Cambridge U. P., Cambridge, 1949), Section 54.

[2] P. A. M. Dirac, J. Math. Phys. **4**, 901 (1963).

[3] P. L. Nash, J. Math. Phys. **21**, 1024 (1980).

[4] H. Boerner, *Representation of Groups* (North-Holland, Amsterdam, 1969).

[5] D. Lovelock and H. Rund, *Tensors, Differential Forms, and Variational Principles* (Wiley, New York, 1975).

Infinitesimal symmetry transformations of some one-dimensional linear systems[a)]

M. Aguirre
Instituto de Física, Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile
J. Krause
Facultad de Física, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile

(Received 20 July 1983; accepted for publication 16 August 1983)

The converse problem of similarity analysis is solved in general for the infinitesimal symmetry transformations of any given inhomogeneous ordinary differential equation of the second order $\ddot{x} + f_2(t)\dot{x} + f_1(t)x = f_0(t)$. The completely general associated Lie algebra is obtained for equations of this kind, which structure constants depend only on the chosen set of initial values $f_0(0)$, $f_1(0)$, $f_2(0)$, and $\dot{f}_2(0)$. The infinitesimal elements of the dynamical group of a Newtonian one-dimensional linear system are also discussed, and some miscellaneous examples are given.

PACS numbers: 02.20. + b, 03.20. + i, 02.30.Hq

1. INTRODUCTION

Recently there has been an increasing interest in studying the symmetry principles involved in classical mechanics.[1] This interest is motivated by the identity of symmetry groups operating in classical mechanics and quantum mechanics [as, for instance, the O(4) symmetry which operates equally for the hydrogen atom as for the classical Keplerian system[2]]. Indeed, there is the feeling that canonical quantization is not the main point in the transition from a classical model to the corresponding quantum formalism.[3] Rather, this common symmetry structure should be the guide for having a complete and unambiguous quantization procedure. Thus, several "geometric quantization" schemes can be found in the current literature.[4] As a consequence, if one assumes that the complete set of symmetry operations performed on an isolated system uniquely characterizes the system, then a fundamental problem arises, namely, that of finding the symmetry group associated with a given mechanical system.

In this paper we examine some features of this problem by means of the similarity analysis of a single particle's motion in Newtonian mechanics. For the sake of simplicity, in this note we shall only dwell on linear systems with one degree of freedom. Specifically, here we obtain the infinitesimal symmetry groups and present the Lie algebras associated with some second-order ordinary differential equations of inherent relevance in mechanics.[5]

There are several motives for formulating this particular problem. First, the techniques themselves bear some mathematical interest, since they are simple and general. Indeed, we wish to remark that the problem one usually tackles in similarity analysis is the determination of the general form of the differential equations (of some required order), which admit a *given group* as a symmetry group.[6] The converse, and more interesting problem, of finding the Lie group of symmetries of a *given differential equation* is much more difficult to solve. Clearly, it is this converse problem which we enface in this paper, and we claim that some systematic methods exist for handling it.

Furthermore, in classical mechanics one usually visualizes the symmetries of a system by means of transformations which leave the Lagrangian (or the Hamiltonian) invariant.[7] But dynamical systems, in general, have more symmetries than those grasped by the symmetry groups of their Lagrangians or Hamiltonians.[8] One reason for this stems from the fact that there may exist symmetry operations of the equations of motion which induce a gauge transformation of the Lagrangian[9] (and, thus, of the Hamiltonian). One avoids this gauge artifact (inherent in the canonical formalism and contrived to produce only a subgroup of the full symmetry group) by considering the symmetries of the equations of motion themselves.[10]

Moreover, the "cloud of mystery" regarding the origin of the accidental symmetries[11] neatly shows the necessity of having a general procedure for investigating the symmetry group of any given system. It seems that there has been some lack of progress concerning this problem in the past because one has thought that the intimate relation between conservation laws and symmetries has to be traced to the Lagrangian and Hamiltonian formulations.[12] However, the recent progress in the use of methods of continuous groups of transformations and local differential geometry in the study of the equations of motion of classical mechanics[13] reveals that the demand of invariance of equations of motion yields not only the conventional conservation laws, but also the (so-called) accidental symmetries. Further, this approach throws new light on the relation between symmetries in configuration space and phase space,[14] is capable of generalization to an arbitrary system,[15] and may become relevant for quantization.

It is well known, from the history of mathematics, that the enormous bulk of knowledge in the field of differential equations was strikingly coordinated by the work of Sophus Lie.[16] The symmetries considered by Lie's approach to the field of differential equations (i.e., similarity analysis) are of a

[a)] Supported in part (M. A.) by Dirección General de Investigación de la Universidad Católica de Valparaíso, through DGI Grant No. 123.726/82, and in part (J. K.) by Dirección de Investigación de la Pontificia Universidad Católica de Chile, through DIUC Grant No. 50/81.

very restricted kind, to be sure. In effect (for the case of one-dimensional systems), in that approach one considers only the symmetries generated by infinitesimal point transformations, of the form

$$t' = t + \epsilon\eta(t,x)\,,\qquad x' = x + \epsilon\theta(t,x)\,, \tag{1.1}$$

where ϵ denotes a parameter of smallness: $0 < \epsilon \ll 1$ (together with the induced, or extended, transformations on $\dot{x}$, $\ddot{x}$, etc.), in order to study the possibility of transforming among themselves the solutions of an ordinary differential equation:

$$G(t,x,\dot{x},\ddot{x},\cdots) = 0\,, \tag{1.2}$$

say. It must be observed that, in contrast with the well-known general result concerning first-order differential equations,[17] differential equations of order higher than one only exceptionally admit continuous groups of symmetry transformations. In particular, if a second-order differential equation admits continuous symmetry groups like (1.1), these monoparametric groups generate a Lie group having no more than eight essential parameters.[18] Thus, for instance, the equation $\ddot{x} = 0$ admits the eight-parameter projective group of the (t,x) plane,[19] while the equation $\ddot{x} = t^2 + x^2$ admits no proper symmetry transformation of the form (1.1), but the trivial one $t' = t$ and $x' = x$.[20]

Quite generally, one calls a symmetry of Eq. (1.2) *any* prescription for transforming any solution of Eq. (1.2) into another solution of the same equation. It is clear that, by defining the composition of symmetries in the usual manner, one gets associativity, and therefore the set of all symmetries of Eq. (1.2) is a semigroup with identity (i.e., a monoid). It seems, however, that the above definition of symmetry is too broad. One can be led to very complicated and diverse prescriptions giving the analytical descriptions of symmetries (global integral symmetries, local differential symmetries, space-time point symmetries, etc.) pertaining to a concrete differential equation. There is, indeed, a "*debarré d'excess*," since many (perhaps useful) recipes may arise as possible symmetries according to the general definition stated above.

For instance, an interesting approach has been considered by Gonzalez-Gascon,[21] following a generalization of Lie's concept of invariance of differential equations due to Anderson *et al.*,[22] while introducing the notion of local differential symmetries. The local differential symmetry prescriptions on t and x (for one-dimensional systems) implicitly include the transformations on $\dot{x}$, $\ddot{x}$, etc., and are assumed to transform any solution of Eq. (1.2) into another solution. The set of local differential symmetries forms a monoid, which does not behave as an infinitesimal group. Neither does it generate a finite Lie group, since the iterative integration process is impossible, because one would need to handle implicity *all* the derivatives $d^n x/dt^n$, $n = 1,2,\cdots$, in order to integrate. This is in strong contrast with Lie's infinitesimal point transformations, for it can be easily shown that the set of all infinitesimal pointlike transformations generates a continuous Lie group (as is well known, indeed: Lie's theorem), while the set of local differential symmetry transformations does not. We think that this remark is important, because in mechanics one is interested in symmetries not only as a tool for obtaining new solutions of the equations of motion. Rather, one usually extracts dynamical information from a given symmetry, whenever the symmetry is organized within a group structure. (This is so, even when the laws of motion are unknown, as is the case for much of elementary particle physics.) Anyhow, it is the *group*-theoretic viewpoint that makes symmetry a remarkably fruitful aid for our understanding of dynamics, and many of the relevant concepts on this issue arise already in classical mechanics.

This paper is organized as follows. We first discuss the infinitesimal symmetry transformations of a given second-order linear inhomogeneous ordinary differential equation, and formally obtain the associated Lie algebra (Sec. 2). Then we briefly present some miscellaneous examples of the previous formalism for the converse problem of similarity in mechanics (Sec. 3). Finally, we analyze the infinitesimal elements of the dynamical group of some one-dimensional Newtonian linear systems (Sec. 4).

2. THE INFINITESIMAL SYMMETRY TRANSFORMATIONS OF $\ddot{x} + f_2\dot{x} + f_1 x = f_0$ AND THE ASSOCIATED LIE ALGEBRA

Let us assume we are interested in the symmetry properties of an ordinary differential equation of the second order, of the general form

$$\ddot{x} = f(t,x,\dot{x})\,, \tag{2.1}$$

where f is a given function. The symmetry group of this equation is realized by the set of point transformations

$$t' = T(t,x)\,,\qquad x' = S(t,x)\,, \tag{2.2}$$

with nonvanishing Jacobian, endowed with the property of leaving Eq. (2.1) form invariant. Clearly, these transformations are active point transformations which transform one solution of Eq. (2.1) into another.

Of course, to solve this similarity problem one needs to find first the generating functions $\eta(t,x)$ and $\theta(t,x)$ [cf. Eqs. (1.1)] of the symmetry transformation. The corresponding twice extended transformations, up to the first order of approximation, are[23]

$$\dot{x}' = \dot{x} + \epsilon(\theta_t + (\theta_x - \eta_t)\dot{x} - \eta_x\dot{x}^2)\,, \tag{2.3}$$

and

$$\begin{aligned}\ddot{x}' = \ddot{x} &+ \epsilon((\theta_x - 2\eta_t) - 3\eta_x\dot{x})\ddot{x}\\ &+ \epsilon(\theta_{tt} + (2\theta_{xt} - \eta_{tt})\dot{x}\\ &+ (\theta_{xx} - 2\eta_{xt})\dot{x}^2 - \eta_{xx}\dot{x}^3)\,,\end{aligned} \tag{2.4}$$

where we have written $\dot{x}' = dx'/dt'$ and $\ddot{x}' = d\dot{x}'/dt'$. Hence, from the assumed invariance of Eq. (2.1), the following first-order relation obtains:

$$\begin{aligned}\theta_{tt} &+ (2\theta_{xt} - \eta_{tt})\dot{x} + (\theta_{xx} - 2\eta_{xt})\dot{x}^2 - \eta_{xx}\dot{x}^3\\ &+ ((\theta_x - 2\eta_t) - 3\eta_x\dot{x})f(t,x,\dot{x}) - \eta f_t(t,t,\dot{x}) - \theta f_x(t,x,\dot{x})\\ &- (\theta_t + (\theta_x - \eta_t)\dot{x} - \eta_x\dot{x}^2)f_{\dot{x}}(t,x,\dot{x}) \equiv 0\,,\end{aligned} \tag{2.5}$$

which holds equally for all $(t,x,\dot{x})$. This equation is the starting point in the similarity analysis of the second-order differential equation (2.1).

Further analysis then requires the explicit form of the function $f(t,x,\dot{x})$ to be known. Because it corresponds to a sufficiently general situation, of much physical relevance, in this paper we consider the linear inhomogeneous ordinary differential equation of the second order, i.e.,

$$\ddot{x} + f_2(t)\dot{x} + f_1(t)x = f_0(t)\,. \tag{2.6}$$

Let us substitute Eq. (2.6) into Eq. (2.5). We immediately get

$$\eta_{xx} = 0\,, \tag{2.7}$$

$$\theta_{xx} - 2\eta_{xt} + 2f_2\eta_x = 0\,, \tag{2.8}$$

$$2\theta_{xt} - \eta_{tt} - 3(f_0 - f_1x)\eta_x + f_2\eta_t + \dot{f}_2\eta = 0\,, \tag{2.9}$$

$$\begin{aligned}&\theta_{tt} + (f_0 - f_1x)\theta_x + f_2\theta_t + f_1\theta \\ &\quad - 2(f_0 - f_1x)\eta_t - (\dot{f}_0 - \dot{f}_1x)\eta = 0\,.\end{aligned} \tag{2.10}$$

Thus, from the first two equations above it follows that

$$\eta(t,x) = \phi_1(t)x + \phi_2(t) \tag{2.11}$$

and

$$\theta(t,x) = (\dot{\phi}_1(t) - f_2(t)\phi_1(t))x^2 + \phi_3(t)x + \phi_4(t)\,, \tag{2.12}$$

where $\phi_1, \ldots, \phi_4$ are functions of t, which shall be formally determined from the last two equations [Eqs. (2.9) and (2.10)]. Indeed, after performing some reductions, one easily arrives at the following system of linear homogeneous differential equations:

$$\begin{aligned}&\ddot{\phi}_1 - f_2\dot{\phi}_1 + (f_1 - \dot{f}_2)\phi_1 = 0\,, \\ &\dddot{\phi}_2 + (4f_1 - f_2^2 - 2\dot{f}_2)\dot{\phi}_2 + (2\dot{f}_1 - f_2\dot{f}_2 - \ddot{f}_2)\phi_2 \\ &\qquad = (f_0f_2 - \dot{f}_0)\phi_1 - 3f_0\dot{\phi}_1\,, \\ &2\dot{\phi}_3 = 3f_0\phi_1 + \ddot{\phi}_2 - f_2\dot{\phi}_2 - \dot{f}_2\phi_2\,, \\ &\ddot{\phi}_4 + f_2\dot{\phi}_4 + f_1\phi_4 = 2f_0\dot{\phi}_2 + \dot{f}_0\phi_2 - f_0\phi_3\,,\end{aligned} \tag{2.13}$$

wherefrom, clearly Eqs. (2.11) and (2.12) become

$$\eta(t,x) = q^a(\phi_{1.a}(t)x + \phi_{2.a}(t))\,, \tag{2.14}$$

$$\theta(t,x) = q^a(\{\dot{\phi}_{1.a}(t) - f_2(t)\phi_{1.a}(t)\}x^2 + \phi_{3.a}(t)x + \phi_{4.a}(t))\,. \tag{2.15}$$

The set of functions $\phi_{1.a}(t), \ldots, \phi_{4.a}(t)$, $a = 1,\ldots,8$, corresponds to a basis of the linear system (2.13), and the q^a, $a = 1, \ldots, 8$, are eight constants of integration, which behave as a set of eight essential parameters of the Lie group; i.e., in the neighborhood of the identity we have $\delta q^a = \epsilon q^a$, with $a = 1, \ldots, 8$ and $0 < \epsilon \ll 1$.

Of course, in order to determine a set of basis functions $\phi_{1.a}(t), \ldots, \phi_{4.a}(t)$ the functions $f_0(t)$, $f_1(t)$, $f_2(t)$, should be known. However, we have already enough information to formally obtain the Lie algebra of this symmetry group, as we shall see in what follows.

From the previous discussion we conclude that the full symmetry group of the linear differential equation (2.6) has the following set of infinitesimal operators:

$$X_a(t,x) = \eta_a(t,x)\partial_t + \theta_a(t,x)\partial_x\,, \tag{2.16}$$

where

$$\eta_a(t,x) = \phi_{1.a}(t)x + \phi_{2.a}(t)\,, \tag{2.17}$$

$$\begin{aligned}\theta_a(t,x) = {}&\{\dot{\phi}_{1.a}(t) - f_2(t)\phi_{1.a}(t)\}x^2 \\ &+ \phi_{3.a}(t)x + \phi_{4.a}(t)\,,\end{aligned} \tag{2.18}$$

$a = 1, \ldots, 8$, are the infinitesimal generators of this group in the (t,x)-realization and in some q-parametrization. These operators satisfy the algebra

$$[X_a, X_b] = f^c_{ab}X_c\,; \tag{2.19}$$

so we have

$$f^c_{ab}\eta_c = [\eta_a, \eta_{bt}] + [\theta_a, \eta_{bx}]\,, \tag{2.20}$$

$$f^c_{ab}\theta_c = [\eta_a, \theta_{bt}] + [\theta_a, \theta_{bx}]\,, \tag{2.21}$$

where, of course, the square brackets denote antisymmetrization of the indices "a" and "b" only. From Eqs. (2.17) and (2.18), we have

$$\begin{aligned}\eta_{at} &= \dot{\phi}_{1.a}x + \dot{\phi}_{2.a}\,, \\ \eta_{ax} &= \phi_{1.a}\,, \\ \theta_{at} &= -f_1\phi_{1.a}x^2 + \dot{\phi}_{3.a}x + \dot{\phi}_{4.a}\,, \\ \theta_{ax} &= 2(\dot{\phi}_{1.a} - f_2\phi_{1.a})x + \phi_{3.a}\,,\end{aligned} \tag{2.22}$$

where we have also made use of Eqs. (2.13). Hence, a straightforward calculation gives

$$f^c_{ab}\phi_{1.c} = [\phi_{1.a}, \dot{\phi}_{2.b} - \phi_{3.b}] - [\dot{\phi}_{1.a}, \phi_{2.b}]\,, \tag{2.23}$$

$$f^c_{ab}\phi_{2.c} = [\phi_{2.a}, \dot{\phi}_{2.b}] - [\phi_{1.a}, \phi_{4.b}]\,, \tag{2.24}$$

$$\begin{aligned}f^c_{ab}\phi_{3.c} = {}&[\phi_{1.a}, \dot{\phi}_{4.b} + 2f_2\phi_{4.b} - \tfrac{3}{2}f_0\phi_{2.b}] \\ &- 2[\dot{\phi}_{1.a}, \phi_{4.b}] + \tfrac{1}{2}[\phi_{2.a}, \ddot{\phi}_{2.b} - f_2\dot{\phi}_{2.b}]\,,\end{aligned} \tag{2.25}$$

$$f^c_{ab}\phi_{4.c} = [\phi_{2.a}, \dot{\phi}_{4.b}] - [\phi_{3.a}, \phi_{4.b}]\,, \tag{2.26}$$

and

$$\begin{aligned}&f^c_{ab}(\dot{\phi}_{1.c} - f_2\phi_{1.c}) \\ &\quad = \tfrac{1}{2}[\phi_{1.a}, \ddot{\phi}_{2.b} - f_2\dot{\phi}_{2.b} + (2f_1 - \dot{f}_2)\phi_{2.b} \\ &\qquad + 2f_2\phi_{3.b}] - [\dot{\phi}_{1.a}, \phi_{3.b}]\,.\end{aligned} \tag{2.27}$$

These are identities which hold for all t; thus let us consider them at $t = 0$.

This brings into the fore the initial value problem of Eqs. (2.13). For reasons which shall become clear presently, we introduce the following parametrization to represent the Lie algebra; we first define the initial data:

$$\begin{aligned}&q^1 = \eta(0,0)\,, \quad q^2 = \theta(0,0)\,, \\ &q^3 = \eta_t(0,0)\,, \quad q^4 = \theta_x(0,0)\,, \\ &q^5 = \eta_x(0,0)\,, \quad q^6 = \theta_t(0,0)\,, \\ &q^7 = \tfrac{1}{2}\eta_{tt}(0,0)\,, \quad q^8 = \tfrac{1}{2}\theta_{xx}(0,0)\,,\end{aligned} \tag{2.28}$$

and then use these q's as the essential parameters of the group. So, one easily finds

$$\begin{aligned}&\phi_{1.a}(0) = \delta_{a5}\,, \quad \dot{\phi}_{1.a}(0) = \delta_{a8} + f_2(0)\delta_{a5}\,, \\ &\phi_{2.a}(0) = \delta_{a1}\,, \quad \dot{\phi}_{2.a}(0) = \delta_{a3}\,, \\ &\ddot{\phi}_{2.a}(0) = 2\delta_{a7}\,, \quad \phi_{3.a}(0) = \delta_{a4}\,, \\ &\phi_{4.a}(0) = \delta_{a2}\,, \quad \dot{\phi}_{4.a}(0) = \delta_{a6}\,,\end{aligned} \tag{2.29}$$

as the chosen values of the linear basis.

Of course, a glance at Eqs. (2.23)–(2.27) shows that, in order to take advantage of the whole set of chosen initial data given in Eqs. (2.29), we have to make some derivatives in these equations [(2.23)–(2.27)], and substitute from Eqs.

(2.13) where needed. In this manner, we obtain

$$f^{c}_{ab}\,\dot{\phi}_{1.c} = \tfrac{1}{2}\,[\,\phi_{1.a},\dddot{\phi}_{2.b} + f_2\ddot{\phi}_{2.b} + (2f_1 - \dot{f}_2)\,\dot{\phi}_{2.b}\,] - [\,\dot{\phi}_{1.a},\phi_{3.b} + f_2\phi_{2.b}\,]\,, \tag{2.30}$$

$$f^{c}_{ab}\,\dot{\phi}_{2.c} = -[\,\phi_{1.a},\dot{\phi}_{4.b}\,] - [\,\dot{\phi}_{1.a},\phi_{4.b}\,] + [\,\phi_{2.a},\ddot{\phi}_{2.b}\,]\,, \tag{2.31}$$

$$\begin{aligned} f^{c}_{ab}\,\ddot{\phi}_{2.c} &= [\,\phi_{1.a}, f_0(\,\phi_{3.b} - 2\dot{\phi}_{2.b} - f_2\phi_{2.b}) + f_2\dot{\phi}_{4.b} + (2f_1 - \dot{f}_2)\,\phi_{4.b}\,] \\ &\quad + [\,\dot{\phi}_{1.a}, 3f_0\phi_{2.b} - 2\dot{\phi}_{4.b} - f_2\phi_{4.b}\,] \\ &\quad - (4f_1 - f_2^{\,2} - 2\dot{f}_2)\,[\,\phi_{2.a},\dot{\phi}_{2.b}\,] + [\,\dot{\phi}_{2.a},\ddot{\phi}_{2.b}\,]\,, \end{aligned} \tag{2.32}$$

$$\begin{aligned} f^{c}_{ab}\,\dot{\phi}_{4.c} &= f_0\,[\,\phi_{2.a}, 2\dot{\phi}_{2.b} - \phi_{3.b}\,] + [\,\dot{\phi}_{2.a} - f_2\phi_{2.a} - \phi_{3.a},\dot{\phi}_{4.b}\,] \\ &\quad + \tfrac{1}{2}\,[\,-3f_0\phi_{1.a} - \ddot{\phi}_{2.a} + f_2\dot{\phi}_{2.a} - (2f_1 - \dot{f}_2)\,\phi_{2.a},\phi_{4.b}\,]\,. \end{aligned} \tag{2.33}$$

Therefore, upon substituting from the initial data (2.29) into Eqs. (2.23)–(2.26) and Eqs. (2.30)–(2.33), the structure constants may be evaluated, as referred to the chosen basis of the algebra. We present our results in Table I. In Table II we display the Lie algebra associated with the symmetry point transformations of the general linear inhomogeneous ordinary differential equation of the second order. A detailed study of this algebra will be published elsewhere.[5]

Interesting enough, the algebra is completely general for differential equations of the form (2.6), and one obtains the structure constants without having a detailed knowledge of the infinitesimal operators $X_a(t,x)$, which, of course, would require the knowledge of the functions $f_0(t)$, $f_1(t)$, and $f_2(t)$. The algebra, however, depends formally and exclusively on the initial values $f_0(0)$, $f_1(0)$, $f_2(0)$, and $\dot{f}_2(0)$.

The outcoming commutation relations also depend on the chosen set of initial conditions; but this change is inher-

TABLE I. The nonzeroth structure constants of the Lie algebra associated with the linear differential equation $\ddot{x} + f_2(t)\dot{x} + f_1(t)x = f_0(t)$.

$f^{1}_{13} = 1,\quad f^{1}_{25} = 1$
$f^{2}_{16} = 1,\quad f^{2}_{24} = 1$
$f^{3}_{17} = 2,\quad f^{3}_{25} = f_2(0),\quad f^{3}_{28} = 1,\quad f^{3}_{56} = -1$
$f^{4}_{13} = -\frac{1}{2}f_2(0),\quad f^{4}_{15} = \frac{3}{2}f_0(0)$ $f^{4}_{17} = 1,\quad f^{4}_{28} = 2,\quad f^{4}_{56} = 1$
$f^{5}_{15} = f_2(0),\quad f^{5}_{18} = 1,\quad f^{5}_{35} = -1,\quad f^{5}_{45} = 1$
$f^{6}_{12} = -f_1(0) + \frac{1}{2}\dot{f}_2(0),\quad f^{6}_{13} = 2f_0(0)$ $f^{6}_{14} = -f_0(0),\quad f^{6}_{16} = -f_2(0)$ $f^{6}_{23} = -\frac{1}{2}f_2(0),\quad f^{6}_{25} = \frac{3}{2}f_0(0)$ $f^{6}_{27} = 1,\quad f^{6}_{36} = 1,\quad f^{6}_{46} = -1$
$f^{7}_{13} = -2f_1(0) + \dot{f}_2(0) + \frac{1}{2}f^{2}_{2}(0),\quad f^{7}_{15} = -f_0(0)f_2(0)$ $f^{7}_{18} = -\frac{3}{2}f_0(0),\quad f^{7}_{25} = -f_1(0) + \frac{1}{2}\dot{f}_2(0) + \frac{1}{2}f^{2}_{2}(0)$ $f^{7}_{28} = \frac{1}{2}f_2(0),\quad f^{7}_{35} = f_0(0),\quad f^{7}_{37} = 1$ $f^{7}_{45} = -\frac{1}{2}f_0(0),\quad f^{7}_{56} = -\frac{1}{2}f_2(0),\quad f^{7}_{68} = 1$
$f^{8}_{15} = -f_1(0) + \frac{1}{2}\dot{f}_2(0),\quad f^{8}_{35} = \frac{1}{2}f_2(0)$ $f^{8}_{48} = 1,\quad f^{8}_{57} = 1$

TABLE II. The Lie algebra associated with the differential equation $\ddot{x} + f_2(t)\dot{x} + f_1(t)x = f_0(t)$. One gets the commutator $[X_a, X_b]$ at the intersection of the ath row with the bth column.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	0	$\{\frac{1}{2}\dot{f}_2(0) - f_1(0)\}X_6$	$X_1 - \frac{1}{2}f_2(0)X_4 + 2f_0(0)X_6 + \{\frac{1}{2}f^{2}_{2}(0) + \dot{f}_2(0) - 2f_1(0)\}X_7$	$-f_0(0)X_6$	$\frac{3}{2}f_0(0)X_4 + f_2(0)X_5 - f_0(0)f_2(0)X_7 + \{\frac{1}{2}\dot{f}_2(0) - f_1(0)\}X_8$	$X_2 - f_2(0)X_6$	$2X_3 + X_4$	$X_5 - \frac{3}{2}f_0(0)X_7$
X_2	$\{f_1(0) - \frac{1}{2}\dot{f}_2(0)\}X_6$	0	$-\frac{1}{2}f_2(0)X_6$	X_2	$X_1 + f_2(0)X_3 + \frac{3}{2}f_0(0)X_6 + \{\frac{1}{2}f^{2}_{2}(0) + \frac{1}{2}\dot{f}_2(0) - f_1(0)\}X_7$	0	X_6	$X_3 + 2X_4 + \frac{1}{2}f_2(0)X_7$
X_3	$-X_1 + \frac{1}{2}f_2(0)X_4 - 2f_0(0)X_6 + \{2f_1(0) - \dot{f}_2(0) - \frac{1}{2}f^{2}_{2}(0)\}X_7$	$\frac{1}{2}f_2(0)X_6$	0	0	$-X_5 + f_0(0)X_7 + \frac{1}{2}f_2(0)X_8$	X_6	X_7	0
X_4	$f_0(0)X_6$	$-X_2$	0	0	$X_5 - \frac{1}{2}f_0(0)X_7$	$-X_6$	0	X_8
X_5	$-\frac{3}{2}f_0(0)X_4 - f_2(0)X_5 + f_0(0)f_2(0)X_7 + \{f_1(0) - \frac{1}{2}\dot{f}_2(0)\}X_8$	$-X_1 - f_2(0)X_3 - \frac{3}{2}f_0(0)X_6 + \{f_1(0) - \frac{1}{2}\dot{f}_2(0) - \frac{1}{2}f^{2}_{2}(0)\}X_7$	$X_5 - f_0(0)X_7 - \frac{1}{2}f_2(0)X_8$	$-X_5 + \frac{1}{2}f_0(0)X_7$	0	$-X_3 + X_4 - \frac{1}{2}f_2(0)X_7$	X_8	0
X_6	$-X_2 + f_2(0)X_6$	0	$-X_6$	X_6	$X_3 - X_4 + \frac{1}{2}f_2(0)X_7$	0	0	X_7
X_7	$-2X_3 - X_4$	$-X_6$	$-X_7$	0	$-X_8$	0	0	0
X_8	$-X_5 + \frac{3}{2}f_0(0)X_7$	$-X_3 - 2X_4 - \frac{1}{2}f_2(0)X_7$	0	$-X_8$	0	$-X_7$	0	0

ent in Lie's formalism, since different sets of initial conditions correspond to different parametrizations of the group and, thus, to a mere change of the basis of the algebra.[24]

Our choice of initial conditions [cf. Eqs. (2.29)], anyhow, is rather simple, since it is attached to the generators η and θ themselves [cf. Eqs. (2.28)]. Indeed, we have preferred this parametrization because: (i) we want to have a uniform formalism in order to be able to compare the symmetry groups of different linear and nonlinear systems; (ii) the initial value definition for the q's is given directly in terms of $\eta(t,x)$ and $\theta(t,x)$, and derivatives thereof, at (0,0), and not in terms of the initial values of the linear basis $\phi_1(t),\ldots,\phi_4(t)$, and their derivatives; (iii) clearly, for a nonlinear system there is no linear basis at all [however, definition (2.28) still works the same]; (iv) finally, the chosen parametrization is precisely the one that brings the infinitesimal operators of the projective group (in two dimensions) in the standard basis adopted in the current literature.[25]

3. SOME MISCELLANEOUS EXAMPLES

With the aim of exhibiting the explicit form adopted by the elements of the infinitesimal symmetry group of an equation of the type of Eq. (2.6), in this section we present four interesting instances, taken from elementary mechanics. For the sake of concreteness we describe this matter in a sketchy manner.

Let us briefly consider the symmetry properties of the following linear systems in one dimension: the free particle, the free falling particle, the simple harmonic oscillator, and the damped harmonic oscillator.

(a) *Free particle*: We have $\ddot{x} = 0$; i.e.,

$$f_0 = f_1 = f_2 = 0\,, \tag{3.1}$$

wherefrom, according to the previous formalism, the well-known generators of the infinitesimal symmetry transformation follow:

$$\begin{aligned} \eta(t,x) &= q^1 + q^3 t + q^7 t^2 + (q^5 + q^8 t)x\,, \\ \theta(t,x) &= q^2 + q^6 t + (q^4 + q^7 t)x + q^8 x^2\,. \end{aligned} \tag{3.2}$$

Hence, the eight infinitesimal operators of the Lie algebra in the chosen q-representation are

$$\begin{aligned} &X_1 = \partial_t\,, \quad X_2 = \partial_x\,, \quad X_3 = t\partial_t\,, \\ &X_4 = x\partial_x\,, \quad X_5 = x\partial_t\,, \quad X_6 = t\partial_x\,, \\ &X_7 = t^2\partial_t + tx\partial_x\,, \quad X_8 = tx\partial_t + x^2\partial_x\,. \end{aligned} \tag{3.3}$$

Clearly, these are the well-known infinitesimal operators of the group of projective transformations in the (t,x) plane,[25] where we recognize the *special* projective tranformation operators of translations (X_1 and X_2), of stretching ($X_3 + X_4$), and of rotation ($X_6 - X_5$), which also pertain in the conformal transformation of the plane.

The q-parametrization, as shown in Eq. (3.2), looks somewhat clumsy; however, this is precisely the parametrization that brings the infinitesimal operators of the projective group (in two dimensions) in the ordered scheme presented in Eq. (3.3), which corresponds with the standard basis of this algebra adopted in the current literature. As we have already said, this is one of the main reasons we have for adopting this particular representation. For the sake of completeness we include herein the well-known Lie algebra of the equation $\ddot{x} = 0$; see Table III.

(b) *Free falling particle*: Now we set $\ddot{x} + g = 0$; i.e.,

$$f_0 = -g\,, \quad f_1 = f_2 = 0\,. \tag{3.4}$$

So we get

$$\eta(t,x) = q^1 + q^3 t + q^7 t^2 + \tfrac{1}{2} g q^8 t^3 + (q^5 + q^8 t)x\,, \tag{3.5}$$

$$\begin{aligned} \theta(t,x) = q^2 + q^6 t + \tfrac{1}{2} g t^2 ((q^4 - 2q^3) - \{q^7 + \tfrac{1}{2} g q^5\}\, t \\ - \tfrac{1}{2} g q^8 t^2) + (q^4 + \{q^7 - \tfrac{3}{2} g q^5\} t)x + q^8 x^2\,, \end{aligned} \tag{3.6}$$

and, thus,

$$\begin{aligned} &X_1 = \partial_t\,, \quad X_2 = \partial_x\,, \quad X_3 = t(\partial_t - gt\partial_x)\,, \\ &X_4 = (x + \tfrac{1}{2} g t^2)\partial_x\,, \\ &X_5 = x(\partial_t - gt\partial_x) - \tfrac{1}{2} g t (x + \tfrac{1}{2} g t^2)\partial_x\,, \\ &X_6 = t\partial_x\,, \quad X_7 = t\,\{t\partial_t + (x - \tfrac{1}{2} g t^2)\partial_x\}\,, \\ &X_8 = (x + \tfrac{1}{2} g t^2)\{t\partial_t + (x - \tfrac{1}{2} g t^2)\partial_x\}\,, \end{aligned} \tag{3.7}$$

wherefrom the Lie algebra easily follows (cf. Table IV).

(c) *Linear harmonic oscillator*: Now we consider the symmetries of the differential equation

$$\ddot{x} + \omega_0^2 x = 0\,; \tag{3.8}$$

TABLE III. The well-known Lie algebra of the projective group in the (t, x) plane.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	0	0	X_1	0	0	X_2	$2X_3 + X_4$	X_5
X_2	0	0	0	X_2	X_1	0	X_6	$X_3 + 2X_4$
X_3	$-X_1$	0	0	0	$-X_5$	X_6	X_7	0
X_4	0	$-X_2$	0	0	X_5	$-X_6$	0	X_8
X_5	0	$-X_1$	X_5	$-X_5$	0	$-X_3 + X_4$	X_8	0
X_6	$-X_2$	0	$-X_6$	X_6	$X_3 - X_4$	0	0	X_7
X_7	$-2X_3 - X_4$	$-X_6$	$-X_7$	0	$-X_8$	0	0	0
X_8	$-X_5$	$-X_3 - 2X_4$	0	$-X_8$	0	$-X_7$	0	0

TABLE IV. The Lie algebra of a free falling particle.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	0	0	$X_1 - 2gX_6$	gX_6	$-\frac{3}{2}gX_4$	X_2	$2X_3 + X_4$	$X_5 + \frac{3}{2}gX_7$
X_2	0	0	0	X_2	$X_1 - \frac{3}{2}gX_6$	0	X_6	$X_3 + 2X_4$
X_3	$-X_1 + 2gX_6$	0	0	0	$-X_5 - gX_7$	X_6	X_7	0
X_4	$-gX_6$	$-X_2$	0	0	$X_5 + \frac{1}{2}gX_7$	$-X_6$	0	X_8
X_5	$\frac{3}{2}gX_4$	$-X_1 + \frac{3}{2}gX_6$	$X_5 + gX_7$	$-X_5 - \frac{1}{2}gX_7$	0	$-X_3 + X_4$	X_8	0
X_6	$-X_2$	0	$-X_6$	X_6	$X_3 - X_4$	0	0	X_7
X_7	$-2X_3 - X_4$	$-X_6$	$-X_7$	0	$-X_8$	0	0	0
X_8	$-X_5 - \frac{3}{2}gX_7$	$-X_3 - 2X_4$	0	$-X_8$	0	$-X_7$	0	0

that is, we set

$$f_1 = \omega_0^2\,, \quad f_0 = f_2 = 0\,. \tag{3.9}$$

Accordingly, we calculate the infinitesimal generators of the symmetry transformations. We obtain

$$\begin{aligned}\eta(t,x) = {} & q^1 + (1/\omega_0)q^3 \sin(\omega_0 t)\cos(\omega_0 t) \\ & + (1/\omega_0)^2 q^7 \sin^2(\omega_0 t) + \{q^5 \cos(\omega_0 t) \\ & + (1/\omega_0) q^8 \sin(\omega_0 t)\}x\,,\end{aligned} \tag{3.10}$$

$$\begin{aligned}\theta(t,x) = {} & q^2 \cos(\omega_0 t) + (1/\omega_0) q^6 \sin(\omega_0 t) + \{q^4 \\ & - q^3 \sin^2(\omega_0 t) + (1/\omega_0) q^7 \sin(\omega_0 t)\cos(\omega_0 t)\}x \\ & + \{q^8 \cos(\omega_0 t) - \omega_0 q^5 \sin(\omega_0 t)\}x^2\,,\end{aligned} \tag{3.11}$$

while the infinitesimal operators come out as follows:

$$\begin{aligned}& X_1 = \partial_t\,, \quad X_2 = (\cos(\omega_0 t))\partial_x\,, \\ & X_3 = (1/\omega_0)\sin(\omega_0 t)\{(\cos(\omega_0 t))\partial_t - (\omega_0 \sin(\omega_0 t))x\partial_x\}\,, \\ & X_4 = x\partial_x\,, \quad X_5 = x\{(\cos(\omega_0 t))\partial_t - (\omega_0 \sin(\omega_0 t))x\partial_x\}, \\ & X_6 = ((1/\omega_0)\sin(\omega_0 t))\partial_x\,, \\ & X_7 = (1/\omega_0)\sin(\omega_0 t)\{((1/\omega_0)\sin(\omega_0 t))\partial_t + (\cos(\omega_0 t))x\partial_x\}\,, \\ & X_8 = x\{((1/\omega_0)\sin(\omega_0 t))\partial_t + (\cos(\omega_0 t))x\partial_x\}\,.\end{aligned} \tag{3.12}$$

The Lie algebra corresponding to the classical linear harmonic oscillator is presented in Table V.

(d) Damped harmonic oscillator: Finally, we also consider the equation of motion

$$\ddot{x} + 2\lambda\dot{x} + \omega_0^2 x = 0\,; \tag{3.13}$$

i.e.,

$$f_0 = 0\,, \quad f_1 = \omega_0^2\,, \quad f_2 = 2\lambda\,. \tag{3.14}$$

Hence, after a straightforward calculation, we get

$$\begin{aligned}\eta(t,x) = {} & q^1 + (1/\omega)q^3 \sin(\omega t)\cos(\omega t) + (1/\omega)^2 q^7 \sin^2(\omega t) \\ & + e^{\lambda t}\{q^5(\cos(\omega t) + (\lambda/\omega)\sin(\omega t)) \\ & + (1/\omega)q^8 \sin(\omega t)\}x\,,\end{aligned} \tag{3.15}$$

$$\begin{aligned}\theta(t,x) = {} & e^{-\lambda t}\{q^2(\cos(\omega t) + (\lambda/\omega)\sin(\omega t)) + (1/\omega)q^6 \sin(\omega t)\} \\ & + \{q^4 - q^3(\sin(\omega t) + (\lambda/\omega)\cos(\omega t))\sin(\omega t) \\ & + (1/\omega)q^7(\cos(\omega t) - (\lambda/\omega)\sin(\omega t))\sin(\omega t)\}x \\ & + e^{\lambda t}\{q^8(\cos(\omega t) - (\lambda/\omega)\sin(\omega t)) \\ & - (\omega_0^2/\omega)q^5 \sin(\omega t)\}x^2\,,\end{aligned} \tag{3.16}$$

TABLE V. The Lie algebra of a classical linear harmonic oscillator.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	0	$-\omega_0^2 X_6$	$X_1 - 2\omega_0^2 X_7$	0	$-\omega_0^2 X_8$	X_2	$2X_3 + X_4$	X_5
X_2	$\omega_0^2 X_6$	0	0	X_2	$X_1 - \omega_0^2 X_7$	0	X_6	$X_3 + 2X_4$
X_3	$-X_1 + 2\omega_0^2 X_7$	0	0	0	$-X_5$	X_6	X_7	0
X_4	0	$-X_2$	0	0	X_5	$-X_6$	0	X_8
X_5	$\omega_0^2 X_8$	$-X_1 + \omega_0^2 X_7$	X_5	$-X_5$	0	$-X_3 + X_4$	X_8	0
X_6	$-X_2$	0	$-X_6$	X_6	$X_3 - X_4$	0	0	X_7
X_7	$-2X_3 - X_4$	$-X_6$	$-X_7$	0	$-X_8$	0	0	0
X_8	$-X_5$	$-X_3 - 2X_4$	0	$-X_8$	0	$-X_7$	0	0

TABLE VI. The Lie algebra of the damped harmonic oscillator.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	0	$-\omega_0^2 X_6$	$X_1-\lambda X_4 - 2\omega^2 X_7$	0	$2\lambda X_5-\omega_0^2 X_8$	$X_2-2\lambda X_6$	$2X_3+X_4$	X_5
X_2	$\omega_0^2 X_6$	0	$-\lambda X_6$	X_2	$X_1+2\lambda X_3 - (\omega^2-\lambda^2)X_7$	0	X_6	$X_3+2X_4 + \lambda X_7$
X_3	$-X_1+\lambda X_4 + 2\omega^2 X_7$	λX_6	0	0	$-X_5+\lambda X_8$	X_6	X_7	0
X_4	0	$-X_2$	0	0	X_5	$-X_6$	0	X_8
X_5	$-2\lambda X_5+\omega_0^2 X_8$	$-X_1-2\lambda X_3 + (\omega^2-\lambda^2)X_7$	$X_5-\lambda X_8$	$-X_5$	0	$-X_3+\lambda X_4 - \lambda X_7$	X_8	0
X_6	$-X_2+2\lambda X_6$	0	$-X_6$	X_6	$X_3-X_4 + \lambda X_7$	0	0	X_7
X_7	$-2X_3-X_4$	$-X_6$	$-X_7$	0	$-X_8$	0	0	0
X_8	$-X_5$	$-X_3-2X_4 - \lambda X_7$	0	$-X_8$	0	$-X_7$	0	0

where we have introduced $\omega = (\omega_0^2 - \lambda^2)^{1/2}$. In this case, for the infinitesimal operators, we obtain

$$
\begin{aligned}
&X_1 = \partial_t\,, \quad X_2 = (1/\omega)e^{-\lambda t}(\omega\cos(\omega t) + \lambda\sin(\omega t))\partial_x\,,\\
&X_3 = (1/\omega)\sin(\omega t)\{(\cos(\omega t))\partial_t - (\omega\sin(\omega t)\\
&\qquad + \lambda\cos(\omega t))x\partial_x\}\,,\\
&X_4 = x\partial_x\,,\\
&X_5 = (1/\omega)e^{\lambda t}x\{(\omega\cos(\omega t) + \lambda\sin(\omega t))\partial_t - (\omega_0^2\sin(\omega t))x\partial_x\}\,, \qquad (3.17)\\
&X_6 = (1/\omega)(e^{-\lambda t}\sin(\omega t))\partial_x\,,\\
&X_7 = (1/\omega^2)\sin(\omega t)\{(\sin(\omega t))\partial_t\\
&\qquad + (\omega\cos(\omega t) - \lambda\sin(\omega t))x\partial_x\}\,,\\
&X_8 = (1/\omega)e^{\lambda t}x\{(\sin(\omega t))\partial_t + (\omega\cos(\omega t) - \lambda\sin(\omega t))x\partial_x\}\,.
\end{aligned}
$$

The Lie algebra is given in Table VI.

Finally, we wish to mention here the obvious fact that the one-dimensional time-independent Schrödinger equation comes quite directly under the scope of the similarity techniques as presented in this paper. However, given its importance, the similarity analysis of the Schrödinger equation deserves a separate treatment.[5]

4. THE INFINITESIMAL ELEMENTS OF THE DYNAMICAL GROUP OF A ONE-DIMENSIONAL LINEAR SYSTEM

Notwithstanding the mechanical interest of the linear examples presented in the previous section, thus far our discussion of the converse problem of similarity analysis, as applied to Eq. (2.6), has been purely mathematical. In order to come closer to the spirit of mechanics, some basic considerations on space and time are important in our approach. So let us recall some features of Newtonian mechanics.

As a typical example, we consider a classical system formed by two particles, which only interact internally, and which are constrained to move on a fixed straight line. Because of the Newtonian assumptions on the nature of space and time, the Galilean transformation is a symmetry of the system. However, once we separate the center of mass and introduce the internal configuration variable, $x = x_2 - x_1$ [whose equation of motion is precisely Eq. (2.1), say], then the requirement of Galilean invariance becomes in the trivial symmetry transformation: $t' = t - \tau$ (or $t' = t$) and $x' = x$, while the center of mass performs the Galilean boost. From a mathematical standpoint, it is clear that the differential equation of motion (2.1) has more symmetries, in general, than the trivial one. Clearly, these new [cf. Eqs. (2.2)] symmetries are not attached to the absolute properties of Euclidean space and Newtonian time. They are internal symmetries which transform allowable internal motions into allowable motions of a classical system.

However, time is a sacrosanct affine parameter in Newtonian mechanics. In effect, in order to avoid artificial "forces," which otherwise would be acting on the center of mass, one has to consider instead of the group alluded to in Eqs. (2.2), the subgroup of internal point transformations

$$
\begin{aligned}
&t' = a(t-\tau)\,,\\
&x' = S(t,x)\,,
\end{aligned}
\qquad (4.1)
$$

(where a and τ are constants), which leave the equation of internal motion [cf. Eq. (2.1)] form invariant, and, hence, which are consistent with the Galileo transformation of the center of mass. In Eq. (4.1) we have included the possibility of a change in time scale as a symmetry of the dynamical system.

The transformations (4.1) form a Lie group, denoted by $G_m(2)$. Following Mariwalla,[26] $G_m(2)$ may be called the *dynamical group* of the system; i.e., the dynamical group reflects the internal symmetries of the system which are consistent with the Newtonian time hypothesis and, thus, with the Galilean symmetry of the center of mass.

In this section, then, we study the infinitesimal point

TABLE VII. The associated Lie algebra of the dynamical group of a one-dimensional linear Newtonian system.

	X_1	X_2	X_3	X_4	X_6
X_1	0	$\frac{1}{2}\dot f_2(0)X_6 - f_1(0)X_6$	$X_1 - \frac{1}{2}f_2(0)X_4 + 2f_0(0)X_6$	$-f_0(0)X_6$	$X_2 - f_2(0)X_6$
X_2	$f_1(0)X_6 - \frac{1}{2}\dot f_2(0)X_6$	0	$-\frac{1}{2}f_2(0)X_6$	X_2	0
X_3	$-X_1 + \frac{1}{2}f_2(0)X_4 - 2f_0(0)X_6$	$\frac{1}{2}f_2(0)X_6$	0	0	X_6
X_4	$f_0(0)X_6$	$-X_2$	0	0	$-X_6$
X_6	$-X_2 + f_2(0)X_6$	0	$-X_6$	X_6	0

transformations belonging in $G_m(2)$, for a one-dimensional linear system. In this case, according to Eqs. (4.1), it is clear that for a linear system Eqs. (2.11) and (2.12) become

$$\eta(t) = \phi_2(t), \tag{4.2}$$

$$\theta(t,x) = \phi_3(t)x + \phi_4(t), \tag{4.3}$$

since now we have

$$\phi_1(t) = 0, \quad \phi_2(t) = q^1 + q^3 t, \tag{4.4}$$

while ϕ_3 and ϕ_4 still have to satisfy Eqs. (2.13). Moreover, these equations become now in the identity

$$(q^1 + q^3 t)\dot\Omega(t) + 2q^3\Omega(t) = 0, \tag{4.5}$$

which must hold for all t, and where we have written

$$\Omega(t) = f_1(t) - \tfrac{1}{4}f_2^2(t) - \tfrac{1}{2}\dot f_2(t), \tag{4.6}$$

plus two linear differential equations for ϕ_3 and ϕ_4; namely,

$$\begin{aligned} 2\dot\phi_3 &= -(d/dt)\{(q^1 + q^3 t)f_2(t)\}, \\ \ddot\phi_4 + f_2\dot\phi_4 + f_1\phi_4 &= 2q^3 f_0 + (q^1 + q^3 t)\dot f_0 - f_0\phi_3. \end{aligned} \tag{4.7}$$

The first equation in (4.7) gives

$$\phi_3(t) = q^4 - \tfrac{1}{2}(q^1 + q^3 t)f_2(t), \tag{4.8}$$

and therefore the second equation (4.7) becomes

$$\ddot\phi_4 + f_2\dot\phi_4 + f_1\phi_4 = \tfrac{1}{2}(q^1 + q^3 t)(2\dot f_0 - f_0 f_2) - (q^4 - 2q^3)f_0. \tag{4.9}$$

The identity (4.5) is an additional condition which must be satisfied by the given functions $f_1(t)$, $f_2(t)$, and $\dot f_2(t)$. This brings into the fore an interesting classification of the problem. We distinguish the following cases.

Case 1: When $\Omega(t) = 0$, then q^1 and q^3 are two arbitrary independent parameters.

Case 2: When $\Omega(t) = \text{const} \neq 0$, then necessarily we have $q^3 = 0$. This means that there is no change in time scale allowed as a symmetry transformation.

Case 3: When $\Omega(t) = at^{-2}$, $a \neq 0$, then $q^1 = 0$, which means that the change in time scale may be a symmetry while the time translation is not.

Case 4: If $\Omega(t) = a(1 + bt^2)^{-1}$, with $ab \neq 0$. This requires $q^3 = bq^1$; i.e., $\eta = q^1(1 + bt)$, which means that, actually, there is no symmetry of time translation.

Case 5: Here we consider all the remaining possibilities. Clearly, this means $q^1 = q^3 = 0$, and therefore $t' = t$.

Furthermore, if one considers the general procedure we have to follow in Sec. 2 to obtain the Lie algebra, one immediately observes the well-known rule:

$$q^a = 0 \Longrightarrow X_a = 0, \tag{4.10}$$

for having the subalgebras associated with the subgroups one obtains by the elimination of some of the parameters. (The meaning of this rule is obvious.) Hence, a glance at Eqs. (2.28) shows that, because of Eqs. (4.2), (4.3), and (4.4), in the present case we have, quite generally, $q^5 = q^7 = q^8 = 0$; i.e.,

$$X_5 = X_7 = X_8 = 0. \tag{4.11}$$

In other words, the elimination of these three operators of the general Lie algebra, presented in Table II, immediately gives us the most general Lie algebra associated with the dynamical group of a one-dimensional linear system in the context of Newtonian mechanics. We present this algebra in Table VII. This algebra corresponds to case 1 above (i.e., $\Omega = 0$). As examples of this case, we mention the free particle and the free falling particle, whose dynamical group algebras can be read easily from Table VII. The damped harmonic oscillator and the simple harmonic oscillator pertain in case 2 above (since $\Omega = \omega$); i.e., we have

$$X_3 = X_5 = X_7 = X_8 = 0, \tag{4.12}$$

wherefrom the corresponding algebra follows immediately (see Table VIII).

TABLE VIII. The Lie algebra of the dynamical group of the one-dimensional damped harmonic oscillator (or, when $\lambda = 0$, the simple harmonic oscillator).

	X_1	X_2	X_4	X_6
X_1	0	$-\omega_0^2 X_6$	0	$X_2 - 2\lambda X_6$
X_2	$\omega_0^2 X_6$	0	X_2	0
X_4	0	$-X_2$	0	$-X_6$
X_6	$-X_2 + 2\lambda X_6$	0	X_6	0

[1]K. Mariwalla, Phys. Rep. C **20**, 287–362 (1975).
[2]N. Mukunda, J. Math. Phys. **8**, 1069 (1967), and references therein; cf., also, H. V. McIntosh, "Symmetry and Degeneracy," in *Group Theory and Its Applications, Vol. II*, edited by E. M. Loebl (Academic, New York, 1971).
[3]K. Mariwalla, "Coordinate transformations that form groups in the large," in *Lectures in Theoretical Physics, Vol. 13*, edited by A. O. Barut and W. E. Brittin (Colorado Associated Press, Boulder, 1970).
[4]Cf. R. T. Prosser, J. Math. Phys. **24**, 548 (1983), and references quoted therein.
[5]This is a report of work in progress. The identification and classification of the Lie algebras obtained in this paper, their corresponding global symmetry groups, and other related issues will be published elsewhere.
[6]See, for instance, G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer-Verlag, New York, 1974).
[7]E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).
[8]This point has been strongly emphasized by Mariwalla; cf. Ref. 1.
[9]E. L. Hill, Rev. Mod. Phys. **23**, 253 (1951).
[10]Needless to say, for those special systems which do not admit a Lagrangian [cf. J. Douglas, Trans. Am. Math. Soc. **50**, 71 (1941)], recourse to the equations of motion is the only approach one has for studying their symmetries.
[11]M. Bedner, Ann. Phys. **75**, 305 (1973); C. Fronsdal, Phys. Rev. **156**, 1665 (1967).
[12]Cf., for instance, R. M. F. Houtappel, H. Van Dam, and E. P. Wigner, Rev. Mod. Phys. **37**, 595 (1965).
[13]Cf. G. H. Katzin and J. Levine, J. Math. Phys. **9**, 8 (1968); G. S. Hall, Lett. Nuovo Cimento **6**, 46 (1973); G. H. Katzin, J. Math. Phys. **14**, 1213 (1973); K. Mariwalla, Ref. 1.
[14]K. H. Mariwalla, Lett. Nuovo Cimento **12**, 253 (1975).
[15]K. H. Mariwalla, Ref. 3.
[16]S. Lie, *Vorlesungen über Differentialgleichungen mit bekannten infinitesimalen Transformationen* (B. G. Teubner, Leipzig, 1891; reprinted by Chelsea, New York, 1967).
[17]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), Chap. IV.
[18]Cf., for instance, L. Bianchi, *Lezioni sulla Teoria dei Gruppi Continui Finiti di Transformazioni* (Zanichelli, Bologna, 1928), Chap. X.
[19]L. P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961).
[20]J. E. Campbell, *Introductory Treatise on Lie's Theory of Finite Continuous Transformation Groups* (Chelsea, New York, 1966), Chap. II.
[21]F. Gonzalez-Gascon, J. Math. Phys. **18**, 1763 (1977).
[22]R. L. Anderson, S. Kumei, and C. E. Wulfman, Phys. Rev. Lett. **28**, 988 (1972); R. L. Anderson and S. M. Davison, J. Math. Anal. Appl. **48**, 301 (1974).
[23]See, for instance, J. E. Campbell, Ref. 20, p. 23.
[24]G. Racah, Ergeb. Exact. Naturwiss. **37**, 28 (1965).
[25]L. P. Eisenhart, Ref. 19, p. 62; cf., also, G. W. Bluman and J. D. Cole, Ref. 6, p. 51.
[26]K. H. Mariwalla, Ref. 1.

General indices of representations and Casimir invariants

S. Okubo
Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

J. Patera
Centre de Recherche de Mathématiques Appliquées, Université de Montreál, Montréal, Québec, H3C 3J7, Canada

(Received 26 October 1982; accepted for publication 16 September 1983)

A modified definition of the index of degree p of a finite-dimensional representation of a simple Lie algebra is given. The definition applies equally to even and odd p. The correspondence between the earlier definitions of indices (even p) and anomaly numbers (odd p) is pointed out as well as the relation to Casimir invariants of the algebra. A closed formula for the fifth-order index is derived.

PACS numbers: 02.20.Sv

The index of order k of a finite dimensional representation ρ of a simple Lie algebra L with complex or real parameters is a rational number (a suitable normalization makes it an integer) associated with ρ. There are three ways of defining it in the literature: as a certain k th order derivative of the character $\chi(\rho)$ of the representation ρ of the Lie group evaluated at the origin, or in terms of scalar products of weights of ρ, or as the trace of a Casimir operator of degree k. As long as one is interested in indices of degree 2 only, there is little difference or relative advantage in using any of the definitions. However, they are quite different for indices of higher order.

The practical value of the indices is in relations they provide when tensor products or tensor powers (with or without a permutational symmetry) are decomposed into direct sums of irreducible components, or when representations of a Lie group/algebra are reduced to direct sums of representations of subgroups/subalgebras. For exploitation of these properties see Refs. 1–6.

The purpose of this paper is to provide a new definition of the index of the representation ρ of L as the trace of a *suitably chosen* set of Casimir operators of L. The main value of the new definition is that it provides more relations between the "new" indices than there are between the "old" ones, and the structure of these relations is more uniform as to the degree of the index and the type of the Lie algebra. Consequently, the applications are more restrictive, i.e., more powerful. For the first time we derive here an explicit expression for the fifth degree index of the representations of the Lie algebras of type A_n ($n\geqslant 5$) and explain the relation between different definitions of the indices.

1. SOME PROPERTIES OF THE "OLD" INDICES

Here we recall some of the properties of the usual indices which will be needed later on in the paper.

Let L be a simple Lie algebra, and let ρ be a finite representation of L. The (second-order) index $l_2(\rho)$ of the representation ρ, as defined by Dynkin,[1] is the trace of the second-order Casimir operator on the representation ρ. Its value when ρ is irreducible is given by

$$l_2(\rho) = d(\rho)(\Lambda,\Lambda + \delta), \tag{1.1}$$

where $d(\rho)$ is the dimension of ρ, $(\Lambda,\Lambda + \delta)$ is the eigenvalue of the Casimir operator, Λ and δ being respectively the highest weight of ρ and the sum of all positive roots of L. The Dynkin index has many useful properties. Let ρ_A and ρ_B be any two irreducible representations of L. Then the tensor product $\rho_A \otimes \rho_B$ decomposes as a direct sum of N irreducible components ρ_j:

$$\rho_A \otimes \rho_B = \bigoplus_{j=1}^{N} \rho_j. \tag{1.2}$$

The Dynkin index $l_2(\rho)$ satisfies the sum rule

$$d(\rho_A)l_2(\rho_B) + d(\rho_B)l_2(\rho_A) = \sum_{j=1}^{N} l_2(\rho_j). \tag{1.3}$$

As early as in 1965, Biedenharn[2] briefly considered the third-order index $l_3^v(\rho)$ [see Eq. (1.6)] for some v especially in connection with the SU(3) group, and noted its proportionality to the third-order Casimir invariant of the SU(3). A generalization of the Dynkin index was undertaken in Ref. 3. It is based on the observation that $l_2(\rho)$ can also be written as $l_2(\rho) = \Sigma_M(M, M)$, where the summation extends over all weights M of ρ and (M, M) is the standard scalar product in the root space of L. The index $l_{2k}(\rho)$ of order $2k$ of the representation ρ was then defined[3] as

$$l_{2k}(\rho) = \sum_M (M, M)^k, \quad k = 0,1,2,\cdots. \tag{1.4}$$

It was shown[3,4] that $l_{2k}(\rho)$ possesses some of the useful properties of $l_0(\rho) = d(\rho)$ (dimension of ρ) and $l_2(\rho)$ (Dynkin type) for some values of $k > 2$ and/or for Lie algebras of several types. One of the general properties of $l_4(\rho)$ which refers to the product (1.2) is an analog of (1.3):

$$l_4(\rho_A \otimes \rho_B) = d(\rho_A)l_4(\rho_B) + d(\rho_B)l_4(\rho_A) + \frac{2(n+2)}{n} l_2(\rho_A)l_2(\rho_B) = \sum_{j=1}^{N} l_4(\rho_j). \tag{1.5}$$

Here n denotes the rank of L. Extensive tables of $d(\rho)$, $l_2(\rho)$, and $l_4(\rho)$ are found in Ref. 5. Sum rules similar to (1.3) and

(1.5) for l_{2k}, $k>2$, hold only for Lie algebras of certain types.[3] Equation (1.4) cannot be used to define odd-order indices. It turns out,[6] however, that the quantity

$$l'_p(\rho)=\sum_M (v,M)^p, \quad p=0,1,2,\cdots, \tag{1.6}$$

where v is an arbitrary but fixed nonzero vector, has a behavior similar to that of an index for even and odd values of p. In particular, one has for the tensor product (1.2)

$$l'_p(\rho_A\otimes\rho_B)=\sum_{k=0}^{p}\frac{p!}{k!(p-k)!}l'_k(\rho_A)l'_{p-k}(\rho_B)$$
$$=\sum_{j=1}^{N} l'_p(\rho_j). \tag{1.7}$$

Moreover, a suitable choice of the fixed vector v makes $l'_3(\rho)$ equal[6] to the triangle anomaly number[7,8] of ρ. This fact has been implicitly noted also in Ref. 2 for the case of SU(3) in a different context.

Explicit algebraic expressions are computed for $l_2(\rho)$ and $l_4(\rho)$ in Ref. 3, for $l'_3(\rho)$ in Ref. 6, and the relation between $l'_p(\rho)$ and $l_p(\rho)$ for $p=2$ and 4 in Ref. 6 allows us to find also $l'_2(\rho)$ and $l'_4(\rho)$ for any ρ. For other degrees one has to use directly the definitions (1.4) and (1.6). Even then the evaluation of l_p or l'_p does not represent a serious problem. Indeed, the summation over all weights can be replaced by the summation over the Weyl group orbits of weights provided one knows the orbits (i.e., dominant weights of ρ) and their multiplicities. The multiplicities can either be found from the new extensive tables[9] or they can be computed using, for instance, the fast algorithm of Ref. 10.

In Ref. 11, a new definition of the fourth-order index has been proposed. The present work is its generalization.

2. CHOICE OF CASIMIR INVARIANTS AND DEFINITION OF THE 'NEW' INDEX

The new definition of the index of representation of general order p offered here [Eq. (2.10) below] eliminates the drawbacks of the indices l_{2p} and l'_p; namely, dependence of their properties (or even the existence of these properties) on the type of the Lie algebra L and the dependence on the choice of the vector v. It relates naturally the indices to the Casimir invariants, but hinges on a particular choice[11] of the basis of the Casimir invariants.

It is known[12] that any simple Lie algebra L of rank n has exactly n fundamental Casimir invariants $\{I_p\}$. For example, for the Lie algebras A_n, they are I_p, $p=2,3,\ldots,n+1$, where I_p denotes the fundamental Casimir invariant of degree p. Their degrees are recalled in Table I. Thus any other Casimir invariant of L is a polynomial of the n fundamental ones. We notice that the choice of the n fundamental Casimir invariants (i.e., the basis) is by no means unique. For example, we may replace the fourth-degree invariant I_4 by

$$I'_4=I_4+cI_2^2+c'I_2 \tag{2.1}$$

for arbitrary constants c and c'. However, we eliminate most of the arbitrariness of the choice in the following way. Let t_1, $t_2,\ldots,t_d$ be a basis of the Lie algebra L with the Lie multiplication table

TABLE I. Degrees of the fundamental Casimir invariants of simple Lie algebras.

Lie algebra	Degrees
A_{l-1} $(l\geqslant 2)$	2, 3, 4,..., l
B_l, C_l $(l\geqslant 2)$	2, 4, 6,..., $2l$
D_l $(l\geqslant 3)$	2, 4, 6,..., $2l-2$, l
E_6	2, 5, 6, 8, 9, 12
E_7	2, 6, 8, 10, 12, 14, 18
F_4	2, 6, 8, 12
G_2	2, 6
E_8	2, 8, 12, 14, 18, 20, 24, 30

$$[t_\mu,t_\nu]=c^\lambda_{\mu\nu}t_\lambda \quad (\mu,\nu,\lambda=1,2,\ldots,d) \tag{2.2}$$

with the structure constants $c^\lambda_{\mu\nu}$, where a summation over repeated Greek indices is understood. Further, let A_p and B_q be two Casimir invariants of L of order p and q, respectively, which, however, are not necessarily fundamental. In terms of the basis of the Lie algebra, they are written as

$$A_p=a^{\mu_1\mu_2\cdots\mu_p}t_{\mu_1}t_{\mu_2}\cdots t_{\mu_p},$$
$$B_q=b^{\mu_1\mu_2\cdots\mu_q}t_{\mu_1}t_{\mu_2}\cdots t_{\mu_q}, \tag{2.3}$$

where the coefficients $a^{\mu_1\cdots\mu_p}$ and $b^{\mu_1\cdots\mu_q}$ are symmetric with respect to permutations of superscripts. Then we define an inner product (A_p,B_q) as

$$(A_p,B_q)=\delta_{pq}a^{\mu_1\cdots\mu_p}b_{\mu_1\cdots\mu_q}, \tag{2.4}$$

where we have lowered the Greek indices in $b^{\mu_1\cdots\mu_p}$ using the Killing metric tensor $g_{\mu\nu}$ defined by

$$g_{\mu\nu}=c\cdot\mathrm{tr}(\mathrm{ad}\,t_\mu\ \mathrm{ad}\,t_\nu),\quad c\neq 0. \tag{2.5}$$

It is easy to verify that the inner product (A_p,B_q) is independent of a particular choice of the basis t_i, $i=1,\ldots,d$, of L. By the linearity we can extend the definition of the inner product to any two Casimir invariants.

Apart from normalization constants, a unique choice of the fundamental Casimir invariants of L then can be made. Let us illustrate it on the example of A_n. We choose

$$J_2=I_2,\quad J_3=I_3 \tag{2.6}$$

since I_2 and I_3 are unique Casimir invariance of the second and third order, respectively, for A_n. However the fourth-order Casimir invariant has the ambiguity expressed in Eq. (2.1). We choose[11] constants c and c' so that J_4 is a fourth-order Casimir invariant satisfying the orthogonality condition

$$(J_4,(I_2)^2)=0,\quad (J_4,J_2)=0. \tag{2.7}$$

Similarly, constants c' and c'' in expression $J_5=I_5+c'I_2I_3+c''I_3$ are determined by (see Sec. 8 for explicit form of J_5)

$$(J_5,I_2I_3)=0 \tag{2.8}$$

as well as by the requirement that J_5 is of purely fifth order.

As for the sixth-order invariant J_6, we demand the validity of

$$(J_6,(I_2)^3) = (J_6,(I_3)^2) = (J_6,I_2I_4) = 0, \tag{2.9}$$

since $(I_2)^3$, $(I_3)^2$, and I_2I_4 are also Casimir invariants of the sixth order. We can continue this process to determine higher order invariants J_p's at least for compact simple Lie algebras. We note that some Lie algebras may lack a fundamental Casimir invariant I_p for some integer p. Then, we have, in general, $J_p = 0$ identically for any such p. For example, all exceptional Lie algebras G_2, F_4, E_6, E_7, and E_8 as well as A_1 and A_2 do not possess[12] any genuine fundamental fourth-order invariants so that $J_4 = 0$ for these algebras.[11] As we noted elsewhere,[13] this fact implies the validity of a quartic trace identity for these Lie algebras. Also, for a given p, J_p is, in general, unique, apart from an overall normalization constant, except for the Lie algebras D_n for n an even integer, where we have two linearly independent nth-order fundamental Casimir invariants J_n and $\hat{J}_n$ satisfying the orthogonality condition $(J_n, \hat{J}_n) = 0$ as in Ref. 11 for the special case $n = 4$.

We define[11] the fundamental pth-order index $D^{(p)}(\rho)$ for a representation ρ to be the trace of J_p in the representation ρ. When ρ is irreducible, the eigenvalue of J_p in ρ is designated as $J_p(\rho)$. Then we have

$$D^{(p)}(\rho) = d(\rho)J_p(\rho) = \operatorname{tr} J_p. \tag{2.10}$$

Hereafter, ρ designates the generic irreducible representation of L, unless otherwise stated so. We shall prove in the next section that $D^{(p)}(\rho)$ satisfies a sum rule

$$d(\rho_A)D^{(p)}(\rho_B) + d(\rho_B)D^{(p)}(\rho_A) = \sum_{j=1}^{N} D^{(p)}(\rho_j) \tag{2.11}$$

for the Clebsch–Gordan decomposition equation (1.2). Then, $l_2(\rho)$ must be proportional to $D^{(2)}(\rho)$. The validity of Eq. (2.11) for $p = 3$ and 4 has already been noted in Ref. 11. We shall also show in Sec. 3 that the indices $l_p(\rho)$ and $l'_p(\rho)$ defined earlier must be polynomials in those fundamental indices $D^{(q)}(\rho)$'s with $q \leqslant p$. This is the reason we called $D^{(p)}(\rho)$'s fundamental indices. We will prove also the $D^{(p)}(\rho)$'s satisfy some polynomial sum rules for the decomposition equation (1.2) in addition to Eq. (2.11).

Finally, let us consider a branching sum rule. Let L_0 be a semisimple subalgebra of L. Any irreducible representation ω of L, restricted to the subalgebra L_0 will then be, in general, reducible, and will be decomposed as a direct sum

$$\omega \supset \rho = \bigoplus_j \rho_j \tag{2.12}$$

of irreducible components ρ_j's of L_0. Then, for the decomposition equation (2.12), we have

$$\xi_p D^{(p)}(\omega) = \sum_j D_0^{(p)}(\rho_j) = D_0^{(p)}(\rho), \tag{2.13}$$

where ξ_p is a constant which depends upon L and L_0 but not upon ω. Therefore, once ξ_p is computed for a specific representation ω, Eq. (2.13) can be used as a check of the branching rule (2.12) for any other irreducible representation ω. The validity of (2.13) was established for $p = 2$ (Ref. 1), $p = 3$ (Ref. 6), and, for some Lie algebras, also for $p = 4$ (Ref. 3). With the new definition of the indices its validity is further extended in Sec. 6. The numerical expression of $J_p(\rho)$ ($p \leqslant 5$) for the Lie algebra A_n is given in Sec. 8 with some applications.

3. GENERAL INDICES

In this section we introduce a general index $L_p(\rho)$ of the representation ρ of degree p. The old and new indices of previous sections turn out to be different specializations of $L_p(\rho)$.

Let t_μ ($\mu = 1,2,\cdots$) be an ordered basis of the Lie algebra L with the multiplication table (2.2). We denote the matrices representing t_μ in ρ by X_μ. The general index is then given by

$$L_p(\rho) = b^{\mu_1\cdots\mu_p}\operatorname{Tr}(X_{\mu_1}\cdots X_{\mu_p}), \tag{3.1}$$

where $b^{\mu_1\cdots\mu_p}$ are real coefficients, completely symmetric with respect to permutation of the superscripts, and not all equal to zero.

When we note $\operatorname{Tr} X_\mu = 0$ for any semisimple Lie algebra, we can readily see that $L_p(\rho)$ for $p = 2$ or 3 satisfy

$$d(\rho_A)L_p(\rho_B) + d(\rho_B)L_p(\rho_A) = \sum_{j=1}^{N} L_p(\rho_j) \quad (p = 2,3) \tag{3.2}$$

for the decomposition equation (1.2). However, for $p \geqslant 4$, the situation is more involved. For example, we have

$$\begin{aligned} &d(\rho_A)L_4(\rho_B) + d(\rho_B)L_4(\rho_A) \\ &\quad + 6b^{\mu\nu\alpha\beta}\operatorname{Tr}^{(A)}(X_\mu X_\nu)\operatorname{Tr}^{(B)}(X_\alpha X_\beta) \\ &= \sum_{j=1}^{N} L_4(\rho_j), \end{aligned} \tag{3.3}$$

where $\operatorname{Tr}^{(A)}$ and $\operatorname{Tr}^{(B)}$ imply the trace operation in the irreducible representation spaces ρ_A and ρ_B, respectively. As we shall see shortly, we have

$$\operatorname{Tr}(X_\mu X_\nu) = c g_{\mu\nu} L_2(\rho) \tag{3.4}$$

for some constant c, which does not depend upon ρ, provided that $L_2(\rho)$ is not identically zero. Therefore, Eq. (3.3) is rewritten as a mixed sum rule

$$\begin{aligned} &d(\rho_A)L_4(\rho_B) \\ &\quad + d(\rho_B)L_4(\rho_A) + c'L_2(\rho_A)L_2(\rho_B) = \sum_{j=1}^{N} L_4(\rho_j), \end{aligned} \tag{3.5}$$

where c' is a constant depending only on the values of $b^{\mu_1\cdots\mu_p}$.

Relations (1.5) and (1.7) for $p = 4$ are clearly special cases of (3.5). It is useful to see explicitly the choice of coefficients in (3.1) which leads to the indices $l_p(\rho)$ and $l'_p(\rho)$ of (1.4) and (1.6), respectively. For that let us first fix the Cartan–Weyl basis for L. Namely, we choose

$$t_\mu = \{h_j, e_\alpha, e_{-\alpha}\}, \tag{3.6}$$

where h_j, $j = 1,2,\ldots,n$, span the Cartan subalgebra of L. Then, choosing

$$b^{\mu_1\cdots\mu_p} = v^{\mu_1}v^{\mu_2}\cdots v^{\mu_p}, \tag{3.7}$$

where

$$v^\mu t_\mu = \sum_{j=1}^{n} v^j h_j,$$

we have

$$L_p(\rho) = l'_p(\rho)$$

of (1.6). Next, let us choose

$$b^{\mu_1\nu_1\mu_2\nu_2\cdots\mu_p\nu_p} = \frac{1}{p!}\sum_P \tilde{g}^{\mu_1\nu_1}\tilde{g}^{\mu_2\nu_2}\cdots\tilde{g}^{\mu_p\nu_p}, \tag{3.8}$$

where the summation is over $p!$ permutations P over $\nu_1,\dots,\nu_p$ and where $\tilde{g}^{\mu\nu}$ has nonzero components only in Cartan subalgebra sector by

$$\begin{aligned} &\tilde{g}^{jk} = g^{jk} \quad (j,k = 1,2,\dots,n), \\ &\tilde{g}^{j\alpha} = \tilde{g}^{\alpha\beta} = 0 \quad (\alpha \text{ and } \beta \text{ are nonzero roots}). \end{aligned} \tag{3.9}$$

It is easy to observe that $L_{2p}(\rho)$ in this case is precisely $l_{2p}(\rho)$ defined by Eq. (1.4).

Since $b^{\mu_1\cdots\mu_p}$ is completely arbitrary except for its totally symmetric property on index set μ_j, we call any $L_p(\rho)$ the generalized index of order p. At first glance, it may appear that we have an infinite number of pth-order indices $L_p(\rho)$ for any given p. However, this is not really the case. We shall prove shortly that the number of linearly independent pth-order generalized indices is equal to the number of linearly independent pth-order Casimir invariants of L. Let A_p be a pth-order Casimir invariant such that

$$A_p = a^{\mu_1\mu_2\cdots\mu_p} t_{\mu_1} t_{\mu_2}\cdots t_{\mu_p} \tag{3.10}$$

for some completely symmetric coefficients $a^{\mu_1\cdots\mu_p}$. The condition

$$[t_\lambda, A_p] = 0 \tag{3.11}$$

can easily be seen to be equivalent to

$$\sum_{j=1}^{p} c^\alpha_{\lambda\mu_j} a_{\mu_1\cdots\hat{\alpha}_j\cdots\mu_p} = 0, \tag{3.12}$$

where we lowered indices using $g_{\mu\nu}$ and the symbol $\hat{\alpha}_j$ implies that we delete the jth index μ_j and replace it by α inside $a_{\mu_1\cdots\mu_p}$.

Now, we construct fundamental Casimir invariants J_p's as in the previous section. They can be written as

$$J_p = g^{\mu_1\cdots\mu_p} t_{\mu_1} t_{\mu_2}\cdots t_{\mu_p}. \tag{3.13}$$

In particular the orthogonality conditions, Eqs. (2.7)–(2.9), are now rewritten as

$$g^{\mu\nu\alpha\beta} g_{\mu\nu} g_{\alpha\beta} = 0 \quad (p = 4), \tag{3.14a}$$

$$g^{\mu\nu\lambda\alpha\beta} g_{\mu\nu\lambda} g_{\alpha\beta} = 0 \quad (p = 5), \tag{3.14b}$$

as well as

$$\begin{aligned} g^{\mu\nu\lambda\alpha\beta\gamma} g_{\mu\nu\lambda\alpha} g_{\beta\gamma} &= g^{\mu\nu\lambda\alpha\beta\gamma} g_{\mu\nu\lambda} g_{\alpha\beta\gamma} \\ &= g^{\mu\nu\lambda\alpha\beta\gamma} g_{\mu\nu} g_{\lambda\alpha} g_{\beta\gamma} = 0 \quad (p = 6), \end{aligned} \tag{3.14c}$$

for $p = 4,5$ and 6. The fundamental pth-order index $D^{(p)}(\rho)$ is given by

$$D^{(p)}(\rho) = g^{\mu_1\cdots\mu_p} \mathrm{Tr}\,(X_{\mu_1} X_{\mu_2}\cdots X_{\mu_p}), \tag{3.15}$$

which is equal to $d(\rho) J_p(\rho)$ when ρ is irreducible.

4. INDEX OF THE TENSOR PRODUCT OF REPRESENTATIONS

Now, we are in a position to prove the sum rule, Eq. (2.11). We first replace X_μ in Eq. (3.15) by

$$X_\mu \to X_\mu^{(A)} \otimes E_B + E_A \otimes X_\mu^{(B)}, \tag{4.1}$$

which is the generator of the product representation $\rho_A \otimes \rho_B$. Here E_A and E_B are the unit matrices in ρ_A and ρ_B, respectively. Then, the right side of Eq. (3.15) contains many terms such as

$$g^{\mu_1\mu_2\mu_3\cdots\mu_p} \mathrm{Tr}^{(A)}(X_{\mu_1} X_{\mu_2})\, \mathrm{Tr}^{(B)}(X_{\mu_3}\cdots X_{\mu_p}) \tag{4.2}$$

in addition to $d(\rho_A) D^{(p)}(\rho_B) + d(\rho_B) D^{(p)}(\rho_A)$. However, any such term as Eq. (4.2) must be zero for the following reason. We have evidently

$$\mathrm{Tr}(X_\mu X_\nu) = [1/d(\rho_0)] D^{(2)}(\rho) g_{\mu\nu}, \tag{4.3}$$

where ρ_0 hereafter designates the adjoint representation of L. Next, for simplicity, we set $q = p - 2$, and replace indices $\mu_3,\dots,\mu_p$ by $\nu_1,\nu_2,\dots,\nu_q$. Consider now

$$a_{\nu_1\cdots\nu_q} = \frac{1}{q!}\sum_P \mathrm{Tr}^{(B)}(X_{\nu_1}\cdots X_{\nu_q})$$

and note that $\mathrm{Tr}^{(B)}([X_\lambda, X_{\nu_1}\cdots X_{\nu_q}]) = 0$, where the summation is over $q!$ permutations P of q indices $\nu_1,\dots,\nu_q$. Then we find that $a_{\nu_1\cdots\nu_q}$ satisfies Eq. (3.12) with replacement of p by q and of μ_j by ν_j. Therefore, $A_q = a^{\nu_1\cdots\nu_q} t_{\nu_1} t_{\nu_2}\cdots t_{\nu_q}$ is a qth-order Casimir invariant of L and must hence be expressed as a polynomial of the fundamental Casimir invariants J_p's. In view of the Poincaré–Birkhoff–Witt theorem this implies that $a_{\nu_1\cdots\nu_q}$ can be expressed as a linear combination of $g_{\nu_1\cdots\nu_q}$, $g_{\nu_1\nu_2} g_{\nu_3\nu_4} g_{\nu_3\nu_4\cdots\nu_q}$, $g_{\nu_1\nu_2\nu_3} g_{\nu_4} g_{\nu_4\cdots\nu_q}$, etc. Then Eq. (4.2) is zero by the orthogonality conditions such as

$$g^{\mu_1\mu_2\mu_3\cdots\mu_p} g_{\mu_1\mu_2} g_{\mu_3\cdots\mu_p} = 0.$$

This proves the validity of the sum rule (2.11).

A natural question which arises in connection with (2.11) is to find expressions for indices L_p of tensor powers $\rho^n = \rho \otimes \cdots \otimes \rho$ of one representation ρ or, preferably, for components of ρ^n with a given permutation symmetry. Such relations for $l_p(\rho)$ and $l'_p(\rho)$ were derived in Refs. 4, 6, and 14. A separate paper is devoted to this problem.[14]

5. INDICES AND CASIMIR INVARIANTS

It is clear from the definition (3.15) of $D^{(p)}(\rho)$ that there is a close relation between the indices and Casimir invariants, and it follows from the general argument [Eqs. (3.10)–(3.12)] that the number of linearly independent indices of a fixed degree coincides with the number of Casimir invariants. Nevertheless, it is useful to work out the lowest nontrivial case, i.e., $p = 4$.

By the same reasoning as above, we can write

$$\frac{1}{4!}\sum_P \mathrm{Tr}(X_\mu X_\nu X_\alpha X_\beta)$$
$$= c_1 g_{\mu\nu\alpha\beta} + \tfrac{1}{3} c_2 (g_{\mu\nu} g_{\alpha\beta} + g_{\mu\alpha} g_{\nu\beta} + g_{\mu\beta} g_{\nu\alpha}) \tag{5.1}$$

for some constants c_1 and c_2. Multiplying both sides of Eq. (5.1) by $g^{\mu\nu\alpha\beta}$ and $g^{\mu\nu}g^{\alpha\beta}$ and noting the orthogonalization condition (3.14a), we find

$$c_1 g^{\mu\nu\alpha\beta} g_{\mu\nu\alpha\beta} = D^{(4)}(\rho) = d(\rho) J_4(\rho) \tag{5.2}$$

and

$$c_2[d(\rho_0)+2]d(\rho_0) = \overline{D}^{(4)}(\rho), \tag{5.3a}$$

$$\overline{D}^{(4)}(\rho) = d(\rho) J_2(\rho)\{J_2(\rho) - \tfrac{1}{6} J_2(\rho_0)\}, \tag{5.3b}$$

where ρ_0 is the adjoint representation. Now $g^{\mu\nu\alpha\beta}g_{\mu\nu\alpha\beta}$ cannot be zero, unless $D^{(4)}(\rho)$ is identically zero for all irreducible representations ρ, and hence unless $J_4 = 0$ identically by Harish-Chandra's theorem.[12] Therefore, we can express c_1 and c_2 in terms of $D^{(4)}(\rho)$ and $\overline{D}^{(4)}(\rho)$, respectively. Inserting these expressions into Eq. (3.1) leads to

$$L_4(\rho) = c_1' D^{(4)}(\rho) + c_2' \overline{D}^{(4)}(\rho), \tag{5.4}$$

where constants c_1' and c_2' are now independent of the generic irreducible representation ρ. We note that $\overline{D}^{(4)}(\rho)$ is the index corresponding to the fourth-order Casimir invariant I_4' given by

$$I_4' = \tfrac{1}{3}(g^{\mu\nu}g^{\alpha\beta} + g^{\mu\alpha}g^{\nu\beta} + g^{\mu\beta}g^{\nu\alpha}) t_\mu t_\nu t_\alpha t_\beta$$
$$= I_2[I_2 - \tfrac{1}{6} I_2(\rho_0)], \tag{5.5}$$

which is *not* fundamental. At any rate, Eq. (5.4) proves that any fourth-order general index $L_4(\rho)$ can be expressed in terms of two indices corresponding to fourth-order Casimir invariants J_4 and $(I_2)^2$. This reasoning is evidently applicable to any $L_p(\rho)$. Also, as we will explain in Sec. 7, the sum rule [involving $\overline{D}^{(4)}(\rho)$] for the decomposition equation (1.2) is equivalent to a quadratic sum rule for $D^{(2)}(\rho)$. Finally, the explicit relation expressing $l_4(\rho)$ in terms of $D_4(\rho)$ and $\overline{D}_4(\rho)$ is given in Ref. 11.

6. INDICES AND BRANCHING RULES

Let us consider the representation (2.12) of L_0 which arises as a result of reduction of a representation ω of $L \supset L_0$, and prove Eq. (2.13).

Suppose that the subalgebra L_0 is spanned by the first m elements $t_1, t_2, \ldots, t_m$ of L, where m is the dimension of L_0. In order to distinguish the basis of L_0 from that of L, we use the notation $\{t_j\}$ $(j = 1,\ldots,m)$ and $\{t_\mu\}$ $(\mu = 1,2,\ldots,d_0)$ for bases of L_0 and L, respectively. Now, the pth-order fundamental index $D_0^{(p)}(\rho)$ of L_0 will be written as

$$D_0^{(p)}(\rho) = g^{(0)j_1 j_2 \cdots j_p} \mathrm{Tr}(X_{j_1} X_{j_2} \cdots X_{j_p}), \tag{6.1}$$

where $g^{(0)j_1 j_2 \cdots j_p} t_{j_1} t_{j_2} \cdots t_{j_p}$ is the Casimir invariant $J_p^{(0)}$ of L_0. If ρ is a direct sum of irreducible components ρ_j, as in Eq. (2.12), we evidently have

$$D_0^{(p)}(\rho) = \sum_j D_0^{(p)}(\rho_j).$$

On the other hand, we may rewrite Eq. (6.1) as

$$D_0^{(p)}(\rho) = g^{(0)j_1 j_2 \cdots j_p} \frac{1}{p!} \sum_P \mathrm{Tr}(X_{j_1} X_{j_2} \cdots X_{j_p}) \tag{6.2}$$

and note that

$$\frac{1}{p!}\sum_P \mathrm{Tr}(X_{\mu_1} X_{\mu_2} \cdots X_{\mu_p})$$

regarded as a p-form over L can be expressed as

$$\frac{1}{p!}\sum_P \mathrm{Tr}(X_{\mu_1} \cdots X_{\mu_p}) = c_p D^{(p)}(\omega) g_{\mu_1 \cdots \mu_p}$$
$$+ c' g_{\mu_1\mu_2} g_{\mu_3\mu_4 \cdots \mu_p} + \cdots \tag{6.3}$$

by the same reasoning used already. Multiplying by $g^{\mu_1 \cdots \mu_p}$ and noting the orthogonality conditions, we find

$$c_p g^{\mu_1 \cdots \mu_p} g_{\mu_1 \cdots \mu_p} = 1 \tag{6.4}$$

so that c_p is independent of ω. Now we restrict the indices $\mu_1 \cdots \mu_p$ to subindices $j_1, j_2, \ldots, j_p$ of L_0, and insert the result of Eq. (6.3) into the right side of (6.2). We now note that $g_{j_1 \cdots j_p}$ can be expressed as a linear combination of $g^{(0)}_{j_1 j_2 \cdots j_p}$, $g^{(0)}_{j_1 j_2} g^{(0)}_{j_3 \cdots j_p}$, etc. by similar reasoning together with the explicit form of I_p to be discussed in Sec. 8. Therefore, the second and higher terms in the right side of Eq. (6.3) will give zero contribution in view of the orthogonality conditions such as

$$g^{(0)j_1 \cdots j_p} g^{(0)}_{j_1 j_2} g^{(0)}_{j_3 \cdots j_p} = 0$$

for L_0. Therefore, we find Eq. (2.13), i.e.,

$$\xi_p D^{(p)}(\omega) = \sum_k D_0^{(p)}(\rho_k) = D_0^{(p)}(\rho), \tag{6.5}$$

where ξ_p is given by

$$\xi_p = c_p g^{(0)j_1 j_2 \cdots j_p} g_{j_1 j_2 \cdots j_p}. \tag{6.6}$$

Evidently ξ_p does not depend upon ω.

We emphasize that the orthogonality conditions for $g^{\mu_1 \cdots \mu_p}$'s are crucial for this derivation. Therefore, the sum rule (6.5) will not hold, in general, for $L_p(\omega)$. The exceptions are, of course, the cases $p = 2, 3$. This is because the second- and third-order Casimir invariants J_2 and J_3 are unique, apart from the normalization constants, so that we have

$$L_2(\omega) = c_1 D^{(2)}(\omega), \qquad L_3(\omega) = c_2 D^{(3)}(\omega) \tag{6.7}$$

for some constants c_1 and c_2 which do not depend upon ω. This fact was previously known.[1,6,8]

We may remark that ξ_p can be identically zero for some choices of p, L, and L_0. We will calculate some explicit values of ξ_p in Sec. 8, for some special cases. As an interesting example of Eq. (6.5), let us identify $L = D_n$, $L_0 = A_{n-1}$, and $p = 3$. Since D_n, for $n > 3$, does not possess any third-order Casimir invariant J_3, we have $D^{(3)}(\omega) = 0$. Therefore, Eq. (6.5) becomes

$$\sum_j D_0^{(3)}(\rho_j) = 0.$$

When we choose ω to be the 2^{n-1} dimensional spinor representation of D_n corresponding to highest weight Λ_n and/or

A_{n-1}, this relation is rewritten more explicitly as

$$\sum_{j\,\text{even}}^{n-1} D_0^{(3)}(\Lambda_j) = 0,$$
$$\sum_{j\,\text{odd}}^{n-1} D_0^{(3)}(\Lambda_j) = 0, \tag{6.8}$$

where $\Lambda_1, \Lambda_2, \ldots, \Lambda_{n-1}$ are fundamental representations of the Lie algebra A_{n-1}. Since $D_0^{(3)}(\rho)$ is the triangle-anomaly coefficient[7,8] in gauge field theory, Eq. (6.8) assures the renormalizability of the theory. The special case of $n = 5$ corresponds to the famous model of Georgi and Glashow[15] for the SU(5) grand unification. We remark that Eq. (6.8) was first observed by Georgi[16] and utilized for SU(7) models of hypercolor theory as well as of grand unification theory by many authors.[17,18]

Also, the special case of $L = \mathrm{sl}(nm)$ and $L_0 = \mathrm{sl}(n) \otimes \mathrm{sl}(m)$ is physically interesting.[19] The sum rule (6.5) for $p = 2$ and 3 and Eq. (2.11) for $p = 2, 3, 4$ were utilized and applied for determination of possible preon models in particle physics by Schellekens *et al.*[20] These sum rules together with the congruence class conservation rule[21] are very useful for many practical purposes.

7. POLYNOMIAL INDEX SUM RULES

In Sec. 3, we have noted that any fourth-order general index $L_4(\rho)$ must satisfy the sum rule (3.3). If we choose $b^{\mu\nu\alpha\beta} = g^{\mu\nu\alpha\beta}$, then it will reproduce Eq. (2.11) for $p = 4$. Here, we shall choose $b^{\mu\nu\alpha\beta}$ to be

$$b^{\mu\nu\alpha\beta} = \tfrac{1}{3}(g^{\mu\nu}g^{\alpha\beta} + g^{\mu\alpha}g^{\nu\beta} + g^{\mu\beta}g^{\nu\alpha}). \tag{7.1}$$

Then it is easy to compute

$$L_4(\rho) = \overline{D}^{(4)}(\rho) = D^{(2)}(\rho)\left\{\frac{D^{(2)}(\rho)}{d(\rho)} - \frac{1}{6}\,\frac{D^{(2)}(\rho_0)}{d(\rho_0)}\right\}, \tag{7.2}$$

where $\overline{D}^{(4)}(\rho)$ is the same as in Eq. (5.3b). The sum rule (3.5) is now explicitly evaluated to give

$$\sum_j \frac{1}{d(\rho_j)}[D^{(2)}(\rho_j)]^2 = 4\,\frac{D^{(2)}(\rho_A)D^{(2)}(\rho_B)}{d(\rho_0)} + d(\rho_A)d(\rho_B)\left[\frac{D^{(2)}(\rho_A)}{d(\rho_A)} + \frac{D^{(2)}(\rho_B)}{d(\rho_B)}\right]^2, \tag{7.3}$$

which reproduces the quadratic sum rule of Ref. 11.

Next, let us consider a fifth-order index

$$L_5(\rho) = b^{\mu\nu\lambda\alpha\beta}\,\mathrm{Tr}(X_\mu X_\nu X_\lambda X_\alpha X_\beta).$$

If we set $b^{\mu\nu\lambda\alpha\beta}$ equal to $g^{\mu\nu\lambda\alpha\beta}$, then $L_5(\rho)$ will reduce to $D^{(5)}(\rho)$. Here, let us choose

$$\begin{aligned}10b^{\mu\nu\lambda\alpha\beta} = {} & g^{\mu\nu\lambda}g^{\alpha\beta} + g^{\mu\nu\alpha}g^{\lambda\beta} \\ & + g^{\mu\nu\beta}g^{\lambda\alpha} + g^{\mu\alpha\lambda}g^{\nu\beta} + g^{\mu\beta\lambda}g^{\nu\alpha} + g^{\alpha\nu\lambda}g^{\mu\beta} \\ & + g^{\beta\nu\lambda}g^{\alpha\mu} + g^{\alpha\beta\lambda}g^{\mu\nu} + g^{\alpha\beta\nu}g^{\mu\lambda} + g^{\alpha\beta\mu}g^{\nu\lambda},\end{aligned} \tag{7.4}$$

which is totally symmetric in five indices μ, ν, λ, α, and β. After some calculation, we find

$$L_5(\rho) = \frac{D^{(2)}(\rho)D^{(3)}(\rho)}{d(\rho)} - \frac{1}{4}\,\frac{D^{(2)}(\rho_0)}{d(\rho_0)}D^{(3)}(\rho). \tag{7.5}$$

Then, we find a sum rule analogous to Eq. (3.5):

$$\begin{aligned}\sum_j \frac{D^{(2)}(\rho_j)D^{(3)}(\rho_j)}{d(\rho_j)} &= \frac{d(\rho_B)}{d(\rho_A)}D^{(2)}(\rho_A)D^{(3)}(\rho_A) \\ &\quad + \frac{d(\rho_A)}{d(\rho_B)}D^{(2)}(\rho_B)D^{(3)}(\rho_B) + \frac{6 + d(\rho_0)}{d(\rho_0)} \\ &\quad \times\{D^{(2)}(\rho_A)D^{(3)}(\rho_B) + D^{(2)}(\rho_B)D^{(3)}(\rho_A)\}.\end{aligned} \tag{7.6}$$

Since $D^{(3)}(\rho) = 0$ identically[8] for all Lie algebras except for A_n $(n \geqslant 2)$, Eq. (7.6) is useful only for the Lie algebra A_n $(n \geqslant 2)$.

If we study $L_6(\rho)$ with suitable choices for $b^{\mu\nu\lambda\alpha\beta\gamma}$, we will find more complicated polynomial sum rules. However, there exists a simpler way of obtaining some of these sum rules. Following the method described in Ref. 13, we know that we have

$$\begin{aligned}\frac{1}{8}\sum_j [I_2(\rho_j) - I_2(\rho_A) - I_2(\rho_B)]^3 d(\rho_j) \\ = \mathrm{Tr}^{(A)}(X^\mu X^\nu X^\alpha)\,\mathrm{Tr}^{(B)}(X_\mu X_\nu X_\alpha)\end{aligned} \tag{7.7}$$

for the decomposition (1.2). We can readily evaluate

$$\mathrm{Tr}(X_\mu X_\nu X_\alpha) = \frac{D^{(3)}(\rho)}{D^{(3)}(\lambda)}g_{\mu\nu\alpha} + \frac{1}{2}\,\frac{D^{(2)}(\rho_0)}{d(\rho_0)}g_{\alpha\beta}c^{\beta}_{\mu\nu}, \tag{7.8}$$

where λ in $D^{(3)}(\lambda)$ is the arbitrary but fixed representation which is called the reference representation in Ref. 11. Here, the normalization of $g_{\mu\nu\alpha}$ is chosen to be

$$g_{\mu\nu\alpha} = \tfrac{1}{2}\,\mathrm{Tr}^{(\lambda)}(x_\mu x_\nu + x_\nu x_\mu)x_\alpha \tag{7.9}$$

in the reference representation λ. This will be explained again in the next section. Inserting (7.8) in the right side of (7.7), we find the following cubic sum rule:

$$\begin{aligned}\sum_j\left\{\frac{D^{(2)}(\rho_j)}{d(\rho_j)} - \frac{D^{(2)}(\rho_A)}{d(\rho_A)} - \frac{D^{(2)}(\rho_B)}{d(\rho_B)}\right\}^3 d(\rho_j) \\ = 8\,\frac{D^{(3)}(\rho^A)D^{(3)}(\rho^B)}{D^{(3)}(\lambda)} \\ - 2\,\frac{D^{(2)}(\rho_0)}{[d(\rho_0)]^2}D^{(2)}(\rho_A)D^{(2)}(\rho_B).\end{aligned} \tag{7.10}$$

For all simple Lie algebras other than A_n $(n \geqslant 2)$, we have $D^{(3)}(\rho) = 0$ identically.[8] In that case, we delete the first term in the right-hand side of Eq. (7.10). Then, it will give a cubic sum rule involving only the second-order index $D^{(2)}(\rho)$.

Similarly, we can find the following quartic identity when we normalize $D^{(4)}(\rho)$ and $D^{(3)}(\rho)$ suitably as in Ref. 11:

$$\begin{aligned}\frac{1}{16}\sum_j\left[\frac{D^{(2)}(\rho_j)}{d(\rho_j)} - \frac{D^{(2)}(\rho_A)}{d(\rho_A)} - \frac{D^{(2)}(\rho_B)}{d(\rho_B)}\right]^4 d(\rho_j) \\ = \frac{1}{2 + d(\rho_0)}\left[\frac{D^{(4)}(\rho_A)D^{(4)}(\rho_B)}{D^{(4)}(\lambda)} + 3\,\frac{d(\rho_A)d(\rho_B)}{d(\rho_0)}H(\rho_A)H(\rho_B)\right] \\ - \frac{1}{2}\,\frac{D^{(2)}(\rho_0)}{d(\rho_0)}\,\frac{D^{(3)}(\rho_A)D^{(3)}(\rho_B)}{D^{(3)}(\lambda)} \\ + \frac{1}{12}\,\frac{[D^{(2)}(\rho_0)]^2}{[d(\rho_0)]^3}D^{(2)}(\rho_A)D^{(2)}(\rho_B),\end{aligned} \tag{7.11}$$

where we have set[11]

$$H(\rho)=\frac{D^{(2)}(\rho)}{d(\rho)}\left[\frac{D^{(2)}(\rho)}{d(\rho)}-\frac{1}{6}\frac{D^{(2)}(\rho_0)}{d(\rho_0)}\right], \quad (7.12)$$

which is identical to $\bar{D}^{(4)}(\rho)$ given by Eq. (5.3b). For the exceptional Lie algebras G_2, F_4, E_6, E_7, and E_8 as well as A_1 and A_2, we have $D^{(3)}(\rho)=D^{(4)}(\rho)=0$ identically, and we omit all terms involving $D^{(3)}(\rho)$ and $D^{(4)}(\rho)$ in Eq. (7.11) for such cases. Then it gives a quartic sum rule involving only $D^{(2)}(\rho)$. Also, for these special algebras, we have proved in Ref. 11 that $l_4(\rho)$ defined by Eq. (1.4) is expressed in terms of a quadratic polynomial of $l_2(\rho)$ and that the sum rule (1.5) is essentially equivalent to the quadratic one of Eq. (7.3). The special case of $\rho_A=\rho_B=\rho_0$ for the exceptional Lie algebras G_2, F_4, E_6, E_7, and E_8 is interesting, since we have many polynomial identities for $D^{(2)}(\rho_j)$ and we can essentially determine[13] them from these sum rules.

8. FIFTH-ORDER CASIMIR INVARIANTS

In order to find explicit forms of fundamental Casimir invariants J_p's, we proceed as follows: Let λ be an arbitrary but fixed nonzero irreducible representation of L, which we called[11] a reference representation. Further, let x_μ be the representation matrix of t_μ in λ and define

$$h_{\mu_1\mu_2\cdots\mu_p}=\frac{1}{p!}\sum_P \mathrm{Tr}(x_{\mu_1}x_{\mu_2}\cdots x_{\mu_p}), \quad (8.1)$$

where the summation is over the $p!$ permutations P of p indices $\mu_1,\mu_2,\ldots,\mu_p$. Then we can construct a pth-order Casimir invariant I_p by

$$I_p=h^{\mu_1\mu_2\cdots\mu_p}t_{\mu_1}t_{\mu_2}\cdots t_{\mu_p}. \quad (8.2)$$

It has recently been shown[22] that essentially all fundamental pth-order Casimir invariants can be constructed from I_p, if we choose the reference representation λ to be the basic (or lowest-dimensional) representation of L. (Adopting the lexicographic ordering convention of the simple root system of Ref. 4, the highest weight of λ is Λ_1.) For this reason, we choose hereafter the reference representation λ to be the basic representation. Moreover, we normalize $g_{\mu\nu}$ by

$$g_{\mu\nu}=h_{\mu\nu}=\mathrm{Tr}(x_\mu x_\nu) \quad (8.3)$$

as in Ref. 11 so that we have

$$D^{(2)}(\lambda)=d(\lambda)I_2(\lambda)=d(\rho_0). \quad (8.4)$$

The Lie algebra D_n possesses one more nth-order fundamental Casimir invariant $\hat{I}_n$ which cannot be obtained in this way. It can, however, be obtained[23] by choosing λ to be the spinor representation of D_n. But the canonical form of $\hat{J}_n$ is well known,[24–26] with its eigenvalue

$$\hat{J}_n(\rho)=l_1l_2\cdots l_n \quad (8.5)$$

in the notation of Ref. 25. The corresponding nth-order fundamental index is then defined by

$$\hat{D}^{(n)}(\rho)=d(\rho)\hat{J}_n(\rho) \quad (8.6)$$

for this case, which will satisfy the sum rules (2.11) and (2.13).

Now, as in Sec. 1, J_2 and J_3 can be readily identified as

$$J_2=I_2,\quad J_3=I_3 \quad (8.7)$$

for this case, which will satisfy the sum rules (2.11) and (2.13). When n is an even integer, then D_n possesses two fundamental nth-order Casimir invariants $\hat{J}_n$ and J_n. Otherwise, J_p is essentially unique.

Before going into detail, we may remark the following. Let L_0 be a semisimple subalgebra of L. We label the basis of L and L_0 as $\{t_\mu\}$ and $\{t_j\}$ as in Sec. 6. Then, if we restrict the Greek indices μ's to subindices j's, the pth-order Casimir invariant I_p of L given by

$$I_p=h^{\mu_1\cdots\mu_p}t_{\mu_1}\cdots t_{\mu_p}$$

will induce a pth-order Casimir invariant $\tilde{I}_p^{(0)}$ of L_0 by

$$\tilde{I}_p^{(0)}=h^{j_1j_2\cdots j_p}t_{j_1}t_{j_2}\cdots t_{j_p}, \quad (8.8)$$

which, however, may not necessarily be fundamental. This is the reason $g^{j_1\cdots j_p}$ may be decomposed into a sum of $g^{(0)j_1\cdots j_p}$, etc. as in the discussion of Sec. 6 for the derivation of the branching sum rule (6.5).

Returning to the original discussion, we chose J_2 and J_3 to be

$$J_2=I_2,\quad J_3=I_3$$

as in Eq. (8.7). For any Lie algebra other than A_n ($n\geqslant 2$), we have of course $J_3(\rho)=0$ identically.[8]

For $p\geqslant 4$, the situation is more involved. The explicit form of J_4 with an appropriate normalization is given by[11]

$$J_4=[2+d(\rho_0)]I_4 - 3\frac{d(\rho_0)}{d(\lambda)}\left\{1-\frac{1}{6}\frac{I_2(\rho_0)}{I_2(\lambda)}\right\}\left\{I_2-\frac{1}{6}I_2(\rho_0)\right\}I_2. \quad (8.9)$$

We may similarly construct J_5. The coefficient $g_{\mu\nu\lambda\alpha\beta}$ satisfying the orthogonality condition (3.14b) is given by

$$g_{\mu\nu\lambda\alpha\beta}=[6+d(\rho_0)]h_{\mu\nu\lambda\alpha\beta} - 10\frac{d(\rho_0)}{d(\lambda)}\left\{1-\frac{1}{4}\frac{I_2(\rho_0)}{I_2(\lambda)}\right\}b_{\mu\nu\lambda\alpha\beta}, \quad (8.10)$$

where $b_{\mu\nu\lambda\alpha\beta}$ is defined in Eq. (7.4). Then J_5 is calculated to be

$$J_5=g^{\mu\nu\lambda\alpha\beta}t_\mu t_\nu t_\lambda t_\alpha t_\beta=[6+d(\rho_0)]I_5 - \frac{5}{8}\frac{d(\rho_0)}{d(\lambda)}\left\{4-\frac{I_2(\rho_0)}{I_2(\lambda)}\right\}\{4I_2-I_2(\rho_0)\}I_3. \quad (8.11)$$

Next, let us evaluate explicit values of $J_p(\rho)$. For $p\leqslant 4$, they are calculated in Ref. 11, so that we will consider here the case of $p=5$. Since the only simple Lie algebras with a fifth-order Casimir invariant are A_n ($n\geqslant 4$), E_6, and D_5, we see that $J_5(\rho)$ is identically zero except for these Lie algebras. For D_5, it is given by Eq. (8.5) with $J_5(\rho)\equiv\hat{J}_5(\rho)$ for this case. For E_6, we note that E_6 has no third-order Casimir invariant. Hence, we have

$$J_5(\rho)=[6+d(\rho_0)]I_5(\rho). \quad (8.12)$$

In order to evaluate $I_5(\rho)$, we introduce[25–27] a nonsymmetrized Casimir invariant $I_5^{(N,S)}$:

$$I_5^{(N,S)}=a^{\mu\nu\lambda\alpha\beta}t_\mu t_\nu t_\lambda t_\alpha t_\beta,$$

$$a_{\mu\nu\lambda\alpha\beta}=\mathrm{Tr}(x_\mu x_\nu x_\lambda x_\alpha x_\beta).$$

Note that $a_{\mu\nu\lambda\alpha\beta}$ is not symmetric for exchanges of their indices. However, the evaluation of $I_5^{(N,S)}(\rho)$ is relatively straightforward.[25]

For any simple Lie algebra other than A_n $(n \geqslant 2)$ and D_4, we can show that we can express $I_5(\rho)$ as

$$[2+d(\rho)]I_5(\rho)=[2+d(\rho_0)]I_5^{(N,S)}(\rho)$$
$$+\frac{1}{12}\left[10I_2(\rho_0)-\frac{D^{(4)}(\rho_0)}{D^{(4)}(\lambda)}\right]J_4(\rho)$$
$$-\frac{1}{12}J_4(\rho_0)I_2(\rho)+\frac{5}{8}I_2(\rho_0)[4I_2(\lambda)-I_2(\rho_0)][I_2(\rho)]^2$$
$$+\frac{5}{144}[I_2(\rho_0)]^2\{[d(\rho_0)+6]I_2(\rho_0)-18I_2(\lambda)\}I_2(\rho) \tag{8.13}$$

by using the method given in Ref. 11. For the Lie algebra E_6, we may set

$$J_4(\rho)=J_4(\rho_0)=0$$

in Eq. (8.13) since E_6 possesses no fundamental fourth-order Casimir invariant. Englefield has computed and tabulated[28] eigenvalues of $I_5^{(N,S)}(\rho)$ for many low-dimensional representations of E_6, so that we can evaluate $J_5(\rho)$ for E_6 from Eqs. (8.12) and (8.13). We remark that both $J_5(\rho)$ and $I_5(\rho)$ change their signs when we replace ρ by its contragradient representation ρ^*. However, this nice property is not shared by $I_5^{(N,S)}(\rho)$. Moreover, $I_5(\rho)$ [but not $I_5^{(N,S)}(\rho)$] is identically zero for all simple Lie algebras other than A_n $(n\geqslant 2)$, E_6, and D_5. Also, we used the normalization condition (8.4) for the derivation of Eq. (8.13). We can rewrite Eq. (8.12) as

$$J_5(\rho)=\tfrac{1}{2}[6+d(\rho_0)]\{I_5(\rho)-I_5(\rho^*)\}=\tfrac{1}{2}[6+d(\rho_0)]$$
$$\times\{I_5^{NS}(\rho)-I_5^{NS}(\rho^*)\}. \tag{8.12}$$

Another way of computing $D^{(5)}(\rho)$ is to calculate directly

$$l_5'(\rho)=\sum_M (v,M)^5$$

as in Eq. (1.6), which must by proportional to $D^{(5)}(\rho)$ for E_6 as we see from arguments given in Sec. 3.

Finally, we have to consider the remaining case of A_n $(n\geqslant 2)$. Following the method of Ref. 25, we embed the Lie algebra A_n into the Lie algebra of the $U(n+1)$ group whose irreducible representation is characterized by $n+1$ integers satisfying

$$f_1\geqslant f_2\geqslant\cdots\geqslant f_n\geqslant f_{n+1}. \tag{8.14}$$

For simplicity, we set

$$N=n+1, \tag{8.15}$$

$$\sigma_j=f_j+\frac{(N+1)}{2}-j-\frac{1}{N}\sum_{k=1}^N f_k, \tag{8.16}$$

as before. Then, the eigenvalues of J_2, J_3, and J_4 have been computed[11] to be

$$J_2(\rho)=\sum_{j=1}^N(\sigma_j)^2-\frac{1}{12}N(N^2-1), \tag{8.17a}$$

$$J_3(\rho)=\sum_{j=1}^N(\sigma_j)^3, \tag{8.17b}$$

$$J_4(\rho)=(N^2+1)\sum_{j=1}^N(\sigma_j)^4-\frac{2N^2-3}{N}\left[\sum_{j=1}^N\sigma_j^2\right]^2$$
$$+\frac{1}{720}N(N^2-1)(N^2-4)(N^2-9). \tag{8.17c}$$

Similarly, we have computed $J_5(\rho)$ here to be

$$J_5(\rho)=(N^2+5)\sum_{j=1}^N(\sigma_j)^5-\frac{5(N^2-2)}{N}\left(\sum_{j=1}^N\sigma_j^2\right)\left(\sum_{k=1}^N\sigma_k^3\right). \tag{8.18}$$

We may easily verify the fact that Eq. (8.18) ensures $J_5(\rho)=0$ identically for $N=2$, 3, and 4, corresponding to the Lie algebra A_1, A_2, and A_3, when we note an identity[13]

$$\sum_{j=1}^4(\sigma_j)^5=\frac{5}{6}\left(\sum_{j=1}^4\sigma_j^2\right)\left(\sum_{k=1}^4\sigma_k^3\right)$$

for any σ_1, σ_2, σ_3, and σ_4 satisfying $\Sigma_{j=1}^4\sigma_j=0$. For possible applications to particle physics, we simply list values of $J_5(\Lambda_k)$ and $J_5(k\Lambda_1)$ here to be

$$J_5(\Lambda_k)=[(N+1)(N+2)(N+3)(N+4)/24N^2]$$
$$\times k(N-k)(N-2k)\{N(N+5)-12k(N-k)\}$$
$$(N-1\geqslant k\geqslant 1), \tag{8.19}$$

$$J_5(k\Lambda_1)=[(N-1)(N-2)(N-3)(N-4)/24N^2]$$
$$\times k(N+k)(N+2k)\{N(N-5)+12k(N+k)\}$$
$$(k\geqslant 1). \tag{8.20}$$

We remark that the eigenvalue $J_5(\Lambda_k)$ corresponding to the completely antisymmetric representation can be obtained[14] from $J_5(k\Lambda_1)$ corresponding to the completely symmetric one by the formal replacement $N\to -N$ in accordance with a general theorem proved[29] by Cvitanović and Kennedy.

We can also derive a fifth-order trace identity. Let t be a generic element so that

$$t=\xi^\mu t_\mu\in L \tag{8.21}$$

for some real or complex numbers ξ^μ's. We then denote by X its representation matrix in the generic irreducible representation, so that

$$X=\xi^\mu X_\mu. \tag{8.22}$$

Following the argument given in Ref. 11, we then find

$$\operatorname{Tr}X^5-A(\rho)\operatorname{Tr}X^2\operatorname{Tr}X^3=c_5(t)D^{(5)}(\rho), \tag{8.23a}$$

$$D^{(5)}(\rho)=d(\rho)J_5(\rho), \tag{8.23b}$$

where $c_5(t)$ may depend upon the form of t but not upon ρ and where $A(\rho)$ is defined by

$$A(\rho)=\frac{5d(\rho_0)}{2[6+d(\rho_0)]d(\rho)}\left(4-\frac{I_2(\rho_0)}{I_2(\rho)}\right). \tag{8.24}$$

If ρ_1 and ρ_2 are two irreducible representations of L, then Eq. (8.23a) implies the validity of

$$\frac{\operatorname{Tr}^{(2)}X^5-A(\rho_2)\operatorname{Tr}^{(2)}X^2\operatorname{Tr}^{(2)}X^3}{\operatorname{Tr}^{(1)}X^5-A(\rho_1)\operatorname{Tr}^{(1)}X^2\operatorname{Tr}^{(1)}X^3}=\frac{D^{(5)}(\rho_2)}{D^{(5)}(\rho_1)}. \tag{8.25}$$

Here, $\operatorname{Tr}^{(j)}$ $(j=1,2)$ designates the trace in the representation space of ρ_j $(j=1,2)$.

This relation can be used to simplify the previous proof[30] of the uniqueness of the grand unified group SU(5).

Since $J_5(\rho) = 0$ for Lie algebras A_2 and A_3, Eq. (8.23) also implies the validity of

$$\mathrm{Tr}\, X^5 = A(\rho)\, \mathrm{Tr}\, X^2\, \mathrm{Tr} X^3 \tag{8.26}$$

for any irreducible representation of these Lie algebras, reproducing the result of Ref. 13. For $N = 16$ and $\rho = \Lambda_2$, we have $J_5(\Lambda_2) = 0$ so that Eq. (8.26) holds also for this case of A_{15}.

Since we now know explicit values of $J_p(\rho)$ ($p \geqslant 5$), we can evaluate values of ξ_p which appear in Eq. (2.13). For example, consider the case of $L = A_{N-1} = \mathrm{su}(N)$ and $L_0 = A_{N-2} = \mathrm{su}(N-1)$. Then choosing ρ to be the basic representation λ, we readily compute ξ_p to be

$$\xi_2 = N(N-2)/(N-1)(N+1), \tag{8.27a}$$

$$\xi_3 = N^2(N-3)/(N-1)^2(N+2), \tag{8.27b}$$

$$\xi_4 = N(N-4)/(N-1)(N+3), \tag{8.27c}$$

$$\xi_5 = N^2(N-5)/(N-1)^2(N+4). \tag{8.27d}$$

We note the fact that ξ_p ($p \leqslant 5$) always contains a factor $N-p$. This is not a coincidence, since we must have $D_0^{(N)}(\rho) = 0$ identically for Lie algebras $L_0 = A_{N-2} = \mathrm{su}(N-1)$ with $p = N$.

Similarly, if we identify $L = D_n$ and $L_0 = A_{n-1}$, we calculate

$$\xi_2 = 2(n^2-1)/n(2n-1), \tag{8.28a}$$

$$\xi_4 = \frac{4(n^2-4)(n^2-9)}{n(4n^2-1)(2n-3)} \tag{8.28b}$$

while we may set $\xi_3 = 0$ identically, for $n \geqslant 4$, and $\xi_5 = 0$ for $n \geqslant 6$. We have checked the validity of our sum rules for many simple special irredicible representations. For example, consider the case of $A_n = \mathrm{su}(n+1)$ with product decomposition

$$\Lambda_1 \otimes \Lambda_1 = \Lambda_2 \oplus 2\Lambda_1.$$

Then, we should have a sum rule

$$2d(\Lambda_1) D^{(p)}(\Lambda_1) = D^{(p)}(\Lambda_1) + D^{(p)}(2\Lambda_1)$$

for any p. This can be explicitly verified with uses of our values of $J_p(\rho)$ ($p \leqslant 5$). Similarly for $L = A_n = \mathrm{su}(n+1)$ and $L_0 = A_{n-1} = \mathrm{su}(n)$, the branching rule for $A_n \to A_{n-1}$,

$$\{\Lambda_2\} \to \{\Lambda_2\} \oplus \{\Lambda_1\},$$

$$\{2\Lambda_1\} \to \{2\Lambda_1\} \oplus \{\Lambda_1\} \oplus \{0\}$$

induces the branching sum rules

$$\xi_p D^{(p)}(\Lambda_2) = D_0^{(p)}(\Lambda_2) + D_0^{(p)}(\Lambda_1),$$

$$\xi_p D^{(p)}(2\Lambda_1) = D_0^{(p)}(2\Lambda_1) + D_0^{(p)}(\Lambda_1),$$

which can be again verified by our explicit formulas for ξ_p and $J_p(\rho)$.

In ending this note, we may remark the following. The calculation of $J_6(\rho)$ will be very interesting, since practically all simple Lie algebras possess fundamental sixth-order Casimir invariants. However, the direct explicit evaluation of $J_p(\rho)$ for $p \geqslant 6$ is very complicated. In a subsequent paper,[14] we will utilize a different approach for direct evaluation of $D^{(p)}(\rho)$ rather than $J_p(\rho)$ itself from character formula of classical groups. Also, our genral Dynkin indices are useful for purposes of decomposing symmetric or antisymmetric tensor products of the same irreducible representation into irreducible components. These problems will be discussed in a subsequent paper.[14]

ACKNOWLEDGMENTS

We are grateful to Dr. R. T. Sharp for a number of useful comments. One of the authors (S. O.) is supported in part by the U.S. Department of Energy under contract No. DE-AC02-76ER13065. The other (J. P.) is supported by the National Science and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.

[1] E. B. Dynkin, Math. Sb. SSSR **30**, 349 (1952) [Am. Math. Soc. Trans., Ser. 2, **6**, 111 (1957)]. The definition of the Dynkin equation (2.27) differs from the present one by a factor equal to the dimension of L.
[2] L. C. Biedenharn, "Group Theory and the Classification of Elementary Particles," CERN Report 65-41 (1965), Sec. B (unpublished).
[3] J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1972 (1976); Erratum **18**, 1519 (1977). We have replaced the symbol $I^{(2p)}$ by $l_{2p}(\rho)$ here since we reserve $I_p(\rho)$ for the eigenvalues of the pth-order Casimir operator.
[4] J. McKay, J. Patera, and R. T. Sharp, J. Math. Phys. **22**, 2770 (1981).
[5] W. G. McKay and J. Patera, *Tables of Dimensions, Indices and Branching Rules for Representations of Simple Lie Algebras* (Dekker, New York, 1981).
[6] J. Patera and R. T. Sharp, J. Math. Phys. **22**, 2352 (1981).
[7] J. Banks and H. Georgi, Phys. Rev. D **14**, 1159 (1976).
[8] S. Okubo, Phys. Rev. D **16**, 3528 (1977).
[9] M. Bremner, R. V. Moody, and J. Patera, *Tables of dominant weight multiplicities for representations of simple Lie algebras of rank* $\leqslant 8$ (Marcel Dekker, New York, 1984).
[10] R. V. Moody and J. Patera, Bull. Am. Math. Soc. 7, 237 (1982).
[11] S. Okubo, J. Math. Phys. **23**, 8 (1982).
[12] J. Dixmier, *Enveloping Algebras* (North-Holland, Amsterdam, 1977).
[13] S. Okubo, J. Math. Phys. **20**, 586 (1979).
[14] S. Okubo and J. Patera, J. Math. Phys. **24**, 2722 (1983).
[15] H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).
[16] H. Georgi, Nucl. Phys. B **156**, 126 (1979).
[17] E. Farhi and L. Susskind, Phys. Rev. D **20**, 3404 (1979). I. Umemura and K. Yamaoto, Phys. Lett. B **100**, 34 (1981); M. Claudson, A. Yildiz, and P. M. Cox, Phys. Lett. B **97**, 224 (1980).
[18] J. E. Kim, Phys. Rev. Lett. **45**, 1916 (1980); Phys. Rev. D **23**, 2706 (1981); P. Cox, P. Frampton, and A. Yildiz, Phys. Rev. Lett. **46**, 1051 (1981).
[19] C. Itzykson and M. Nauenberg, Rev. Mod. Phys. **38**, 95 (1966).
[20] A. N. Schellekens, I.-G. Koh, and K. Kang, J. Math. Phys. **23**, 2244 (1982); A. N. Schellekens, K. Kang, and I.-G. Koh, Phys. Rev. D **26**, 658 (1982).
[21] F. W. Lemire and J. Patera, J. Math. Phys. **21**, 2026 (1980).
[22] F. Berdjis, J. Math. Phys. **22**, 1851 (1981); F. Berdjis and E. Beslmüller, *ibid.* **22**, 1857 (1981).
[23] B. Gruber and L. O'Raifeartaigh, J. Math. Phys. **4**, 436 (1963).
[24] J. D. Louck, "Theory of angular momentum in N-dimensional space," Los Alamos Scientific Laboratory 1960, unpublished.
[25] S. Okubo, J. Math. Phys. **18**, 2382 (1977).
[26] V. S. Popov and A. M. Perelomov, Yad. Fiz. **5**, 693 (1967) [Sov. J. Nucl. Phys. **5**, 489 (1967)].
[27] See Ref. 25. Also, S. A. Edwards, J. Math. Phys. **19**, 164 (1978); M. J. Englefield and R. C. King, J. Phys. A **13**, 2297 (1980).
[28] M. J. Englefield, "Tabulation of Kronecker Products of Representations of F_4, E_6, and E_7," University of Southhampton Report No. 57, 1980.
[29] P. Cvitanović and A. D. Kennedy, Phys. Scripta **26**, 5 (1982).
[30] Y. Tosa and S. Okubo, Phys. Rev. D **23**, 3058 (1981). S. Okubo, Hadronic J. **5**, 7 (1981).

Variational bounds on the temperature distribution

Kalman Kalikstein
Physics Department, Hunter College, City University of New York, New York, New York 10021[a] *and Physics Department, New York University, New York, New York 10003*

Larry Spruch
Physics Department, New York University, New York, New York 10003

Alberto Baider
Mathematics Department, Hunter College, City University of New York, New York, New York 10021

(Received 1 March 1983; accepted for publication 26 August 1983)

Upper and lower stationary or variational bounds are obtained for functions which satisfy parabolic linear differential equations. (The error in the bound, that is, the difference between the bound on the function and the function itself, is of second order in the error in the input function, and the error is of known sign.) The method is applicable to a range of functions associated with equalization processes, including heat conduction, mass diffusion, electric conduction, fluid friction, the slowing down of neutrons, and certain limiting forms of the random walk problem, under conditions which are not unduly restrictive: in heat conduction, for example, we do not allow the thermal coefficients or the boundary conditions to depend upon the temperature, but the thermal coefficients can be functions of space and time and the geometry is unrestricted. The variational bounds follow from a maximum principle obeyed by the solutions of these equations.

PACS numbers: 02.30. + g

I. INTRODUCTION

We will address ourselves to the question of the construction of upper and lower variational bound principles for functions which are the solutions of differential equations of equalization processes. These include heat conduction, mass diffusion, electric conduction, fluid friction, the slowing down of neutrons, and certain limiting forms of the random walk problem. Before doing so, however, it will be useful to make a few remarks on terminology, to establish some of the notation, and to record some of the properties of the time translation operator $U(t,t')$ and of the Green's function $G(\mathbf{r},t;\mathbf{r}',t')$, and to review briefly some previous results on variational bounds in other areas.

A. Some terminology

We firstly consider terminology. By a stationary principle one means a principle which provides an estimate of some quantity Q of interest for which the error in the estimate is of the order of the *square* of the input error. (In a Rayleigh–Ritz estimate of the energy E of a system, for example, the error δE in the estimate of E is a weighted average of the square of $\delta\psi = \psi_{\rm tr} - \psi$, where ψ is the exact normalized wave function and $\psi_{\rm tr}$ is a trial normalized wave function.) By a variational principle, one means a stationary principle in which the trial input function contains open parameters which are to be determined by demanding that the estimate Q_v of Q be stationary with respect to variation of each parameter. The distinction between stationary and variational is irrelevant for our purposes, and we will follow common usage, in which the terms are used interchangeably. By an upper (lower) variational bound, one means an estimate which is not only variational but for which the sign of $Q_v - Q$ is known to be positive (negative). The Rayleigh–Ritz estimate of the ground state energy of a system is an example of an upper variational bound. We will use VP for variational principle and V Bd for variational bound.

B. Some notation and some properties of $U(t,t')$ and $G(\mathbf{r},t;\mathbf{r}',t')$

We now give a very brief discussion of some properties of U and G. We are not here concerned with proofs; we wish simply to recall and record some properties which will be needed later. We consider a linear operator $A(t)$ of the form

$$A(t) = K(t) - \frac{\partial}{\partial t}. \tag{1.1a}$$

We could be far more general, but we will assume that $K(t)$ is of the form appropriate to the Schroedinger equation, for which $K(t) = -iH(t)/\hbar$, with $H(t)$ a Hamiltonian, or of a form appropriate to an equalization process, for which

$$\begin{aligned}\langle \mathbf{r}|A(t)|\mathbf{r}'\rangle &= \delta(\mathbf{r}-\mathbf{r}')\Big(a(\mathbf{r},t)\nabla\cdot b(\mathbf{r},t)\nabla - \frac{\partial}{\partial t}\Big)\\ &\equiv \delta(\mathbf{r}-\mathbf{r}')\Big(K(\mathbf{r},t) - \frac{\partial}{\partial t}\Big)\\ &\equiv \delta(\mathbf{r}-\mathbf{r}')A(\mathbf{r},t),\end{aligned} \tag{1.1b}$$

where ∇ operates in $\mathbf{r}$ space. We now assume that the "state vector" $F(t)$ is defined by

$$A(t)F(t) = 0, \quad t \geqslant \tau, \tag{1.2}$$

and by an initial condition, its value $F(\tau)$ at the initial time τ. $F(t)$ has components defined by $F(\mathbf{r},t) \equiv \langle \mathbf{r}|F(t)\rangle$. $F(t)$ can be expressed in terms of the state vector $F(t')$ at an earlier time t' by means of the relationship

$$F(t) = U(t,t')F(t'), \quad t' \geqslant \tau, \tag{1.3}$$

a relationship which defines the time-translation operator

[a] Permanent address.

$U(t,t')$. It follows immediately that

$$U(t',t') = 1, \tag{1.4}$$

where 1 is the identity operator, and that

$$U(t,t')U(t',t'') = U(t,t''), \tag{1.5}$$

where $t \geqslant t' \geqslant t''$. Inserting Eq. (1.3) into Eq. (1.2) and noting that $F(t')$ is arbitrary, it follows that

$$A(t)U(t,t') = 0. \tag{1.6}$$

Equations (1.4) and (1.6) provide an alternative definition of $U(t,t')$. The determination of the "wave function" $F(\mathbf{r},t)$ requires a knowledge of $F(\mathbf{r}',t')$ at *all* values of $\mathbf{r}'$, that is,

$$F(\mathbf{r},t) = \int G(\mathbf{r},t;\mathbf{r}',t')F(\mathbf{r}',t')d\mathbf{r}'. \tag{1.7}$$

This relationship defines the Green's function, $G(\mathbf{r},t;\mathbf{r}',t')$. On setting $t = t'$ in Eq. (1.7) and noting that $F(\mathbf{r}',t')$ is arbitrary, one finds

$$G(\mathbf{r},t';\mathbf{r}',t') = \delta(\mathbf{r}-\mathbf{r}'). \tag{1.8}$$

The completeness of the set of functions $|\mathbf{r}'\rangle$ enables us to rewrite Eq. (1.3) as

$$\langle \mathbf{r}|F(t)\rangle = \int \langle \mathbf{r}|U(t,t')|\mathbf{r}'\rangle d\mathbf{r}' \langle \mathbf{r}'|F(t')\rangle;$$

comparison with Eq. (1.7) gives

$$\langle \mathbf{r}|U(t,t')|\mathbf{r}'\rangle = G(\mathbf{r},t;\mathbf{r}',t'). \tag{1.9}$$

C. Some previous results on variational bounds

There exists a unified formulation for the construction of VP's which is applicable to just about any problem in mathematical physics.[1,2] This formulation can be used,[3] in particular, to obtain a VP for the temperature distribution $T(\mathbf{r},t)$ in a heat conduction problem and for functionals $F(T)$. Perhaps not surprisingly, there does not exist a unified formulation for the construction of V Bd's; the construction of a V Bd is a much more difficult task, for it includes the construction of a VP *and* the determination of the sign of the error. Nevertheless, V Bd's do exist in a number of areas. Thus, for example, there exist variational bounds on classical quantities such as power dissipation and capacitance.[4] Further, as noted earlier, the Rayleigh–Ritz principle provides an upper V Bd on the ground state energy E_{gd} of a system; the existence of this upper V Bd originates in the fact that the Hamiltonian H is bounded from below and that E_{gd} is the lowest point in the spectrum. The literature also contains an upper V Bd on the scattering length which characterizes scattering at zero incident relative energy.[5,6] The upper V Bd on the scattering length, which is as simple to use as the Rayleigh–Ritz V Bd on E_{gd}, again originates in the fact that H is bounded from below. In this case zero incident relative kinetic energy represents the lowest point in the continuous spectrum; if there exist any discrete eigenvalues (below the continuum), these must be and can be accounted for.

A further comment on the methods used in the development of a V Bd on E_{gd} or on the scattering length will provide an insight into the procedure to be used to derive a V Bd on $T(\mathbf{r},t)$. Letting Q represent either E_{gd} or the scattering length, it is possible to obtain a variational identity, an expression of the form

$$Q = Q_v + \Delta Q,$$

where Q_v is an explicit calculable VP for Q. ΔQ is a formal expression, for it contains the unknown wave function ψ—the normalized ground state wave function if one is considering E_{gd}, or the appropriately normalized scattering wave function. Even though ΔQ is formal it is simple to prove that ΔQ is of second order in $\delta\psi$ (the error in ψ)—which proves that Q_v is indeed a VP—and that $\Delta Q \geqslant 0$. Thus, in the ground state case one has $\Delta Q = \langle \delta\psi|H - E_{gd}|\delta\psi\rangle$, and $H - E_{gd}$ is non-negative in the space of square integrable functions. In the scattering case we restrict ourselves for simplicity only to potential scattering, and, further, to a potential which cannot support any bound states. H is then non-negative in the space of square integrable functions. In this case one has $\Delta Q = \langle \delta\psi|H|\delta\psi\rangle$, where $\delta\psi$ is *not* square integrable but rather approaches a constant at large distances; however, one can easily prove that H is non-negative for this class of functions too.

We turn now to the development of V Bd's in time-dependent scattering theory, filling in a few of the details because much of the development of V Bd's in the area of equalization processes will proceed along very similar lines. In might seem from the above discussion that one could not obtain a V Bd on the parameters which characterize time-dependent scattering problems, where the relevant operator is

$$A(t) = \left(\frac{-i}{\hbar}\right)H(t) - \frac{\partial}{\partial t},$$

for even though $H(t)$ is bounded from below, $\partial/\partial t$ is not bounded from below (nor from above). However, time-dependent scattering problems have a simplifying feature which time-independent scattering problems do not, namely, the wave functions do not contain plane waves but are localized and can therefore be normalized. As a consequence, the relevant operator is not the singular time-independent resolvent $(H-E)^{-1}$, or in coordinate space $G(E;\mathbf{r},\mathbf{r}')$, with E the total energy, but the nonsingular $U(t,t')$ or, in coordinate space, $G(\mathbf{r},t;\mathbf{r}',t')$. [For the time-independent case, $U(t,t')$ can be written as $U(t-t')$, and $G(E;\mathbf{r},\mathbf{r}')$ is the Fourier transform of $G(t-t';\mathbf{r},\mathbf{r}') = \langle \mathbf{r}|U(t-t')|\mathbf{r}'\rangle$. Note too that $G(E;\mathbf{r},\mathbf{r}')$ is not singular for E off the real axis. While physical problems normally involve E on the real axis, amplitudes there can be obtained by analytic continuation from amplitudes off the real axis, and since $G(E;\mathbf{r},\mathbf{r}')$ is not then singular one can obtain V Bd's for amplitudes off the real energy axis.[7]] From the fact that the scattering wave function satisfies

$$\langle \psi(t)|\psi(t)\rangle = \langle \psi(t')|\psi(t')\rangle,$$

it follows immediately for all finite t and t', on using Eq. (1.3), that

$$U^{\dagger}(t,t')U(t,t') = 1. \tag{1.10}$$

That Eq. (1.10) remains valid as $t \sim +\infty$ and $t' \sim -\infty$ is anything but obvious but is well known to be true.[8] On the other hand, while the relationship

$$U(t,t')U^{\dagger}(t,t') = 1$$

is also valid for finite times, it need *not* be valid for infinite times so that U need not be unitary. The isometric property, Eq. (1.10) extended to include infinite times, will however be sufficient for our purposes. One can write[9,10]

$$U(t,\tau) = U_v(t,\tau) + \Delta U(t,\tau), \tag{1.11}$$

where the variational estimate U_v is calculable but where ΔU is the formal expression—it contains U—

$$\Delta U(t,\tau) = \int_\tau^t ds \int_s^t ds' [A(s')\bar{U}(s',t)]^\dagger U(s',s)[A(s)\bar{U}(s,\tau)], \tag{1.12}$$

where $\bar{U}$ is a trial estimate of U. (One can replace one of the two $\bar{U}$'s in the above expression by a second independent trial estimate.[9,10]) $A\bar{U}$ is of first order since $AU = 0$ and since $\bar{U} - U$ is by definition of first order, and it follows that ΔU is of second order. The transition amplitude $a_{fi}(t,\tau)$ for going from an initial state $\psi_i(\tau)$ to a final state $\psi_f(t)$ in the presence of the known external time-dependent potential is

$$a_{fi}(t,\tau) = \langle \psi_f(t) | U(t,\tau) | \psi_i(\tau) \rangle,$$

where $\psi_i(t)$ and $\psi_f(t)$ are the (known) exact time-dependent solutions in the absence of the external potential. Since the amplitude of interest is normally

$$a_{fi} \equiv \lim_{t\to\infty} \lim_{\tau\to-\infty} a_{fi}(t,\tau),$$

the determination of both upper and lower V Bd's on a_{fi} reduces to the determination of just an upper bound (one need *not* obtain a V Bd) on the absolute magnitude of

$$\Delta a_{fi}(t,\tau) \equiv \langle \psi_f(t) | \Delta U(t,\tau) | \psi_i(\tau) \rangle. \tag{1.13}$$

We now proceed to obtain such a bound. With δ representing i or f, we have

$$\psi_\delta(t') = U(t',t'')\psi_\delta(t''),$$

and it is therefore natural to introduce the notation

$$\bar{\psi}_\delta(t') \equiv \bar{U}(t',t'')\psi_\delta(t''). \tag{1.14}$$

Inserting ΔU defined by Eq. (1.12) into Eq. (1.13) and using Eq. (1.14), we then have

$$\Delta a_{fi}(t,\tau) = \int_\tau^t ds \int_s^t ds' \langle A(s')\bar{\psi}_f(s') | U(s',s) | A(s)\bar{\psi}_i(s) \rangle.$$

We now use the Schwarz inequality and the isometry of U to obtain the sought-for bound,

$$|\Delta a_{fi}(t,\tau)| \leqslant \int_\tau^t ds \int_s^t ds' N_f(s') N_i(s),$$

where

$$N_\delta(t') \equiv \langle A(t')\bar{\psi}_\delta(t') | A(t')\bar{\psi}_\delta(t') \rangle^{1/2}.$$

Note that this expression for N_δ is explicit, not formal. Since N_f and N_i are each of first order, the bound on Δa_{fi} is of second order, as is required if we are to obtain a V Bd. (See the remark due to Percival in the Note added in proof in Ref. 9.) With Δa_{fi}^{bd} the bound on $|\Delta a_{fi}(t,\tau)|$ in the limit of infinite times, and with a_{fiv} the variational estimate of a_{fi} obtained by taking the matrix element of U_v, we have

$$a_{fiv} - \Delta a_{fi}^{bd} \leqslant a_{fi} \leqslant a_{fiv} + \Delta a_{fi}^{bd}.$$

The V Bd's are of course useless unless they lie between 0 and 1.

[We note parenthetically that the V Bd on a_{fi} has not thus far proved to be terribly useful in numerical estimates. It gave a rather poor estimate of the probability of ionization in a proton-hydrogen atom collision, $p + H \to p + p + e^-$, but this is a very difficult problem.[11] The approach did prove to be very useful in an analysis at asymptotically high incident velocities of the probability for the transfer of a light particle between two heavy particles. Assuming all interactions were of a certain class of short-ranged potentials, and assuming the validity of the impact-parameter approximation, the external potential seen by the light particle is a known time-dependent interaction and the problem reduces to a transition amplitude analysis. It was shown[10] that the *second*-Born contribution dominates. (The analogous result for the important charge-transfer problem $p + H \to H + p$, where the interactions are not short ranged but Coulombic, has not been obtained.)]

There are two main differences between the development of V Bd's in time-dependent scattering problems and in equalization processes. The first difference is that in the latter case we are concerned with the development of a V Bd on a function, such that $T(\mathbf{r},t)$, rather than on a matrix element a_{fi}. We will see that this difference is by no means unimportant, but the second and essential difference is the fact that the functions in equalization processes are real and of immediate physical interest, while the functions in time-dependent scattering theory are neither real nor of direct physical interest but rather are complex amplitudes [from which, of course, one can calculate (real) transition rates]. Related to this second difference is the fact that one does not have conservation of probability and the consequent isometry of U for equalization processes. We can have conservation of energy in the latter case—we do not have energy conservation for time-dependent external potentials—but the relevant property for our purposes is the "maximum principle," the monotonically nonincreasing behavior with time of the maximum value over space of $T(\mathbf{r},t)$—for the case of heat conduction—presumably the (partial) origin of the terminology equalization process. This will enable us to define a norm of U which will give the desired V Bd.

II. EQUALIZATION PROCESSES

A. The maximum principle

In our analysis of heat conduction we restrict our attention to a solid or to a fluid at rest, within which no heat sources are present, and we allow the thermal coefficients to be functions of $\mathbf{r}$ and t but not of T. Appendix A treats the case for which heat sources are present. The (linear homogeneous) differential heat conduction equation is then given by Eq. (1.2), with the identifications $F(t) = T(t)$ and

$$K(\mathbf{r},t) = \lambda(\mathbf{r},t)\nabla\cdot(k(\mathbf{r},t)\nabla)$$

in Eq. (1.1a), where k is the thermal conductivity and $\lambda = C^{-1}$ with C the heat capacity per unit volume. The temperature T is in degrees Kelvin and is therefore positive. For diffusion in a medium at rest and at uniform temperature one has

$$A(t)u(t) = 0,$$

where

$$K(\mathbf{r},t) = \nabla\cdot(\mathscr{D}(\mathbf{r},t)\nabla),$$

with $\mathscr{D}$ the diffusivity and u the concentration. There are similar equations for the other equalization processes. Our discussion will be in the context of heat conduction but can be transcribed immediately to the other processes.

The operator A is said to be uniformly parabolic in a domain of the four-dimensional $\mathbf{r},t$ space if there exists a positive constant μ for which $\lambda(\mathbf{r},t)\mathbf{k}(\mathbf{r},t)\geqslant\mu$ for all points in that domain. The maximum principle states[12] that if in a domain λk and $\lambda\nabla k$ are bounded and A is uniformly parabolic, conditions which we will assume to be satisfied, then for any solution T the maximum must occur either at the initial time or on the boundary.

A linear operator W is said to be bounded in a space if there exists a finite number R such that

$$\|Wq\|\leqslant R\,\|q\|$$

for any function q in that space, where the norm $\|q\|$ of q in a domain is defined as the maximum value of $|q|$ in that domain. The norm $\|W\|$ of W can be defined in the space of q by its supremum,

$$\|W\| = \sup_{\|q\|\neq 0}\frac{\|Wq\|}{\|q\|}.$$

Identifying W with U, t and t' fixed,

$$\|U(t,t')q(t')\|\leqslant\|q(t')\| = q_{\max}(t'), \tag{2.1}$$

where $q_{\max}(t')$ is the maximum value of $|q(\mathbf{r},t')|$ over all allowed values of $\mathbf{r}$; rewriting (2.1) as

$$\operatorname*{Max}_{\mathbf{r}}\left|\int\langle\mathbf{r}|U(t,t')|\mathbf{r}'\rangle d\mathbf{r}'\langle\mathbf{r}'|q(t')\rangle\right|\leqslant\operatorname*{Max}_{\mathbf{r}}|\langle\mathbf{r}|q(t')\rangle|,$$

we have, using Eq. (1.9),

$$\left|\int d\mathbf{r}'G(\mathbf{r},t;\mathbf{r}',t')q(\mathbf{r}',t')\right|\leqslant\operatorname*{Max}_{\mathbf{r}}|q(\mathbf{r},t')|. \tag{2.2}$$

Note that the bound is independent of the later time t. The inequality (2.1), or (2.2), is the statement of the maximum principle and will play the role for equalization processes which was played by the isometry of U for time-dependent scattering. To use the maximum principle it will be necessary to obtain a variational identity for U. This will be very similar to that given in Eq. (1.11) for time-dependent scattering.

B. The variational identity for U

We here restrict ourselves to homogeneous boundary conditions. (The more general case is treated in Appendix A.) By Duhamel's principle,[13,14] it follows from Eqs. (1.6) and (1.4), where A is defined by Eqs. (1.1a) and (1.1b), that

$$f(t) = U(t,\tau)f(\tau) - \int_\tau^t U(t,s)A(s)f(s)ds$$

for any $f(t)$. Choosing $f(t')$ for $t'=t$, τ or s to be an estimate $\overline{U}(t',\tau)$ of $U(t',\tau)$, subject to

$$\overline{U}(\tau,\tau) = 1, \tag{2.3}$$

we have an integral equation for $U(t,\tau)$,

$$U(t,\tau) = \overline{U}(t,\tau) + \int_\tau^t U(t,s)A(s)\overline{U}(s,\tau)ds. \tag{2.4}$$

(In Appendix B, we provide a proof of this equation which some physicists may find more "physical" than the usual proofs of Duhamel's principle.) Equation (2.4), which is an identity for arbitrary $\overline{U}$, was also the starting point for the time-dependent scattering problem. If we approximate U in the integrand by $\overline{U}$ we get a Lippmann–Schwinger-type VP for U. If in Eq. (2.4) we replace τ by s and s by s', we obtain an integral equation for $U(t,s)$. The insertion of this expression for $U(t,s)$ into Eq. (2.4) gives the variational identity

$$U(t,\tau) = U_v(t,\tau) + \Delta U(t,\tau), \tag{2.5}$$

where

$$U_v(t,\tau) = \overline{U}(t,\tau) + \int_\tau^t \overline{U}(t,s)A(s)\overline{U}(s,\tau)ds \tag{2.6}$$

is an explicit VP for $U(t,\tau)$, while the (formal) second-order error term

$$\Delta U(t,\tau) = \int_\tau^t ds\int_s^t ds'\,U(t,s')A(s')\overline{U}(s',s)A(s)\overline{U}(s,\tau) \tag{2.7}$$

contains the unknown U. The elimination of U by means of the maximum principle gives

$$\|\Delta U(t,\tau)\|\leqslant\int_\tau^t ds\int_s^t ds'\|A(s')\overline{U}(s',s)A(s)\overline{U}(s,\tau)\|. \tag{2.8}$$

We now let U as given by Eqs. (2.5), (2.6), and (2.7) operate on $T(\tau)$ and we introduce the zeroth-order estimate

$$T_0(t) = \overline{U}(t,\tau)T(\tau), \tag{2.9}$$

the first-order correction

$$T_1(t) = \int_\tau^t \overline{U}(t,s)A(s)\overline{U}(s,\tau)T(\tau)ds = \int_\tau^t \overline{U}(t,s)A(s)T_0(s)ds, \tag{2.10}$$

the variational estimate

$$T_v(t) = T_0(t) + T_1(t), \tag{2.11}$$

and the second-order error term

$$\Delta T(t)\equiv\Delta U(t,\tau)T(\tau) = \int_\tau^t ds\int_s^t ds'\,U(t,s')A(s')\overline{U}(s',s)A(s)T_0(s). \tag{2.12a}$$

On interchanging the order of integration and using Eq. (2.10), it follows that

$$\Delta T(t) = \int_\tau^t ds\;U(t,s)A(s)T_1(s), \tag{2.12b}$$

where we changed variables from s' to s, and therefore that

$$|\Delta T(t)|\leqslant\Delta T^{\mathrm{Bd}}(t)\equiv\int_\tau^t ds\|A(s)T_1(s)\|. \tag{2.13}$$

In the coordinate representation, we then have

$$T_v(\mathbf{r},t) - \Delta T^{\mathrm{Bd}}(t)\leqslant T(\mathbf{r},t)\leqslant T_v(\mathbf{r},t) + \Delta T^{\mathrm{Bd}}(t), \tag{2.14}$$

where, using Eqs. (2.11), (2.9), (2.10), (2.12b), (2.13), (1.9), and (1.1b),

$$T_v(\mathbf{r},t) = T_0(\mathbf{r},t) + T_1(\mathbf{r},t), \tag{2.15}$$

$$T_0(\mathbf{r},t) = \int\overline{G}(\mathbf{r},t;\mathbf{r}',\tau)T(\mathbf{r}',\tau)d\mathbf{r}', \tag{2.16}$$

$$T_1(\mathbf{r},t) = \int_\tau^t ds \int d\mathbf{r}'\overline{G}(\mathbf{r},t;\mathbf{r}',s)A(\mathbf{r}',s)T_0(\mathbf{r}',s), \tag{2.17}$$

$$\Delta T(\mathbf{r},t) = \int_\tau^t ds \int d\mathbf{r}'G(\mathbf{r},t;\mathbf{r}',s)A(\mathbf{r}',s)T_1(\mathbf{r}',s), \tag{2.18}$$

$$\Delta T^{\mathrm{Bd}}(t) = \int_\tau^t ds\, \underset{\mathbf{r}}{\mathrm{Max}}|A(\mathbf{r},s)T_1(\mathbf{r},s)|. \tag{2.19}$$

All spatial integrations are over the same volume V, and $\overline{G}$ is an estimate of G.

Before proceeding to a discussion of the choice of $\overline{G}$, we comment on the differences between the procedure just used and that used in time-dependent scattering. Because of the isometry of U in the one case as opposed to the property defined by Eq. (2.1) (which follows from the maximum principle) in the other, the quantities to be calculated are here much simpler; it is far simpler to bound a function than to use the Schwarz inequality to bound an integral. In the scattering problem, we had to manipulate the variational identity into a form in which $\overline{U}$, U, and $\overline{U}$ appeared in that order, which necessitated the introduction of adjoints, for had we used the form which involved the order, U, $\overline{U}$, and $\overline{U}$, as in the present Eq. (2.7), the subsequent use of the Schwarz inequality would have led to a horrendously complicated bound.

As in the scattering problem,[9] we can also obtain a much simpler but nonvariational bound on $T(\mathbf{r},t)$. Operating on $T(\tau)$ with $U(t,\tau)$ of Eq. (2.4), we have, using Eq. (2.9),

$$T(t) = T_0(t) + \int_\tau^t U(t,s)A(s)T_0(s)ds,$$

so that

$$T_0(\mathbf{r},t) - \int_\tau^t T^{(1)}(s)ds \leqslant T(\mathbf{r},t) \leqslant T_0(\mathbf{r},t) + \int_\tau^t T^{(1)}(s)ds, \tag{2.20}$$

where $T^{(1)}(s)$ is the maximum over the range of $\mathbf{r}$ of $A(\mathbf{r},s)T_0(\mathbf{r},s)$. A nonvariational bound was also obtained by Protter and Weinberger.[12] They too used the maximum principle, but otherwise their approach was rather different. See also Eidel'man.[15]

III. THE CHOICE OF THE TRIAL GREEN'S FUNCTION

If the upper and lower V Bd's and $T(\mathbf{r},t)$ are to be useful, it will be necessary to choose a moderately accurate but reasonably simple form for $\overline{G}$. An obvious possibility is the choice $\overline{G}$ equal to an exact solution of a similar problem. Though it will be seen that some caution must be exercised in making such a choice, it will be useful to begin with a discussion of some known exact solutions.

If the thermal properties λ and k of the medium are constant, exact solutions for steady and nonsteady heat conduction exist for many simple geometries, domains, and boundary conditions.[16,17] The results were obtained by classical methods; separation of variables, Laplace, Fourier and Hankel transforms, and Green's functions. Many problems in fluid dynamics of viscous liquids, heat conduction, convective diffusion, etc., reduce with suitable substitutions to differential equations of heat conduction with variable coefficients. For these cases, exact solutions can be found in a limited number of special situations. Solutions have been obtained by means of Laplace transforms for simple boundary conditions for one-dimensional problems with k and λ simple functions of the coordinate z, namely,

$$k = k_0(1+bz)^n, \quad \lambda^{-1} = c_0(1+\bar{b}z)^m,$$

and

$$k = k_0 z^n, \quad \lambda^{-1} = c_0 z^m,$$

where k_0, c_0, b, $\bar{b}$, n, and m are constants.

For $\lambda(\mathbf{r},t) = \lambda_0$ and $k(\mathbf{r},t) = k_0$, we set

$$\alpha = \lambda_0 k_0 \tag{3.1}$$

and introduce the Green's function G_α defined by

$$A_\alpha(\mathbf{r},t)G_\alpha(\mathbf{r},t;\mathbf{r}',t') = 0, \quad t > t', \tag{3.2}$$

where

$$A_\alpha(\mathbf{r},t) \equiv \alpha\nabla^2 - \frac{\partial}{\partial t} \tag{3.3}$$

and

$$G_\alpha(\mathbf{r},t;\mathbf{r}',t) = \delta(\mathbf{r}-\mathbf{r}'). \tag{3.4}$$

In the three-dimensional infinite case, $0 \leqslant r \leqslant \infty$, G is given exactly for $t > t'$ by

$$G_\alpha(\mathbf{r},t;\mathbf{r}',t') = [4\pi\alpha(t-t')]^{-3/2}\exp\left(\frac{-|\mathbf{r}-\mathbf{r}'|^2}{4\alpha(t-t')}\right). \tag{3.5}$$

There are many cases for which it might seem simple and natural to choose $\overline{G}$ to be G_α, or G_α plus some slowly varying function. This can lead to difficulties, however, since the integrands of V Bd's on T contain factors of the form $A\overline{G}$, and $A\overline{G}$ will often behave more singularly than $\overline{G}$ does; this can cause serious difficulties when taking norms. Consider, for example, the case $\lambda(\mathbf{r},t) = \lambda_0, k(\mathbf{r},t) = k_0 r$. Then one readily finds that $A(\mathbf{r},t)G_\alpha(\mathbf{r},t;\mathbf{r}',t')$ has a term proportional to $G_\alpha/(t-t')$ and is clearly therefore more singular at $t = t'$ than is G_α.

We will discuss the possibility of bypassing this difficulty in Sec. IV. For the remainder of Sec. III, we will consider the interesting but limited cases for which no difficulty arises.

From Eq. (2.17), we have

$$T_1(\mathbf{r},s) = \int_\tau^s ds' \int d\mathbf{r}'\overline{G}(\mathbf{r},s;\mathbf{r}',s')B(\mathbf{r}',s'), \tag{3.6}$$

where

$$B(\mathbf{r}',s') \equiv A(\mathbf{r}',s')T_0(\mathbf{r}',s'). \tag{3.7}$$

Assuming that

$$k(\mathbf{r},s) = k_0 = \text{const}$$

and

$$\lambda(\mathbf{r},s) = \lambda(\mathbf{r}),$$

one has, using Eq. (1.1b) and the first equation in Sec. II,

$$\begin{aligned} A(\mathbf{r},s) &= k_0\lambda(\mathbf{r})\nabla^2 - \frac{\partial}{\partial s} \\ &= \left(\frac{k_0}{\alpha}\right)\lambda(\mathbf{r})\left(\alpha\nabla^2 - \frac{\partial}{\partial s} + \frac{\partial}{\partial s}\right) - \frac{\partial}{\partial s}. \end{aligned} \tag{3.8}$$

With the aid of Eqs. (3.1) and (3.3), we can also write

$$A(\mathbf{r},s) = \frac{\lambda(\mathbf{r})}{\lambda_0}A_\alpha(\mathbf{r},s) + \frac{[\lambda(\mathbf{r})-\lambda_0]}{\lambda_0}\frac{\partial}{\partial s}, \tag{3.9}$$

where λ_0 is a constant of arbitrary choice. We will use the second form of $A(\mathbf{r},s)$ when operating on $G(\mathbf{r},s;\mathbf{r}',s')$, and the first form of $A(\mathbf{r},s)$ when operating on the upper limit s in Eq. (3.6) for T_1. Equation (2.19) can be written as

$$\Delta T^{\rm Bd}(t) = \int_\tau^t ds \operatorname*{Max}_{\mathbf{r}} \Big| \int_\tau^s ds' \int d\mathbf{r}' \times\Big(\frac{\lambda(\mathbf{r})}{\lambda_0} A_\alpha(\mathbf{r},s) + \frac{\lambda(\mathbf{r})-\lambda_0}{\lambda_0}\frac{\partial}{\partial s}\Big) \times \overline{G}(\mathbf{r},s;\mathbf{r}',s') B(\mathbf{r}',s')\Big| - \int_\tau^t ds \operatorname*{Max}_{\mathbf{r}} \Big| \int d\mathbf{r}' \overline{G}(\mathbf{r},s;\mathbf{r}',s) B(\mathbf{r}',s)\Big|. \quad (3.10)$$

Letting

$$\overline{G} = G_\alpha,$$

and using Eq. (3.4) with t replaced by s, and

$$A_\alpha(\mathbf{r},s) G_\alpha(\mathbf{r},s;\mathbf{r}',s') = -\delta(\mathbf{r}-\mathbf{r}')\delta(s-s'), \quad (3.11)$$

which follows from Eqs. (3.2) and (3.4), one obtains

$$\Delta T^{\rm Bd}(t) = \int_\tau^t ds \operatorname*{Max}_{\mathbf{r}} \Big| \frac{\lambda(\mathbf{r})}{\lambda_0} \int_\tau^{s+o} ds' \int d\mathbf{r}' \delta(\mathbf{r}-\mathbf{r}')\delta(s-s') B(\mathbf{r}',s')\Big| + \int_\tau^t ds' \operatorname*{Max}_{\mathbf{r}} \Big| \frac{\lambda(\mathbf{r})-\lambda_0}{\lambda_0} \int d\mathbf{r}' \int_{s'}^t ds \frac{\partial \overline{G}_\alpha(\mathbf{r},s;\mathbf{r}',s')}{\partial s} B(\mathbf{r}',s')\Big| + \int_\tau^t ds \operatorname*{Max}_{\mathbf{r}} |B(\mathbf{r},s)|. \quad (3.12)$$

In the first term the limit $s+o$ indicates that the upper limit s is to be approached from above and in the second term we used

$$\int_\tau^t ds \int_\tau^s ds' = \int_\tau^t ds' \int_{s'}^t ds.$$

We now integrate by parts to obtain

$$\int d\mathbf{r}' \int_{s'}^t ds \frac{\partial G_\alpha(\mathbf{r},s;\mathbf{r}',s')}{\partial s} B(\mathbf{r}',s') = \int d\mathbf{r}' [G_\alpha(\mathbf{r},t;\mathbf{r}',s') - G_\alpha(\mathbf{r},s';\mathbf{r}',s')] B(\mathbf{r}',s') = \int G_\alpha(\mathbf{r},t;\mathbf{r}',s') B(\mathbf{r}',s') d\mathbf{r}' - B(\mathbf{r},s). \quad (3.13)$$

Thus, we can write

$$\Delta T^{\rm Bd}(t) = \int_\tau^t ds \operatorname*{Max}_{\mathbf{r}} \Big| \frac{\lambda(\mathbf{r})}{\lambda_0} B(\mathbf{r},s)\Big| + \int_\tau^t ds \operatorname*{Max}_{\mathbf{r}} \Big| \frac{\lambda(\mathbf{r})-\lambda_0}{\lambda_0} \times\Big[\int G_\alpha(\mathbf{r},t;\mathbf{r}',s) B(\mathbf{r}',s) d\mathbf{r}' - B(\mathbf{r},s)\Big]\Big| + \int_\tau^t ds \operatorname*{Max}_{\mathbf{r}} |B(\mathbf{r},s)|. \quad (3.14)$$

Note that

$$\int_\tau^t ds \operatorname*{Max}_{\mathbf{r}} \Big| \frac{\lambda(\mathbf{r})-\lambda_0}{\lambda_0} \Big(\int G_\alpha(\mathbf{r},t;\mathbf{r}',s) B(\mathbf{r}',s) d\mathbf{r}' - B(\mathbf{r},s)\Big)\Big| \leqslant \int_\tau^t ds \Big[\operatorname*{Max}_{\mathbf{r}} \Big|\frac{\lambda(\mathbf{r})-\lambda_0}{\lambda_0}\Big| \operatorname*{Max}_{\mathbf{r}} |B(\mathbf{r},s)| + \operatorname*{Max}_{\mathbf{r}} \Big| \frac{\lambda(\mathbf{r})-\lambda_0}{\lambda_0} B(\mathbf{r},s)\Big|\Big]. \quad (3.15)$$

Equation (3.14), in conjunction with Eq. (3.15), is a useful result when $\mathrm{Max}_{\mathbf{r}} |(\lambda(\mathbf{r})-\lambda_0)/\lambda_0|$ is finite.

Another case for which no difficulty arises, with $\lambda(\mathbf{r})$ now a constant, λ_0, and $k(\mathbf{r})$ varying, is

$$k(\mathbf{r}) = k_0 + k_1 r^n, \quad n \geqslant 2, \quad (3.16)$$

with k_0 and k_1 constants. One then has for $\overline{G} = G_\alpha$ and $t > t'$,

$$AG_\alpha = \Big(\alpha\nabla^2 + \lambda_0 k_1 \nabla\cdot(r^n\nabla) - \frac{\partial}{\partial t}\Big) G_\alpha = \lambda_0 k_1 \nabla\cdot(r^n \nabla G_\alpha),$$

and no singularity higher than the one possessed by G_α is introduced.

IV. AVOIDANCE OF A SINGULARITY BY A TRANSFORMATION OF COORDINATES

We now show that there are many important situations for which the difficulty noted below Eq. (3.5) can be avoided. We assume, as is normally the case, that λ and k are time independent—but we make the further more restrictive assumption that we are concerned with an infinite one-dimensional problem, that is,

$$\lambda(\mathbf{r},t) = \lambda(z), \quad k(\mathbf{r},t) = k(z), \quad -\infty \leqslant z \leqslant \infty,$$

or with an infinite spherically symmetric problem, that is,

$$\lambda(\mathbf{r},t) = \lambda(r), \quad k(\mathbf{r},t) = k(r), \quad 0 \leqslant r \leqslant \infty.$$

[The proof goes through under the slightly less restrictive condition $\lambda(\mathbf{r},t) = \lambda(z)M(t)$ and $k(\mathbf{r},t) = k(z)M(t)$, or $\lambda(\mathbf{r},t) = \lambda(r)M(t)$ and $k(\mathbf{r},t) = k(r)M(t)$, that is, we allow an arbitrary time dependence but the time dependence must be the same for each function.] The operator $K(\mathbf{r},t)$ reduces in the one-dimensional situation to

$$K(z) = \lambda(z) \frac{\partial}{\partial z} k(z) \frac{\partial}{\partial z}.$$

We make the change of variables from z to Z, with Z as yet unspecified, so that

$$\frac{\partial}{\partial z} = \frac{dZ}{dz}\frac{\partial}{\partial Z},$$

and write

$$\lambda[z(Z)] = \tilde\lambda(Z), \quad k[z(Z)] = \tilde k(Z), \quad \frac{dZ}{dz} = \Big(\frac{dz}{dZ}\Big)^{-1}. \quad (4.1)$$

We then have

$$K(z) = \tilde\lambda \tilde k \Big(\frac{dZ}{dz}\Big)^2 \frac{\partial^2}{\partial Z^2} + \tilde\lambda\Big[\tilde k'\Big(\frac{dZ}{dz}\Big)^2 + \tilde k \frac{dZ}{dz}\frac{d}{dZ}\Big(\frac{dz}{dZ}\Big)^{-1}\Big]\frac{\partial}{\partial Z} \equiv \widetilde K(Z), \quad (4.2)$$

where

$$\tilde k' \equiv \frac{\partial \tilde k}{\partial Z}.$$

It is useful to have the coefficient of the second partial equal to a constant, since $\widetilde{K}(Z)$ then contains the operator which appears when λ and k are constants. We therefore make the choice

$$\frac{dZ}{dz} = \left(\frac{\lambda(z)k(z)}{\lambda_0 k_0}\right)^{-1/2} = \alpha^{1/2}[\lambda(z)k(z)]^{-1/2}, \qquad (4.3)$$

where λ_0 and k_0 are characteristic values of $\lambda(z)$ and $k(z)$, respectively. Z is given explicitly, ignoring a constant of integration, by

$$Z(z) = \alpha^{1/2}\int_0^z [\lambda(z')k(z')]^{-1/2}\,dz'. \qquad (4.4)$$

With $\tilde{\lambda}' = \equiv \partial\tilde{\lambda}/\partial Z$, the substitution of Eq. (4.3) into (4.2) leads to

$$\widetilde{K}(Z) = \alpha\left[\frac{\partial^2}{\partial Z^2} + \frac{1}{2}\left(\frac{\tilde{k}'}{\tilde{k}} - \frac{\tilde{\lambda}'}{\tilde{\lambda}}\right)\frac{\partial}{\partial Z}\right]. \qquad (4.5)$$

For our trial Green's function $\overline{G}(z,t;z',t') = \overline{\widetilde{G}}(Z,t;Z',t')$, we choose

$$\overline{\widetilde{G}}(Z,t;Z',t') = \frac{G_\alpha(Z,t;Z',t')}{g(Z)}\,h(Z') + \Gamma, \qquad (4.6)$$

where Γ is a smooth function which vanishes at $t' = t$ and which will therefore cause no difficulty; $g(Z)$ and $h(Z')$ are thus far arbitrary and will be chosen to counterbalance the higher singularities introduced on operating on G_α with $\widetilde{K}$. Ignoring the Γ term, we have, with

$A[z(Z),t] \equiv \widetilde{A}(Z,t)$,

$$\begin{aligned}\widetilde{A}(Z,t)(G_\alpha gh) &= \left(\widetilde{K}(Z) - \frac{\partial}{\partial t}\right)(G_\alpha gh)\\ &= \alpha h(Z')\left[2\frac{\partial G_\alpha}{\partial Z}\frac{\partial g}{\partial Z} + G_\alpha\frac{\partial^2 g}{\partial Z^2}\right.\\ &\quad \left. + \frac{1}{2}\left(\frac{\tilde{k}'}{\tilde{k}} - \frac{\tilde{\lambda}'}{\tilde{\lambda}}\right)\left(G_\alpha\frac{\partial g}{\partial Z} + \frac{\partial G_\alpha}{\partial Z}g\right)\right],\end{aligned}$$

where we have used the one-dimensional form of Eq. (3.2),

$$\left(\alpha\frac{\partial^2}{\partial Z^2} - \frac{\partial}{\partial t}\right)G_\alpha = 0.$$

The terms which involve G_α itself need not concern us; it is the terms in $\partial G_\alpha/\partial Z$ which can cause difficulties. With $g' \equiv \partial g/\partial Z$, the coefficient of $\partial G_\alpha/\partial Z$ is

$$2\alpha g(Z)h(Z')\left[\frac{g'}{g} + \frac{1}{4}\left(\frac{\tilde{k}'}{\tilde{k}} - \frac{\tilde{\lambda}'}{\tilde{\lambda}}\right)\right],$$

which vanishes if we choose $g(Z)$ to be

$$g(Z) = \left(\frac{\tilde{\lambda}(Z)}{\lambda_0}\frac{k_0}{\tilde{k}(Z)}\right)^{1/4}. \qquad (4.7)$$

The function $h(Z')$ will be chosen to have $\overline{G}(z,t;z',t')$ satisfy the initial condition

$$\begin{aligned}\delta(z - z') &= \overline{G}(z,t;z',t) \equiv g(Z)G_\alpha(Z,t;Z',t)h(Z')\\ &= g(Z)\delta(Z - Z')h(Z') \end{aligned} \qquad (4.8)$$

since $\langle\zeta|\overline{U}(t,t')|\zeta'\rangle = \overline{G}(\zeta,t;\zeta',t')$ for $\zeta = z$ and Z, and, from Eq. (2.3), $\overline{U}(t,t) = 1$. Restricting our attention to transformations which are single valued, we have

$$\delta[Z(z) - Z(z')] = \delta\left(\frac{dZ}{dz}(z - z')\right) = \frac{dz}{dZ}\delta(z - z').$$

Comparison with Eq. (4.8) gives

$$h(Z) = \left(g(Z)\frac{dz}{dZ}\right)^{-1}. \qquad (4.9)$$

We are at liberty to introduce $\overline{G}(z,t;z',t')$ in a form analogous to that used in (4.8), with $h(Z)$ chosen as in (4.9). We then have, using Eq. (4.7),

$$\begin{aligned}\overline{G}&(z,t;z',t')\\ &= \left(\frac{\tilde{\lambda}(Z)}{\tilde{k}(Z)}\right)^{1/4} G_\alpha(Z,t;Z',t')\\ &\quad\times\left(\frac{\tilde{k}(Z')}{\tilde{\lambda}(Z')}\right)^{1/4}\times\left(\frac{dz'}{dZ'}\right)^{-1}.\end{aligned} \qquad (4.10)$$

To proceed further, it would seem simplest not to transform back to the variable z but, since the most complicated function present will normally be $\overline{G}$, which is simplest when expressed as a function of Z, to remain in Z space. Thus, in performing an integration over z, one would change the variable from z to Z.

For example, $T_1(z,t)$, defined by Eq. (2.17), is transformed into

$$\begin{aligned}T_1(Z,t) &= \int_\tau^t dt'\int_{-\infty}^{\infty} dZ'\\ &\quad\times\int_{-\infty}^{\infty} dZ''g(Z)G_\alpha(Z,t;Z',t')g^{-1}(Z')\\ &\quad\times\widetilde{A}(Z',t')g(Z')G_\alpha(Z',t',Z'',\tau)g^{-1}(Z'')\widetilde{T}(Z'',\tau),\end{aligned} \qquad (4.11)$$

where

$$\widetilde{T}(Z,\tau) \equiv T[z(Z),\tau].$$

V. CONCLUSION

To gain some insight into the difficulties involved in evaluating the integrals and determining the norms, and to obtain some feeling for the quality of the bounds to be obtained, we applied the method to the problem of a one-dimensional infinite solid with k a constant k_0 and with

$$\lambda(z)k = [\lambda_0 + \Delta\lambda(z)]k_0 = [\alpha + \beta(z^2/l^2)e^{-z^2/l^2}], \qquad (5.1)$$

where λ_0, α, β, and l are constants. The initial temperature distribution was chosen to be—with τ set equal to zero—

$$T_{\text{in}}(z,o) = T_I e^{-z^2/L^2}, \qquad (5.2)$$

with T_I and L constants. As the trial Green's function we chose the one-dimensional version of Eq. (3.5), namely,

$$\begin{aligned}\overline{G} &= G_\alpha(z,t;z',t') = [4\pi\alpha(t - t')]^{-1/2}\\ &\quad\times\exp\{-(z - z')^2/[4\alpha(t - t')]\}.\end{aligned} \qquad (5.3)$$

From Eqs. (2.16), (2.17), (3.14), and (3.15) in conjunction with Eqs. (5.1)–(5.3), we obtain

$$T_0(z,t) = T_I g(t)\exp[-(z/L)^2 g^2(t)], \qquad (5.4)$$

$$\begin{aligned}T_1(z,t) &= \pi^{-1/2}T_I\left(\frac{2\beta}{l^2}\right)\int_0^t ds\int_{-\infty}^{\infty} dz' q e^{-q^2(z-z')^2}\\ &\quad\times e^{-[(L/l)^2 + g^2(s)](z'/L)^2}\\ &\quad\times[2(z'/L)^4 g^5(s) - (z'/L)^2 g^3(s)],\end{aligned} \qquad (5.5)$$

and

$$\Delta T^{\mathrm{Bd}}(t) = 1.66\left(\frac{(\beta/l^2)}{(\alpha/L^2)}\right) \times\left[1 + 1.5e^{-1}\left(\frac{\beta}{\alpha}\right)\right][g^{-1}(t) - 1\}T_I, \qquad (5.6\mathrm{a})$$

where we have introduced

$$g(t) \equiv [1 + 4(\alpha/L^2)t\,]^{-1/2}, \quad q \equiv [4\alpha(t-s)]^{-1/2}. \qquad (5.6\mathrm{b})$$

The z' integration in Eq. (5.5) is easily performed, but the complicated result is not shown. The subsequent s integration can be performed analytically for $z = 0$.

We come now to the choice of reasonable values for the parameters which appear in Eq. (5.6a), namely, β/l^2, α/L^2, and β/α. The problem is uninteresting if the variation of $\lambda(z)$ is insignificant; we choose the maximum fractional variation of $\Delta\lambda(z)/\lambda_0$ to be $1/e$. It follows from Eq. (5.1) that $\beta/\alpha = 1$. The value of L is arbitrary and simply sets a scale of length. If L/l were very small, there would be negligible variation of $\lambda(z)k$ in the region over which T_{in} is not very small, while if L/l were very large the variation of $\lambda(z)k$ would be over a very small region, and we chose $l = L$. For $l = L$, the coefficient 1.36 in (5.6a) could have been replaced by 1.14. We now have, since $\beta = \alpha$,

$$\Delta T^{\mathrm{Bd}}(t)/T_I = 1.77[g^{-1}(t) - 1] = 1.77[(1 + 4\alpha L^{-2}t)^{1/2} - 1]. \qquad (5.7)$$

For $\beta = 0$, the exact solution is given by $T_0(z,t)$, Eq. (5.4). We want to choose a time t such that $T_0(z,t)$ does not differ insignificantly from the initial T. We therefore must choose $g(t)$ so that it isn't too close to unity. If, for example, we choose $4\alpha t/L^2 = 1/4$, we have $\Delta T^{\mathrm{Bd}}/T_I = 0.21$. In other words, having chosen a value of t such that there have been significant changes in $T(z,t)$ from $T(z,o)$, the maximum error in $T(z,t)$, for any z, is of order 21%. We note that in arriving at Eq. (5.6), we made a number of unnecessary simplifications; furthermore, we have not introduced any variational parameters in our choice of $\bar{G}$.

Apart from the greater power of a V Bd formulation than of a VP formulation, one might have occasion to be more interested in a rigorous bound than in a VP of somewhat greater accuracy but with an error of unspecified sign. For example, one might want to be certain that in some region of space T never exceeds some local melting temperature T_{mel}. If initially $T_{\max}(\mathbf{r},\tau)$ is everywhere lower than the T_{mel} under consideration there is no problem. Suppose, however, that we have regions 1 and 2, with $T_{\mathrm{mel}\,1} < T_{\mathrm{mel}\,2}$, and that at $t = \tau$ the temperature in region 1 is below $T_{\mathrm{mel}\,1}$ but that the temperature in region 2 is below $T_{\mathrm{mel}\,2}$ but above $T_{\mathrm{mel}\,1}$. We would want to be certain that the temperature in region 1 never rose above $T_{\mathrm{mel}\,1}$. We thank Professor E. Gerjuoy for this observation.

Having obtained a V Bd for the time-dependent problem, it is natural to ask if the method applies to the time-independent problem. The answer is no. Thus, in the time-dependent problem we are given the initial temperature distribution, while in the time-independent problem the initial distribution remains constant and is exactly what we are trying to determine. We can of course make a time-dependent problem out of a time-independent problem by choosing an initial temperature distribution from the stationary distribution; the temperature will gradually relax to the time-independent case, but it will take an infinite time to do so, and our methods become poorer as the time interval increases.

In closing, we comment on the fact that the bound we obtained on $\Delta T(\mathbf{r},t)$ of Eq. (2.18), namely, $\Delta T^{\mathrm{Bd}}(t)$ of Eq. (2.19), is independent of $\mathbf{r}$. This has the desirable feature that for a given t one bound on $\Delta T(\mathbf{r},t)$ suffices to give upper and lower V Bd's on $T(\mathbf{r},t)$ for all $\mathbf{r}$, and those upper and lower V Bd's are *not* independent of $\mathbf{r}$ because of the presence of $T_v(\mathbf{r},t)$—see Eq. (2.14). Nevertheless, there might be some advantages to having a bound on $\Delta T(\mathbf{r},t)$ which was a function of $\mathbf{r}$, for if $\Delta T(\mathbf{r},t)$ varies considerably with $\mathbf{r}$, such a bound could be much closer. We are looking into this possibility.

This research was supported in part by the ONR under contract No. N00014-76-C-0317 and by the NSF under grant No. PHY7910413.

APPENDIX A. VARIATIONAL BOUNDS ON THE SOLUTION OF THE *INHOMOGENEOUS* HEAT EQUATION

Our considerations thus far have been based on the assumption that no sources (or sinks) are present. We will now show that one can bound the temperature $T(\mathbf{r},t)$ when that assumption is dropped.

With $C(\mathbf{r},t)\sigma(\mathbf{r},t)$ the heat generated per unit volume per unit time, $T(\mathbf{r},t)$ is defined by

$$A(\mathbf{r},t)T(\mathbf{r},t) = -\sigma(\mathbf{r},t), \qquad (\mathrm{A1})$$

subject to specified initial conditions and boundary conditions. It will now be convenient to introduce a temperature $T_{\mathrm{in}}(\mathbf{r},t)$ which satisfies the homogeneous differential equation

$$AT_{\mathrm{in}}(\mathbf{r},t) = 0 \qquad (\mathrm{A2})$$

subject to the initial condition

$$T_{\mathrm{in}}(\mathbf{r},\tau) = p(\mathbf{r}), \qquad (\mathrm{A3})$$

where $p(\mathbf{r})$ is specified, and to *homogeneous* boundary conditions. $T_{\mathrm{in}}(\mathbf{r},t)$ therefore satisfies conditions the same as had been used throughout the paper. As previously, U is the solution operator, satisfying $AU = 0$. We can now choose a function $\Theta(\mathbf{r},t)$ which is twice differentiable and which satisfies

$$\Theta(\mathbf{r},\tau) = T(\mathbf{r},\tau), \qquad (\mathrm{A4a})$$

$$\Theta(t)|^{\mathrm{surface}} = T(t)|^{\mathrm{surface}}. \qquad (\mathrm{A4b})$$

It follows immediately that the function T_h defined by

$$T_h(\mathbf{r},t) = T(\mathbf{r},t) - \Theta(\mathbf{r},t) \qquad (\mathrm{A5})$$

satisfies

$$T_h(\mathbf{r},\tau) = 0 \qquad (\mathrm{A6a})$$

and that

$$T_h(t)|^{\mathrm{surface}} = 0, \qquad (\mathrm{A6b})$$

that is, that T_h satisfies homogeneous initial conditions and homogeneous boundary conditions. One also immediately finds that

$$AT_h(\mathbf{r},t) = -\sigma(\mathbf{r},t) - A\Theta(\mathbf{r},t) \equiv w(\mathbf{r},t). \qquad (\mathrm{A7})$$

Although T_h satisfies an inhomogeneous differential equation, one can determine T_h because of the simple form of its initial conditions and boundary conditions. One has

$$T_h = -\int_\tau^t U(t,s)w(s)ds. \tag{A8}$$

We can now bound T_h in the usual fashion, using Eq. (2.5) to write T_h as the sum of a variational estimate and a second-order term which can be bounded. The bound on $T = T_h + \Theta$ follows immediately.

APPENDIX B. DERIVATION OF THE VARIATIONAL IDENTITY FOR $U(t,\tau)$

We derive in a relatively simple fashion the integral equation for $U(t,\tau)$, given in Eq. (2.4), a special case of Duhamel's principle. The proof proceeds along lines very similar to those used in the time-dependent scattering problem. The starting point is the relationship (to be proved)

$$G(\mathbf{r},t;\mathbf{r}',t') = G(\mathbf{r}',t';\mathbf{r},t)^\dagger \tag{B1a}$$

or, from Eq. (1.9),

$$U(t,t') = U(t',t)^\dagger, \tag{B1b}$$

where henceforth the † represents the transpose. To simplify the notation, we let $P(\mathbf{x})$ denote $P(\mathbf{r},t)$ for any function P. We can then rewrite Eq. (B1a) as

$$G(\mathbf{x},\mathbf{x}') = G(\mathbf{x}',\mathbf{x})^\dagger. \tag{B1c}$$

We will also use $\delta(\mathbf{x}-\mathbf{x}') = \delta(\mathbf{r}-\mathbf{r}')\delta(t-t')$ and $d\mathbf{x} = d\mathbf{r}\,dt$. A slight extension of the conditions under which G was defined, namely, Eqs. (3.2) and (3.4), gives as our definition

$$\left(K(\mathbf{x}) - \frac{\partial}{\partial t}\right)G(\mathbf{x},\mathbf{x}') = -\delta(\mathbf{x}-\mathbf{x}'), \tag{B2a}$$

$$G(\mathbf{x},\mathbf{x}') = 0, \quad t' > t, \tag{B2b}$$

$$[k(\mathbf{x})\nabla + l(\mathbf{x})]G(\mathbf{x},\mathbf{x}') = 0 \quad \text{on} \quad S, \tag{B2c}$$

where S is the surface and $l(\mathbf{x})$ is the transfer coefficient. $K^\dagger$, defined by $(\alpha, K\beta) = (K^\dagger\alpha,\beta)$ for arbitrary functions $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$, with the integration over $d\mathbf{x}$, is found to be

$$K^\dagger(\mathbf{x})\alpha(\mathbf{x}) = \nabla\cdot\{k(\mathbf{x})\nabla[\lambda(\mathbf{x})\alpha(\mathbf{x})]\},$$

while $(\partial/\partial t)^\dagger = -(\partial/\partial t)$. It is then natural to introduce the transposed Green's function $G^\dagger$ defined by

$$\left(K^\dagger(\mathbf{x}) + \frac{\partial}{\partial t}\right)G^\dagger(\mathbf{x},\mathbf{x}') = -\delta(\mathbf{x}-\mathbf{x}') \tag{B3a}$$

$$G^\dagger(\mathbf{x},\mathbf{x}') = 0, \quad t > t' \tag{B3b}$$

$$[k(\mathbf{x})\nabla + l(\mathbf{x})][G^\dagger(\mathbf{x},\mathbf{x}')\lambda(\mathbf{x})] = 0 \quad \text{on} \quad S. \tag{B3c}$$

With $\mathbf{y} \equiv (\mathbf{r}'',s)$, we now introduce the expression

$$I \equiv \int\left[G^\dagger(\mathbf{y},\mathbf{x})\left(K(\mathbf{y}) - \frac{\partial}{\partial s}\right)G(\mathbf{y},\mathbf{x}') - G(\mathbf{y},\mathbf{x}')\left(K^\dagger(\mathbf{y}) + \frac{\partial}{\partial s}\right)G^\dagger(\mathbf{y},\mathbf{x})\right]d\mathbf{y}, \tag{B4}$$

where

$$\int d\mathbf{y} \equiv \int_{t'-\epsilon}^{t+\epsilon} ds \int_V d\mathbf{r}'',$$

ϵ is a small positive number and V is the spatial volume of integration. Using Eqs. (B2a) and (B3a), we obtain

$$I = -G^\dagger(\mathbf{x}',\mathbf{x}) + G(\mathbf{x},\mathbf{x}'). \tag{B5}$$

We can also evaluate I differently. The terms in K and $K^\dagger$ cancel from the way in which K and $K^\dagger$ were defined, and we have

$$I = -\int\frac{\partial}{\partial s}[G^\dagger(\mathbf{y},\mathbf{x})G(\mathbf{y},\mathbf{x}')]d\mathbf{y} = -\int G^\dagger(\mathbf{y},\mathbf{x})G(\mathbf{y},\mathbf{x}')\Big|_{t'-\epsilon}^{t+\epsilon} d\mathbf{r}'' = 0. \tag{B6}$$

The last step follows from Eqs. (B2b) and (B3b); at the upper limit $G^\dagger$ is zero, and at the lower limit G is zero. We have therefore proved Eq. (B1).

To derive the variational identity for $U(t,\tau)$ we introduce the expression

$$J = \int_{\tau+0}^{t-0}\left\{U^\dagger(s,t)\left(K(s) - \frac{\partial}{\partial s}\right)\overline{U}(s,\tau) - \left[\left(K^\dagger(s) + \frac{\partial}{\partial s}\right)U^\dagger(s,t)\right]\overline{U}(s,\tau)\right\}ds. \tag{B7}$$

The terms in K and $K^\dagger$ cancel as above, and J reduces to

$$J = -\int_{\tau+0}^{t-0}\frac{\partial}{\partial s}[U^\dagger(s,t)\overline{U}(s,\tau)]ds = -\overline{U}(t,\tau) + U(t,\tau), \tag{B8}$$

using Eqs. (B1b), (1.4), and (2.3). Equating (B7) and (B8) gives the integral equation, Eq. (2.4).

[1] E. Gerjuoy, A. R. P. Rau, and L. Spruch, Phys. Rev. A **8**, 662 (1973); E. Gerjuoy, A. R. P. Rau, L. Rosenberg, and L. Spruch, J. Math. Phys. **16**, 1104 (1975).
[2] E. Gerjuoy, A. R. P. Rau, and L. Spruch, Rev.Mod. Phys. **55**, 725 (1983).
[3] K. Kalikstein and L. Spruch, Phys. Rev. A (in press).
[4] See, for example, P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[5] L. Spruch and L. Rosenberg, Phys. Rev. **116**, 1034 (1959).
[6] L. Rosenberg, L. Spruch, and T. F. O'Malley, Phys. Rev. **118**, 184 (1960).
[7] L. Rosenberg, Phys. Rev. D **1**, 1019 (1970).
[8] See, for example, J. R. Taylor, *Scattering Theory, The Quantum Theory of Nonrelativistic Collisions* (Wiley, New York, 1972).
[9] L. Spruch, in *Lectures in Theoretical Physics—Atomic Collisions*, edited by S. Geltman, K. T. Mahanthappa, and W. E. Brittin (Gordon and Breach, New York, 1969), Vol. XIC, p. 77.
[10] R. Shakeshaft and L. Spruch, Phys. Rev. A **8**, 206 (1973).
[11] R. Shakeshaft, Phys. Rev. A **12**, 2230 (1975).
[12] M. H. Protter and H. F. Weinberger, *Maximum Principles in Differential Equations* (Prentice Hall, Englewood Cliffs, New Jersey, 1967), Chap. 3. This chapter includes a discussion of the weak (Levi, Picone) and strong (Nierenberg) maximum principles.
[13] K. Yosida, *Functional Analysis*, 4th ed. (Springer, New York, 1974).
[14] A. Friedman, *Partial Differential Equations* (Holt, Rinehart, and Winston, New York, 1969).
[15] S. D. Eidel'man, *Parabolic Systems* (North-Holland, Amsterdam, 1969).
[16] H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids*, 2nd ed. (Clarendon, Oxford, 1959).
[17] A. V. Luikov, *Analytical Heat Diffusion Theory*, edited by J. P. Hartnell (Academic, New York, 1968).

On Lie–Bäcklund vector fields of the evolution equations $\partial^2 u/\partial x\,\partial t = f(u)$ and $\partial u/\partial t = \partial^2 u/\partial x^2 + f(u)$

W.-H. Steeb
Universität Paderborn, Theoretsche Physik, D-4790 Paderborn, West Germany

(Received 26 April 1983; accepted for publication 30 September 1983)

For the evolution equations $\partial^2 u/\partial x\,\partial t = f(u)$ and $\partial u/\partial t = \partial^2 u/\partial x^2 + f(u)$ we derive the analytic functions f where Lie–Bäcklund vector fields are admitted.

PACS numbers: 02.30. – f, 02.30.Jr

In the present paper we derive with the help of the jet bundle technique[1] a class of analytic functions f where the evolution equations

$$\frac{\partial^2 u}{\partial x\,\partial t} = f(u) \tag{1}$$

and

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(u) \tag{2}$$

admit Lie–Bäcklund vector fields.

Consider first Eq. (1). It is well known that this equation admits Lie–Bäcklund vector fields when the function f is given by $f(u) = e^u$ or $f(u) = \sin u$ or $f(u) = \sinh u$. The Bäcklund problem of Eq. (1) has been studied by Shadwick.[2]

We prove the following: "If the function f satisfies the ordinary differential equation $f'' + (2a/3)f = 0$, where a is an arbitrary real parameter, then Eq. (1) admits Lie–Bäcklund vector fields." The cases given above are included. For example, if we choose $a = \frac{3}{2}$, then $f(u) = \sin u$ fulfills this equation. On the other hand, if we choose $a = -\frac{3}{2}$, then $f(u) = \sinh u$ fulfills it. Before solving the equation $f'' + (2a/3)f = 0$ and discussing the general solution, we give the proof of our statement. Notice that also the wave equation $\partial^2 u/\partial x\,\partial t = e^u + e^{-2u}$ is integrable.

For describing Lie–Bäcklund vector fields of evolution equations the jet bundle technique is a suitable approach. We introduce the abbreviations $u_x = u_1$, $u_{xx} = u_2$ and so on. Within this approach we consider the submanifold

$$F \equiv u_{1t} - f(u) = 0 \tag{3}$$

and all its differential consequences with respect to the space coordinate. This means

$$F_1 \equiv u_{2t} - u_1 f' = 0\,, \tag{4a}$$

$$F_2 \equiv u_{3t} - u_1 f'' - u_2 f' = 0\,, \tag{4b}$$

$$F_3 \equiv \quad u_{4t} - u_1 f''' - 3u_1 u_2 f'' - u_3 f' = 0\,, \tag{4c}$$

$$\vdots$$

Let

$$V = g(u,u_1,u_2,u_3)\frac{\partial}{\partial u} \tag{5}$$

be a Lie–Bäcklund vector field. The assumption that the analytic function g depends also on x and t does not affect the results (this means the existence of Lie–Bäcklund vector fields) and therefore, for the sake of simplicity, they will be omitted. It is well known that Eq. (1) admits the Lie vector field $-x\partial/\partial x + t\partial/\partial t$ (scaling invariance). Chen *et al.*[3] have shown that the one-dimensional sine-Gordon equation admits a hierarchy of time-dependent Lie–Bäcklund vector fields. Neither does the assumption that the function g depends on $u_4,\ldots,u_n$ affect the result. Notice, however, that there is in general a hierarchy of Lie–Bäcklund vector fields if at least one exists. The nonlinearity in the evolution equation (1) only appears in the function f. The function f depends only on u. The term where the derivative appears is linear. Consequently, we can assume that the vector field V is linear in u_3, namely

$$V = (g_1(u,u_1,u_2) + u_3)\frac{\partial}{\partial u}\,. \tag{6}$$

Furthermore, we can assume, without loss of generality, that the function g_1 does not depend on u. If we include the dependence of u, then our calculations show that g_1 does not depend on u. Consequently,

$$V = (g_2(u_1,u_2) + u_3)\frac{\partial}{\partial u}\,. \tag{7}$$

The invariance requirement is expressed as

$$L_{\bar V} F \mathrel{\hat=} 0\,, \tag{8}$$

where $L_{\bar V}(\cdot)$ denotes the Lie derivative and $\hat=$ stands for the restriction to solutions of Eq. (1). $\bar V$ is the extended vector field of V. Due to the structure of Eq. (1) we are only forced to include the term of the form $(\cdots)\partial/\partial u_{1t}$ in the extended vector field $\bar V$. From condition (8) it follows that

$$u_3 u_1 \frac{\partial^2 g_2}{\partial u_2^2} f' + u_3 \frac{\partial^2 g_2}{\partial u_1\,\partial u_2} f + u_2 \frac{\partial^2 g_2}{\partial u_1^2} f$$
$$+ u_2 u_1 \frac{\partial^2 g_2}{\partial u_1\,\partial u_2} f' + (u_1^2 f'' + u_2 f')\frac{\partial g_2}{\partial u_2} + 3u_2 u_1 f''$$
$$+ u_1^3 f''' + u_1 \frac{\partial g_2}{\partial u_1} f' - g_2 f' = 0\,, \tag{9}$$

where we have taken into account Eqs. (4a)–(4c). Separating out the terms with the factors $u_3 u_1$ and u_3 we obtain

$$u_3 u_1 \frac{\partial^2 g_2}{\partial u_2^2} f' = 0 \tag{10}$$

and

$$u_3 \frac{\partial^2 g_2}{\partial u_1\,\partial u_2} f = 0\,. \tag{11}$$

If we assume that the function g_2 does not depend on u_2, then

 0022-2488/84/020237-03$02.50

Eqs. (10) and (11) are satisfied. From Eq. (9) it follows that

$$u_2\left(3u_1 f'' + f\frac{\partial^2 g_2}{\partial u_1^2}\right) = 0 \tag{12}$$

and

$$u_1^3 f''' + u_1 f' \frac{\partial g_2}{\partial u_1} - g_2 f' = 0 . \tag{13}$$

Since we assume that the function f is a nonlinear analytic function of u, it follows that the function g_2 must be of the form

$$g_2(u_1) = a u_1^3/3 , \tag{14}$$

where a is an arbitrary real parameter ($a \neq 0$). We obtain

$$u_2 u_1 (3f'' + 2af) = 0 \tag{15}$$

and

$$u_1^3 (f''' + (2a/3)f') = 0 . \tag{16}$$

For solving both Eqs. (15) and (16) simultaneously, we have to solve

$$f'' + (2a/3)f = 0 . \tag{17}$$

Consequently, the statement given above has been proved.

Let us now study the solutions to Eq. (17). We have to distinguish between the cases $a > 0$ and $a < 0$. First let $a > 0$ and we put $a = \frac{3}{2}$. Then we obtain

$$f(u) = C_1 \sin u + C_2 \cos u , \tag{18}$$

where C_1 and C_2 are two real constants. Second let $a < 0$ and we put $a = -\frac{3}{2}$. Then we find

$$f(u) = C_1 \cosh u + C_2 \sinh u . \tag{19}$$

To summarize, the evolution equations

$$\frac{\partial^2 u}{\partial x\, \partial t} = C_1 \sin u + C_2 \cos u \tag{20}$$

and

$$\frac{\partial^2 u}{\partial x\, \partial t} = C_1 \sinh u + C_2 \cosh u \tag{21}$$

admit Lie–Bäcklund vector fields. The simplest one is given by the vector field (7) together with Eq. (14). Moreover, within the technique described above we also find that the evolution equations of the form $\partial^2 u/\partial x\, \partial t = f(u)$, where f is a polynomial in u do not admit Lie–Bäcklund vector fields.

Some comments are in order. We have shown that the evolution equations (20) and (21) admit at least one Lie–Bäcklund vector field. It can be proved that they admit a hierarchy of Lie–Bäcklund vector fields. They can be found with the help of a recursion operator. A simpler approach for finding the hierarchy of Lie Bäcklund vector fields is given by Chen *et al.*[3] where the Lie point-symmetry vector fields, which depend on space coordinate and time coordinate, come into play. The evolution equations (20) and (21) have one-parameter families of Bäcklund transformations. Notice that the sine-Gordon equation has an auto-Bäcklund transformation whereas the Liouville equation has a Bäcklund transformation (the equation can be linearized). As is well known, formal power series expansion in the Bäcklund parameter lead, in such cases, to the hierarchy of Lie–Bäcklund vector fields. Both hierarchies are equal. The question whether the Lie–Bäcklund vector fields generate Lie–Bäcklund transformation groups is not clear (compare Ref. 4 for details of this question).

Consider now Eq. (2). Recently, several authors (compare for example Ref. 5) have studied the nonlinear diffusion equation $\partial u/\partial t = \partial(f(u)\partial u/\partial x)/\partial x$ and Lie–Bäcklund vector fields. They found that only in the case where $f(u) = u^{-2}$ does this equation admit Lie–Bäcklund vector fields. Like the Lie point vector fields the Lie–Bäcklund vector fields can also be used for finding solutions to the underlying partial differential equation.[6] Moreover, a mapping to the linear diffusion equation can be given. In the following we discuss whether Lie–Bäcklund vector fields of Eq. (2) exist. Recently, the Lie point vector fields of Eq. (2) have been given where $f(u) = ku^n$.[7]

As described above by our first example, within the jet bundle approach we consider the submanifold

$$f \equiv u_t - u_2 - f(u) = 0 \tag{22}$$

and all its differential consequences with respect to the space coordinate. This means

$$\begin{aligned} F_1 &\equiv u_{1t} - u_3 - u_1 f' = 0 , \\ F_2 &\equiv u_{2t} - u_4 - u_1 f'' - u_2 f' = 0 , \\ F_3 &\equiv u_{3t} - u_5 - 3u_1 u_2 f'' - u_1 f''' - u_3 f' = 0 , \\ &\vdots \end{aligned} \tag{23}$$

Let

$$V = g(x,t,u,u_1,u_2,u_3) \frac{\partial}{\partial u} \tag{24}$$

be a Lie–Bäcklund vector field, where g is an analytic function. Due to the structure of the evolution equation (2) we can simplify our vector field V without loss of generality, namely,

$$V = (g_1(u,u_1,u_2) + u_3) \frac{\partial}{\partial u} . \tag{25}$$

Notice that, if we study the diffusion equation $\partial u/\partial t = \partial(u^{-2}\partial u/\partial x)/\partial x$, then for the vector field V we must make the ansatz $V = (g_1(u,u_1,u_2) + g_2(u)u_3)\partial/\partial u$.

As described above, the invariance requirement is expressed as $L_{\bar V} F \mathrel{\hat=} 0$. Due to the structure of Eq. (2) we are only forced to include the terms of the form $(\cdots)\partial/\partial u_t$ and $(\cdots)\partial/\partial u_2$ in the extended vector field $\bar V$. From the condition $L_{\bar V} F \mathrel{\hat=} 0$ it follows that

$$\begin{aligned} &\frac{\partial g_1}{\partial u} f + u_1 \frac{\partial g_1}{\partial u_1} f' + (u_1^2 f'' + u_2 f') \frac{\partial g_1}{\partial u_2} \\ &\quad + 3u_1 u_2 f'' + u_1^3 f''' - u_1^2 \frac{\partial^2 g_1}{\partial u^2} - 2u_1 u_2 \frac{\partial^2 g_1}{\partial u\, \partial u_1} \\ &\quad - 2u_1 u_3 \frac{\partial^2 g_1}{\partial u\, \partial u_2} - u_2^2 \frac{\partial^2 g_1}{\partial u_1^2} - 2u_2 u_3 \frac{\partial^2 g_1}{\partial u_1\, \partial u_2} \\ &\quad - u_3^2 \frac{\partial^2 g_1}{\partial u_2^2} - g_1 f' = 0 . \end{aligned} \tag{26}$$

Separating out the term with the factor u_3^2 we obtain

$$u_3^2 \frac{\partial^2 g_1}{\partial u_2^2} = 0 . \tag{27}$$

Consequently the function g_1 takes the form

$$g_1(u,u_1,u_2) = g_2(u,u_1)u_2 + g_3(u,u_1)\,. \tag{28}$$

From Eq. (26) it also follows that

$$u_1u_3 \frac{\partial^2 g_1}{\partial u\,\partial u_2} = 0 \tag{29}$$

and

$$u_2u_3 \frac{\partial^2 g_1}{\partial u_1\,\partial u_2} = 0\,. \tag{30}$$

With the help of Eq. (28) we find that the function g_2 does not depend on u_1 and u. Consequently,

$$g_1(u,u_1,u_2) = C_1u_2 + g_3(u,u_1)\,. \tag{31}$$

Inserting Eq. (31) into Eq. (28) it follows that

$$\begin{aligned} C_1u_1^2 f'' + 3u_1u_2 f'' + u_1 f''' - g_3 f' + f\frac{\partial g_3}{\partial u} + u_1 f' \frac{\partial g_3}{\partial u_1} \\ - u_1^2 \frac{\partial^2 g_3}{\partial u^2} - 2u_1u_2 \frac{\partial^2 g_3}{\partial u\,\partial u_1} - u_2^2 \frac{\partial^2 g_3}{\partial u_1^2} = 0\,. \end{aligned} \tag{32}$$

From Eq. (32) we have

$$u_2^2 \frac{\partial^2 g_3}{\partial u_1^2} = 0 \tag{33}$$

and therefore the function g_3 takes the form

$$g_3(u,u_1) = g_4(u)u_1 + g_5(u)\,. \tag{34}$$

Then from Eq. (32) we obtain

$$u_1u_2(3f'' - 2g_4') = 0\,, \tag{35}$$

and therefore $g_4 = \frac{3}{2}f' + C_2$. It follows that

$$\begin{aligned} C_1u_1^2 f'' + u_1 f''' - g_5 f' + \tfrac{3}{2} u_1 f f'' + f g_5' \\ - \tfrac{3}{2} u_1^3 f''' - u_1^2 g_5'' = 0\,. \end{aligned} \tag{36}$$

From Eq. (36) we see that the following statement holds: The diffusion equation (2) admits Lie–Bäcklund vector fields if and only if $f''(u) = 0$. Hence Eq. (2) becomes linear. The vector field V takes the form

$$V = (u_3 + C_1u_2 + C_2u_1)\frac{\partial}{\partial u}\,. \tag{37}$$

Notice that we obtain the same result when we extend the vector field (24) to

$$V = g(x,t,u,u_1,\ldots,u_n)\frac{\partial}{\partial u}\,. \tag{38}$$

Equation (2) belongs to the following class of partial differential equations which admit a hierarchy of Lie–Bäcklund vector fields, namely,

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f_1(u)\left(\frac{\partial u}{\partial x}\right)^2 + f_2(u)\frac{\partial u}{\partial x} + f_3(u)\,, \tag{39}$$

and the functions f_1, f_2, and f_3 satisfy the system of differential equations

$$f_2' f_3 = 0\,, \quad f_2' f_1 = f_2''\,, \quad f_3'' + (f_1 f_3)' = 0\,. \tag{40}$$

Thus if $f_3(u) = 0, f_1(u) = 0$, and $f_2(u) = u$, then Eq. (40) is satisfied and we obtain the well-known Burgers equation. If we put $f_1(u) = 0$ and $f_2(u) = 0$, then it follows that $f_3''(u) = 0$. Consequently, $f_3(u) = au + b$ $(a,b \in R)$.

Finally, we mention that we find "nonlinear" diffusion equations, which admit Lie–Bäcklund vector fields, when we consider systems of diffusion equations. For example, the system of diffusion equations

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + v^2\,, \quad \frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} \tag{41}$$

admits Lie–Bäcklund vector fields.

[1] W.-H. Steeb and W. Strampp, Physica A **114**, 95 (1982).
[2] W. F. Shadwick, J. Math. Phys. **19**, 2317 (1978).
[3] H. H. Chen, Y. C. Lee, and J.-E. Lin, "On a new hierarchy of symmetries for the integrable evolution equations" (presented at the Conference on Nonlinear Waves and Integrable Systems, 1982, East Carolina University).
[4] R. L. Anderson and N. H. Ibragimov, *Lie Bäcklund Transformations in Applications* (SIAM, Philadelphia, 1979).
[5] G. Bluman and S. Kumei, J. Math. Phys. **21**, 1019 (1980).
[6] W.-H. Steeb and W. Oevel, Z. Naturforschung A **38**, 86 (1983).
[7] T. Branson and W.-H. Steeb, J. Phys. A: Math. Gen. **16**, 469 (1983).

Singular solutions of the axially symmetric Bogomolny equations

Hedley C. Morris and Roger K. Dodd
School of Mathematics, Trinity College, Dublin, Ireland

(Received 11 May 1982; accepted for publication 22 December 1982)

A singular Bäcklund transformation is constructed for the Ernst equation and used to construct singular solutions to the axially symmetric Bogomolny equations.

PACS numbers: 02.30.Hq

I. INTRODUCTION

The Bogomolny equations[1] for the SU(2) Yang–Mills theory, in the limit of vanishing Higgs potential, can be written in a vector form by using the Pauli matrices as the SU(2) basis. The equations then connect a triplet of vector fields $(\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3)$ to a single vector field $\mathbf{\Phi}$ according to the equations

$$\partial_a \mathbf{A}_b - \partial_b \mathbf{A}_a + \mathbf{A}_a \times \mathbf{A}_b = -\epsilon_{abc}(\partial_c \mathbf{\Phi} + \mathbf{A}_c \times \mathbf{\Phi}), \tag{1.1}$$

where ∂_a denotes the partial derivative with respect to the coordinate x_a and the symbol $\times$ denotes the normal vector product in R^3. The quantity ϵ_{abc} is the permutation symbol and arises from the commutation relations of the Pauli matrices.

For the axially symmetric configurations Manton[2] introduced the ansatz for the solutions to Eqs. (1.1) expressed in axial polar coordinates,

$$\mathbf{\Phi} = (0, \phi_1, \phi_2), \quad \mathbf{A}_\phi = -(0, \eta_1, \eta_2),$$
$$\mathbf{A}_z = -(w_1, 0, 0), \quad \mathbf{A}_\rho = -(w_2, 0, 0), \tag{1.2}$$

where w_i, ϕ_i, and η_i are functions of the axial polar coordinates ρ and z alone. As a result, he was able to reduce Eqs. (1.1) to the five equations

$$\partial_z\phi_1 - w_1\phi_2 = \rho^{-1}(\partial_\rho \eta_1 - w_2\eta_2), \tag{1.2a}$$
$$\partial_z\phi_2 + w_2\phi_1 = \rho^{-1}(\partial_\rho \eta_2 + w_2\eta_1), \tag{1.2b}$$
$$\partial_\rho w_1 - \partial_z w_2 = \rho^{-1}(\phi_1\eta_2 - \phi_2\eta_1), \tag{1.2c}$$
$$\partial_\rho\phi_1 - w_2\phi_2 = -\rho^{-1}(\partial_z \eta_1 - w_1\eta_2), \tag{1.2d}$$
$$\partial_\rho\phi_2 + w_2\phi_1 = -\rho^{-1}(\partial_z \eta_2 + w_1\eta_1). \tag{1.2e}$$

Forgács *et al.*[3] then observed that if the Manton fields were parametrized in terms of two new functions as

$$\phi_1 = f^{-1}\partial_z\psi, \quad \eta_1 = -\rho f^{-1}\partial_\rho\psi, \quad w_1 = -f^{-1}\partial_z\psi,$$
$$\phi_2 = -f^{-1}\partial_z f, \quad \eta_2 = \rho f^{-1}\partial_\rho f, \quad w_2 = -f^{-1}\partial_\rho\psi \tag{1.3}$$

then the requirement that the Manton fields satisfy (1.2) is confirmed if the complex quantity $E = f + i\psi$ is a solution of the single complex equation

$$(\mathrm{Re}\, E)\Delta E = (\nabla E)^2, \tag{1.4}$$

where Δ is the Laplacian operator $(\partial_\rho^2 + \rho^{-1}\partial_\rho + \partial_z^2)$. Equation (1.4) is the Ernst equation[4] of general relativity. This equation is known to be associated with an inverse scattering problem[5,6] and to possess soliton solutions.

In this paper we use the deformation problem formulated in Ref. 7 to derive a new type of singular Bäcklund transformation for the Ernst and Bogomolny equations.

II. A SINGULAR BÄCKLUND TRANSFORMATION

It has been shown in Ref. 6 that the Ernst equation can be expressed as the integrability requirement for the system

$$L_1\Psi = (A + \eta B)\Psi, \tag{2.1a}$$
$$L_2\Psi = (C + \eta^{-1}D)\Psi, \tag{2.1b}$$

where L_1 and L_2 are the differential operators

$$L_1 = (\rho\partial_\rho + 2\eta\partial_\eta - \eta\partial_z), \tag{2.2a}$$
$$L_2 = (\rho^{-1}\partial_\rho + \eta^{-1}\partial_z), \tag{2.2b}$$

which satisfy the commutator algebra

$$[L_1, L_2] = -2L_2. \tag{2.3}$$

The matrices A, B, C, and D are given by

$$A = \begin{bmatrix} r & 0 \\ s & -r \end{bmatrix}, \quad B = \begin{bmatrix} b & a \\ 0 & -b \end{bmatrix}, \tag{2.4a}$$

where

$$r = \tfrac{1}{2}\eta_2, \quad s = -\eta_1, \quad b = -\tfrac{1}{2}\phi_2, \quad a = \phi_1 \tag{2.4b}$$

and C and D are related to A and B by

$$C = -\rho^{-2}A^\tau, \quad D = B^\tau, \tag{2.4c}$$

where τ denotes the transpose of a matrix.

Equations (1.2) have a useful symmetry property. If Ψ is a solution of (2.1), then so also is $\hat{\Psi}$ defined by

$$\hat{\Psi}^{-1}(\eta, \rho, z) = \Psi^\tau(-\rho^2\eta^{-1}, \rho, z). \tag{2.5}$$

In general, Ψ and $\hat{\Psi}$ are independent solutions.

Let us now suppose that we know a matrix solution Ψ_0 to the system (2.1) for matrices A_0, B_0, C_0, and D_0 corresponded to a known solution of the Ernst equation:

$$L_1\Psi_0 = (A_0 + \eta B_0)\Psi_0, \tag{2.6a}$$
$$L_2\Psi_0 = (C_0 + \eta^{-1}D_0)\Psi_0. \tag{2.6b}$$

We now seek solutions to the system (2.1) in the form

$$\Psi(\eta, \rho, z) = \chi(\eta, \rho, z)\Psi_0(\eta, \rho, z). \tag{2.7}$$

The requirement that this be possible for matrix functions χ having specified analytical structure in the complex η-plane leads to a relationship between the original matrices A_0, B_0, C_0, and D_0 and new matrices A, B, C, and D, which are the Bäcklund transform of the original "seed" matrices A_0, B_0, C_0, and D_0.

In the work of Zakharov and Belinskii,[6] the assumed analytical structure of χ is that of a finite set of simple poles. In this paper we consider the case in which one of the poles is at infinity and a second is at zero. As a result of the symmetry

 0022-2488/84/020240-05$02.50

property (2.5) we are able to impose the constraint

$$\hat{\chi} = \chi. \tag{2.8}$$

This ensures that Ψ_0 and $\hat{\Psi}_0$ both give rise to independent solutions of the transformed equation. However, we do not utilize that fact here, and (2.8) may be viewed as a constraint that can be consistently imposed in the search for exact solutions. It is the condition (2.8) that requires us to have two poles.

We assume χ to have the form

$$\chi = gU(I + S\eta^{-1} + P\eta), \tag{2.9}$$

where g is a scalar function of ρ and z and U and P are matrix functions of ρ and z. The matrix U is chosen to maintain the specific reduction of the general system (2.1) inherent in the specific choices (2.4), which correspond to the Bogomolny equations. In order to maintain the symmetry condition (2.1c), the matrix U is chosen to be orthogonal and parametrized in the form

$$U = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}. \tag{2.10}$$

The requirement (2.8) implies that

$$\chi^{-1} = g(I - \eta S^\tau \rho^{-2} - \eta^{-1} P^\tau \rho^2) U^\tau, \tag{2.11}$$

and the trivial identity

$$\chi\chi^{-1} = I \tag{2.12}$$

gives rise to the equations

$$S = -\rho^2 P^\tau, \tag{2.13a}$$

$$P^2 = 0, \tag{2.13b}$$

$$PP^\tau + P^\tau P = \rho^{-2}(1 - g^{-2})I. \tag{2.13c}$$

The general solution to (2.13b) can be parametrized in the form

$$P = \alpha\begin{bmatrix} \lambda & -\lambda^2 \\ 1 & -\lambda \end{bmatrix}, \tag{2.14}$$

where α and λ are functions of ρ and z. From (2.13c) we obtain

$$g = [1 - \rho^2\alpha^2(1 + \lambda^2)^2]^{-1/2}. \tag{2.15}$$

The final forms for χ and χ^{-1} are then given by

$$\chi = gU(I + \rho^2\eta^{-1}P^\tau + \eta P), \tag{2.16a}$$

$$\chi^{-1} = g(I - \eta P - \rho^2\eta^{-1}P^\tau)U. \tag{2.16b}$$

The appearance of the singular function g means that all of the solutions that we will generate will be singular on the surfaces.

$$\rho\alpha(1 + \lambda^2) = \pm 1. \tag{2.17}$$

Substituting (2.7) into (2.1), we obtain

$$(L_1\chi)\chi^{-1} + \chi(A_0 + \eta B_0)\chi^{-1} = (A + \eta B), \tag{2.18a}$$

$$(L_2\chi)\chi^{-1} + \chi(C_0 + \eta^{-1}D_0)\chi^{-1} = (C + \eta^{-1}D). \tag{2.18b}$$

Using the parametrized forms (2.16) for χ and χ^{-1}, we obtain from (2.18a), by examining the pole structure at 0 and infinity, the equations

$$\eta^3:\quad (\partial_z P)P - PB_0P = 0, \tag{2.19a}$$

$$\eta^2:\quad -\partial_z P - (\rho\partial_\rho P)P + [P, B_0] - PA_0P = 0, \tag{2.19b}$$

$$\eta^{-2}:\quad -(\rho\partial_\rho P^\tau)P^\tau - P^\tau A_0 P^\tau = 0, \tag{2.19c}$$

$$\eta^{-1}:\quad \rho\partial_\rho P^\tau + \rho^2(\partial_z P^\tau)P^\tau + [P^\tau, A_0] - \rho^2 P^\tau B_0 P^\tau = 0, \tag{2.19d}$$

$$\begin{aligned}\eta^0:\quad & g^2[-\rho^2\partial_z P^\tau - \rho^3(\partial_\rho P^\tau)P - (\rho\partial_\rho P + 2P)\rho^2 P^\tau \\ & + A_0 + \rho^2[P^\tau, B_0] - \rho^2 P^\tau A_0 P - \rho^2 P A_0 P^\tau] \\ & = \tilde{A} - (\rho\partial_\rho g)g^{-1}I,\end{aligned} \tag{2.20a}$$

$$\begin{aligned}\eta:\quad & g^2[\rho\partial_\rho P + 2P + \rho^2(\partial_z P^\tau)P + \rho^2(\partial_z P)P^\tau + B_0 \\ & + [P, A_0] - \rho^2 P^\tau B_0 P - \rho^2 P B_0 P^\tau] \\ & = \tilde{B} + (\partial_z g)g^{-1}I,\end{aligned} \tag{2.20b}$$

where $\tilde{A}$ and $\tilde{B}$ are related to A and B by a final gauge transformation

$$A = U\tilde{A}U^\tau + (\rho\partial_\rho U)U^\tau, \tag{2.21a}$$

$$B = U\tilde{B}U^\tau - (\partial_z U)U^\tau. \tag{2.21b}$$

The matrix U must be chosen to put A and B into the lower and upper triangular form, respectively. As a result of the symmetry, (2.4c) and the orthogonality of U equations (2.18b) provide exactly the same equations (2.19)–(2.21).

The substitution of (2.14) into (2.19) yields the following equations for λ and α:

$$\partial_z\lambda = -(2b_0\lambda + a_0), \tag{2.22a}$$

$$\rho\partial_\rho\lambda = -(2r_0\lambda + s_0), \tag{2.22b}$$

$$\partial_z\alpha = 2b_0\alpha - (4r_0\lambda + (1 - \lambda^2)s_0)\alpha^2, \tag{2.23a}$$

$$\rho\partial_\rho\alpha = 2r_0\alpha + \rho^2\alpha^2[4b_0\lambda + a_0(1 - \lambda^2)]. \tag{2.23b}$$

In terms of the Ernst potentials f and g the solution to (2.22) is given by

$$\lambda = f_0^{-1}(c_0 - g_0), \tag{2.24}$$

where c_0 is a constant. By introducing the new dependent variable $q = f\alpha^{-1}$ Eqs. (2.22) can be reduced to the form

$$\partial_z q = -\rho[2\partial_\rho f_0\lambda + (1 - \lambda^2)\partial_\rho g_0], \tag{2.25a}$$

$$\partial_\rho q = \rho[2\partial_z f_0\lambda + (1 - \lambda^2)\partial_z g_0], \tag{2.25b}$$

which can be directly integrated as the right-hand sides are explicitly known. If we write

$$\tilde{A} = \begin{bmatrix} \tilde{r} & \tilde{t} \\ \tilde{s} & -\tilde{r} \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} \tilde{b} & \tilde{a} \\ \tilde{c} & -\tilde{b} \end{bmatrix}, \tag{2.26}$$

we obtain from (2.20) the following expressions for $\tilde{r}, \tilde{t}, \tilde{s}, \tilde{b}, \tilde{a}, \tilde{c}$ in terms of λ and α:

$$\begin{aligned}\tilde{r} = g^2\{& r_0[1 + \alpha^2\rho^2(\lambda^2 + 1)^2] \\ & + a_0\alpha(\lambda^2 + 1)\rho^2 - \alpha^2\rho^2(\lambda^4 - 1)\},\end{aligned} \tag{2.27a}$$

$$\begin{aligned}\tilde{s} = g^2\{& s_0[1 + \alpha^2\rho^2(\lambda^4 - 1)] - 4\lambda(\lambda^2 + 1)\rho^2\alpha^2 r_0 \\ & - (4b_0\lambda + 2a_0)\lambda\alpha\rho^2 - 2\lambda(\lambda^2 + 1)\alpha^2\rho^2\},\end{aligned} \tag{2.27b}$$

$$\begin{aligned}\tilde{t} = g^2\rho^2[& 2\lambda\alpha a_0 - 4\alpha b_0 + 4\alpha^2\lambda r_0(\lambda^2 + 1) \\ & + 2\alpha^2(\lambda^2 + 1)s_0 - 2\alpha^2(\lambda^2 + 1)\lambda],\end{aligned} \tag{2.27c}$$

$$\tilde{b} = g^2\{b_0[1 + \rho^2\alpha^2(\lambda^2 + 1)^2] - \alpha(\lambda^2 + 1)s_0 + 2\alpha\lambda\}, \tag{2.28a}$$

$$\begin{aligned}\tilde{a} = g^2\{& a_0[1 + \rho^2\alpha^2(\lambda^4 - 1)] - 4\lambda(1 + \lambda^2)\rho^2\alpha^2 b_0 \\ & + 4\alpha\lambda^2 r_0 + 2\lambda\alpha s_0 - 2\alpha\lambda^2\},\end{aligned} \tag{2.28b}$$

$$\tilde{c} = g^2\{a_0[2\alpha^2\rho^2(1+\lambda^2)] + b_0[4\alpha^2\rho^2\lambda(1+\lambda^2)] + 4\alpha r_0 - 2\alpha\lambda s_0 + 2\alpha\}. \quad (2.28c)$$

From (2.27) and (2.28) we also obtain the following identities:

$$-2\lambda\tilde{r} + \lambda^2\tilde{t} - \tilde{s} = 2\lambda r_0 - s_0 = \rho\partial_\rho\lambda, \quad (2.29a)$$

$$-2\lambda\tilde{b} + \lambda^2\tilde{c} - \tilde{a} = -2\lambda b_0 - a_0 = \partial_z\lambda, \quad (2.29b)$$

which we shall need in our discussion of the gauge transformation U. From (2.21) and (2.10) we obtain

$$r = \tilde{r}\cos 2\theta + \tfrac{1}{2}\sin 2\theta(\tilde{t}+\tilde{s}), \quad (2.30a)$$

$$s = \tilde{s}\cos^2\theta - \tilde{t}\sin^2\theta - \tilde{r}\sin 2\theta - \rho\partial_\rho\theta, \quad (2.30b)$$

$$0 = \cos^2\theta\tilde{t} - \sin^2\theta\tilde{s} + \rho\partial_\rho\theta, \quad (2.30c)$$

$$b = \tilde{b}\cos 2\theta + \tfrac{1}{2}\sin 2\theta(\tilde{a}+\tilde{c}), \quad (2.31a)$$

$$a = \cos^2\theta\tilde{a} - \sin^2\theta\tilde{c} - \sin 2\theta\tilde{b} - \partial_z\theta, \quad (2.31b)$$

$$0 = \cos^2\theta\tilde{c} - \sin^2\theta\tilde{a} - \sin 2\theta\tilde{b} + \partial_z\theta. \quad (2.31c)$$

Thus we obtain two equations, (2.30c) and (2.31c), to determine θ. In terms of the variable $Y = \tan\theta$ these become the equations

$$\rho\partial_\rho Y = (2\tilde{r}Y + \tilde{s}Y^2 - \tilde{t}), \quad (2.32a)$$

$$\partial_z Y = (2\tilde{b}Y + \tilde{a}Y^2 - \tilde{c}). \quad (2.32b)$$

Fortunately, we are able to solve these equations. Comparison between (2.32) and (2.29) shows that $Y = \lambda^{-1}$ is a solution provided $\lambda \neq 0$.

Equations (2.30a), (2.30b), (2.31a), and (2.31b) become

$$r = \left(\frac{\lambda^2-1}{\lambda^2+1}\right)\tilde{r} + \left(\frac{\lambda}{\lambda^2+1}\right)(\tilde{t}+\tilde{s}), \quad (2.33a)$$

$$s = \left(\frac{\lambda^2-1}{\lambda^2+1}\right)(\tilde{s}+\tilde{t}) - \left(\frac{4\lambda}{\lambda^2+1}\right)\tilde{r}, \quad (2.33b)$$

$$b = \left(\frac{\lambda^2-1}{\lambda^2+1}\right)\tilde{b} + \left(\frac{\lambda}{\lambda^2+1}\right)(\tilde{a}+\tilde{c}), \quad (2.34a)$$

$$a = \left(\frac{\lambda^2-1}{\lambda^2+1}\right)(\tilde{a}+\tilde{c}) - \left(\frac{4\lambda}{\lambda^2+1}\right)\tilde{b}. \quad (2.34b)$$

Together with Eqs. (2.27) and (2.28), these equations define a Bäcklund transformation for the Ernst equation and axially symmetric Bogomolny equations.

Combining (2.27), (2.28), and (2.34), we obtain the Bäcklund transformation for the Bogomolny equations given by

$$r = \left[\left(\frac{\lambda^2-1}{\lambda^2+1}\right)r_0 + \left(\frac{\lambda}{\lambda^2+1}\right)s_0\right]\left(\frac{1+\Delta}{1-\Delta}\right) + [\alpha(\lambda^2-1)a_0\rho^2 - 4\lambda\alpha\rho^2 b_0] \times\left(\frac{1}{1-\Delta}\right) - \left(\frac{\Delta}{1-\Delta}\right), \quad (2.35a)$$

$$b = \left[\left(\frac{\lambda^2-1}{\lambda^2+1}\right)b_0 + \left(\frac{\lambda}{\lambda^2+1}\right)a_0\right]\left(\frac{1+\Delta}{1-\Delta}\right) + [4\lambda\alpha r_0 - \alpha(\lambda^2-1)s_0]\left(\frac{1}{1-\Delta}\right), \quad (2.35b)$$

$$s = \left[\left(\frac{\lambda^2-1}{\lambda^2+1}\right)s_0 - \left(\frac{4\lambda}{\lambda^2+1}\right)r_0\right]\left(\frac{1+\Delta}{1-\Delta}\right) - [4\lambda a_0\rho^2\alpha + 4\alpha\rho^2(\lambda^2-1)b_0]\left(\frac{1}{1-\Delta}\right), \quad (2.35c)$$

$$a = \left[\left(\frac{\lambda^2-1}{\lambda^2+1}\right)a_0 - \left(\frac{4\lambda}{\lambda^2+1}\right)b_0\right]\left(\frac{1+\Delta}{1-\Delta}\right) + [4\lambda\alpha s_0 + 4\alpha(\lambda^2-1)r_0] \times\left(\frac{1}{1-\Delta}\right) - \frac{2\alpha(1+\lambda^2)}{(1-\Delta)}, \quad (2.35d)$$

where $\lambda = f_0^{-1}(c_0 - g_0)$, $\Delta = \alpha^2\rho^2(1+\lambda^2)^2$, and $\alpha = -fq^{-1}$. The function q is completely determined by Eqs. (2.25) once a seed solution $E_0 = f_0 + ig_0$ has been specified.

As a step towards integrating these equations to obtain the Ernst potentials f and g, we introduce the function H defined by

$$f = (1-\Delta)H, \quad (2.36)$$

where f is the real part of the new Bäcklund transformed Ernst potential. From (2.35) we obtain

$$\tfrac{1}{2}\rho H^{-1}\partial_\rho H = r_0\left(\frac{\lambda^2-1}{\lambda^2+1}\right) + s_0\left(\frac{\lambda}{\lambda^2+1}\right) - 4b_0\rho^2\alpha\lambda + a_0\rho^2(\lambda^2-1)\alpha, \quad (2.37a)$$

$$\tfrac{1}{2}H^{-1}\partial_z H = b_0\left(\frac{\lambda^2-1}{\lambda^2+1}\right) + a_0\left(\frac{\lambda}{\lambda^2+1}\right) + 4r_0\lambda\alpha - (\lambda^2-1)\alpha s_0. \quad (2.37b)$$

From Eqs. (2.22) and (2.23) we see that the solution to these equations is given by

$$H = (\lambda^2+1)^{-1}\alpha^{-2}f_0 C, \quad (2.38)$$

and so the new potential f is given by

$$f = (1-\Delta)f_0/\alpha^2(\lambda^2+1), \quad (2.39)$$

where without loss of generality we have set $C = 1$.

Equations (2.35) for the imaginary part g of the Bäcklund transformed Ernst potential now become

$$\rho g_\rho = \frac{f_0}{\alpha^2(\lambda^2+1)}\left[s_0\left(\frac{\lambda^2-1}{\lambda^2+1}\right) - \left(\frac{4\lambda}{\lambda^2+1}\right)r_0\right](1+\Delta) - \frac{4f_0\rho^2}{\alpha(\lambda^2+1)}[a_0\lambda + (\lambda^2-1)b_0], \quad (2.40a)$$

$$g_z = \frac{f_0}{\alpha^2(\lambda^2+1)}\left[a_0\left(\frac{\lambda^2-1}{\lambda^2+1}\right) - \left(\frac{4\lambda}{\lambda^2+1}\right)b_0\right](1+\Delta) - \frac{4f_0}{\alpha(\lambda^2+1)}[\lambda s_0 + (\lambda^2-1)r_0] - \frac{2}{\alpha}f_0. \quad (2.40b)$$

Equations (2.39)–(2.40) represent the neatest form of this Bäcklund map for the Ernst equation.

III. SOME SPECIAL SOLUTIONS

In order to implement the construction of solutions using the Bäcklund map of Sec. II we need an initial "seed solution." The simplest solutions of the Ernst equation correspond to purely real E. If we write

$$f = e^U, \quad g = 0, \quad (3.1)$$

we discover that the function U must satisfy Laplace's equation

$$\partial^2_\rho U + \rho^{-1}\partial_\rho U + \partial_z^2 U = 0. \quad (3.2)$$

Equations (2.25) become, with $\lambda = ue^{-U}$,

$$\partial_z q = -2u(\rho\partial_\rho U), \tag{3.3a}$$

$$\partial_\rho q = 2u(\rho\partial_z U), \tag{3.3b}$$

and the existence of q is guaranteed by the integrability condition, which is (3.2). To obtain explicit solutions, we must solve (3.2). There are many ways we can solve (3.2). For example, we can separate (3.2) in normal axial coordinates by seeking a solution of the form

$$U = R(\rho)Z(z). \tag{3.4}$$

The Laplace equation (3.2) can then be split into the equations

$$R'' + \rho^{-1}R' + \sigma^2 R = 0, \tag{3.5a}$$

$$Z'' - \sigma^2 Z = 0. \tag{3.5b}$$

Given a solution to (3.5) with $\sigma^2 \neq 0$, a solution to (3.3) is given by

$$q = -[2\sigma^{-2}\rho R'(\rho)Z'(z) + C] \ (\sigma \neq 0), \tag{3.6a}$$

and such solutions may clearly be generalized by the linear superposition of solutions to (3.2). When $\sigma = 0$, the solution to (3.5) is given by

$$R = (C_1 \ln\rho + C_2), \quad Z = (k_1 z + k_2), \tag{3.7}$$

and the solution for q is given by

$$q = -uc_1(k_1 z^2 + 2k_2 z) + uk_1 c_2 \rho^2 + uc_1 k_1 \rho^2[\ln\rho - \tfrac{1}{2}] + c. \tag{3.8}$$

A simple example is provided by

$$U = z, \tag{3.9}$$

for which $c_1 = k_2 = 0$, $c_1 = k_1 = 1$, and q is given by

$$q = (u\rho^2 + c). \tag{3.10}$$

The functions λ and α are then given by

$$\lambda = ue^{-z}, \quad \alpha = -e^z(c + u\rho^2)^{-1}. \tag{3.11}$$

If we set $u = 1$ and $c = 0$, the Bäcklund transformation of the previous section generates the solution

$$r = (2\rho^2 - 4\cosh^2 z)(\rho^2 - 4\cosh^2 z)^{-1}, \tag{3.12a}$$

$$s = (-4\sinh z\rho^2)(\rho^2 - 4\cosh^2 z)^{-1}, \tag{3.12b}$$

$$b = \tfrac{1}{2}\tanh z(\rho^2 + 4\cosh^2 z)(\rho^2 - 4\cosh^2 z)^{-1}, \tag{3.12c}$$

$$a = -\rho^2 \operatorname{sech} z(\rho^2 - 4\cosh^2 z)^{-1}, \tag{3.12d}$$

which corresponds to the Ernst potential

$$E = \rho^2 \operatorname{sech} z\,(\rho^2 - 4\cosh^2 z) - i\rho^4 \tanh z. \tag{3.13}$$

The scalar fields of the Bogomolny equation are given by

$$\phi_1 = \frac{-\rho^2 \sec z}{(\rho^2 - 4\cosh^2 z)}, \quad \phi_2 = \frac{\tanh z\,(\rho^2 + 4\cosh^2 z)}{(\rho^2 - 4\cosh^2 z)}. \tag{3.14}$$

An alternative approach is to separate the Laplace equation (3.2) in some alternative orthogonal coordinate system. For example, we may use prolate spheroidal coordinates x and y defined by

$$\rho = (x^2 - 1)^{1/2}(1 - y^2)^{1/2}, \quad z = xy \quad (1 \leqslant x < \infty, \ -1 \leqslant y \leqslant 1). \tag{3.15}$$

For an initial solution of the form

$$f = e^U, \quad g = 0, \quad U = U(x,y), \tag{3.16}$$

Eq. (2.25) takes the form

$$\partial_x q = 2u(1 - y^2)\partial_y U, \tag{3.17a}$$

$$\partial_y q = -2u(x^2 - 1)\partial_x U, \tag{3.17b}$$

and the Laplace equation becomes

$$\partial_y[(1 - y^2)\partial_y U] + \partial_x[(x^2 - 1)\partial_x U] = 0. \tag{3.18}$$

If we seek solution in form

$$U = X(x)Y(y), \tag{3.19}$$

Eq. (3.18) may be decoupled into the pair of equations

$$\partial_x[(x^2 - 1)\partial_x X] + k^2 X = 0, \tag{3.20a}$$

$$\partial_y[(1 - y^2)\partial_y Y] - k^2 Y = 0. \tag{3.20b}$$

For $k^2 \neq 0$ a solution to (3.17) is given by

$$q = -2k^{-2}u(x^2 - 1)(1 - y^2)X'(x)Y'(y) + c. \tag{3.21}$$

As in the previous case, this may be generalized by using the linear superposition principle for (3.18).

When $k = 0$, Eqs. (3.20) have the solutions

$$x = \delta_1 \ln\left(\frac{x-1}{x+1}\right) + k_1, \quad y = \delta_2 \ln\left(\frac{1+y}{1-y}\right) + k_2, \tag{3.22}$$

which correspond to the Weyl metrics

$$f = K\left(\frac{x-1}{x+1}\right)^{\delta_1}\left(\frac{1+y}{1-y}\right)^{\delta_2}. \tag{3.23}$$

The corresponding form of q is given by

$$q = 4u(\delta_2 x - \delta_1 y) + c. \tag{3.24}$$

The initial solution to the Bogomolny equation for solutions to the Ernst equation of the form (3.16) is given in these coordinates by

$$r_0 = \tfrac{1}{2}[(x^2 - 1)(1 - y^2)/(x^2 - y^2)](x\partial_x U - y\partial_y U), \tag{3.25a}$$

$$b_0 = \tfrac{1}{2}[1/(x^2 - y^2)][y(x^2 - 1)\partial_x U + x(1 - y^2)\partial_y U], \tag{3.25b}$$

$$a_0 = 0, \tag{3.25c}$$

$$s_0 = 0. \tag{3.25d}$$

As an example consider the solution corresponding to $\delta_1 = k_2 = u = 1$, $\delta_2 = k_1 = 0$. This gives the initial solution to the Bogomolny equations expressed by

$$r_0 = x(1 - y^2)/(x^2 - y^2), \quad b_0 = y/(x^2 - y^2), \quad a_0 = s_0 = 0. \tag{3.26}$$

From Eqs. (2.27)–(2.33), we obtain

$$r = \left[\frac{\lambda^2 - 1}{\lambda^2 + 1}\left(\frac{1+\Delta}{1-\Delta}\right)r_0 - \frac{4\lambda\alpha\rho^2}{(1-\Delta)}b_0 - \frac{\Delta}{1-\Delta}\right], \tag{3.27a}$$

$$b = \left[\frac{\lambda^2 - 1}{\lambda^2 + 1}\left(\frac{1+\Delta}{1-\Delta}\right)b_0 + \frac{4\lambda\alpha r_0}{1-\Delta}\right], \tag{3.27b}$$

$$s = \left[\frac{-4\lambda}{\lambda^2 + 1}\left(\frac{1+\Delta}{1-\Delta}\right)r_0 - \frac{4\alpha\rho^2(\lambda^2 - 1)}{(1-\Delta)}b_0\right], \tag{3.27c}$$

$$a = \left[\frac{4\alpha(\lambda^2 - 1)}{1-\Delta}r_0 - \frac{4\lambda}{\lambda^2 + 1}\left(\frac{1+\Delta}{1-\Delta}\right)b_0 - \frac{2\alpha(1+\lambda^2)}{1-\Delta}\right], \tag{3.27d}$$

where $\Delta = \alpha^2\rho^2(1+\lambda^2)^2$.

For the initial solution (3.25) the functions λ, α, and Δ are given by

$$\lambda = \left(\frac{x+1}{x-1}\right),\quad \alpha = \frac{1}{4y}\left(\frac{x-1}{x+1}\right),$$
$$\Delta = \frac{(x^2+1)^2}{(x^2-1)}\,\frac{(1-y^2)}{4y^2}. \tag{3.28}$$

If we write $V = \ln f$, we easily find that

$$V_x = \frac{4x}{x^2-1}\left(\frac{1+\Delta}{1-\Delta}\right) - \frac{2x}{x^2-1}\,\frac{\Delta}{1-\Delta} = -\frac{\Delta_x}{1-\Delta} + \left[\log\frac{x^2-1}{x^2+1}\right]_x, \tag{3.29a}$$

$$V_y = \frac{2}{y}\,\frac{1}{(1-\Delta)} + \frac{2y}{1-y^2}\,\frac{\Delta}{1-\Delta} = \frac{-\Delta_y}{1-\Delta} + \frac{2}{y}. \tag{3.29b}$$

This has the general solution

$$V = \log[(1-\Delta)y^2(x^2+1)^{-1}(x^2-1)K], \tag{3.30}$$

where K is a constant. The corresponding function f is therefore given by

$$f = K\,[4y^2(x^2-1) - (x^2+1)^2(1-y^2)]/4(x^2+1). \tag{3.31}$$

The imaginary part of the Bäcklund transformed Ernst potential g satisfies the equations

$$g_x = \{\tfrac{1}{2} - \tfrac{1}{2}y^2[(x^4+6x^2-3)/(x^2+1)^2]\}K, \tag{3.32a}$$
$$g_y = y[(3x-x^3)/(x^2+1)]K. \tag{3.32b}$$

These equations are easily solved to give

$$g = K\{\tfrac{1}{2}y^2[(3x-x^3)/(x^2+1)] + \tfrac{1}{2}x\} + c. \tag{3.33}$$

Therefore, the Bäcklund transform of the initial solution

$$E = (x-1)/(x+1) \tag{3.34}$$

is given by

$$E = K\left\{\frac{[4(x^2-1)y^2 - (x^2+1)^2(1-y^2)] + i[2y^2(3x-x^3) + 2x(x^2+1)]}{4(x^2+1)}\right\} + c. \tag{3.35}$$

If we consider special solution to the Ernst equation having the form

$$E = e^{\delta z}[F(\rho) + iG(\rho)], \tag{3.36}$$

where δ is a nonzero constant, then the Ernst equations reduce to

$$FF'' - (F')^2 + \rho^{-1}FF' = -(G')^2 - \delta^2G^2, \tag{3.37a}$$
$$FG'' - 2G'F' + \rho^{-1}FG' = \delta^2FG. \tag{3.37b}$$

In Ref. 8 we have shown that the deformation problem (2.1) reduces to a monodromy problem for these ordinary differential equations. Various parametrizations of F and G lead to Painlevé equations of types III and V. In this final section we will show that the Bäcklund transformation of the previous sections gives rise to a Bäcklund transformation for this system of ordinary differential equations also.

In order to maintain the special form (3.36), we set $c_0 = 0$ in (2.24) and choose

$$\lambda_0 = -G_0F_0^{-1}. \tag{3.38}$$

As we require $\lambda_0 \neq 0$ this imples $G_0 \neq 0$. Equations (2.25) then provide the solution

$$q = -\rho\delta^{-1}[(1 - G_0^2F_0^{-2})G_0' - 2F_0'G_0F_0^{-1}]e^{\delta z}, \tag{3.39}$$

from which it follows that α_0 and Δ_0 are given by

$$\alpha = \rho^{-1}\delta F_0[(1-G_0^2F_0^{-2})G_0' - 2F_0'G_0F_0^{-1}]^{-1} \equiv \rho^{-1}\delta F_0\Lambda_0^{-1}, \tag{3.40}$$

$$\Delta_0 = \delta^2F_0^{-2}(G_0^2 + F_0^2)\Lambda_0^{-2}. \tag{3.41}$$

If the Bäcklund transformed Ernst potential is written

$$\tilde{E} = e^{\delta z}[\tilde{F}(\rho) + i\tilde{G}(\rho)] \tag{3.42}$$

we obtain from (2.39)

$$\tilde{F} = F_0(1-\Delta_0)\alpha_0^{-2}(\lambda_0^2+1)^{-1} \tag{3.43a}$$

and from (2.40b) we obtain the companion equation

$$\tilde{G} = \frac{F_0}{(\lambda^2+1)}\left[G_0F_0^{-1}\left(\frac{\lambda^2-1}{\lambda^2+1}\right) - \left(\frac{2\lambda}{\lambda^2+1}\right)\right](1+\Delta) - \frac{4\rho\delta^{-1}}{\alpha(\lambda^2+1)}\{\lambda G_0' + \tfrac{1}{2}(\lambda^2-1)F_0'\} - 2\rho\delta^{-2}\Lambda_0. \tag{3.43b}$$

Equations (3.43) define a Bäcklund transformation for Eqs. (3.37).

[1] E. Bogomolny, Sov. J. Nucl. Phys. **24**, 449 (1976).
[2] N. S. Manton, Nucl. Phys. B **126**, 525 (1977).
[3] P. Forgács, Z. Horváth, and L. Palla, Phys. Rev. Lett. **15**, 505 (1980).
[4] E. J. Ernst, Phys. Rev. **167**, 1175 (1968).
[5] B. K. Harrison, Phys. Rev. Lett. **41**, 1197 (1978); D. Kramer and G. Neugebaur, Phys. Lett. A **75**, 259 (1980).
[6] V. E. Zakharov and V. A. Belinskii, Zh. Eksp. Teor. Fiz. **75**, 1955 (1978) [Sov. Phys. JETP **48**, 985 (1978)].
[7] H. C. Morris and R. K. Dodd, Phys. Lett. A **75**, 20 (1980); *Springer Lecture Notes in Mathematics* **810**, edited by R. Martini (Springer–Verlag, Berlin, 1980), p. 63; "Linear deformation problems for the Ernst equation," J. Math. Phys. **23**, 1131 (1982).
[8] R. K. Dodd and H. C. Morris, "Some Classes of Equations Which Give Special Solutions to the Ernst Equation and Their Solution," Proc. R. Irish Acad. (to be published).

SU(3) symmetry of the equations of unidimensional gas flow, with arbitrary entropy distribution

B. Gaffet[a)]
Department of Astronomy, Kyoto University, Kyoto, Japan

(Received 4 February 1983; accepted for publication 5 August 1983)

We have shown in an earlier work that, assuming a particular class of equations of state, the Euler equations of one-dimensional gas flow are invariant under an SU(3) group of transformations, and in fact admit of a Lie group of symmetry of infinite order; they, therefore, possess an infinite number of conservation laws. We show in the present work that the SU(3) symmetrical formalism still brings about tremendous simplification and analytical order in the most general case where the equation of state is arbitrary. The six characteristic equations assume a vector form and relate two conjugate, three-dimensional vectors **U** and **X**. The SU(3) symmetry is only broken to a minor extent through the occurrence of a multiplicative factor Γ in the equations. The conservation laws take the form of the Cauchy integrability condition for the elements of a traceless second rank tensor ϵ_{ij} and, taken all together, form an SU(3) octet; in the most general case, however, there exist four conservation laws only (five if the gas is monatomic) as a result of the breaking of symmetry. Application of these results to the theory of self-similar flow is also discussed. Finally, we show the invariance of the equations of monatomic gas flow under Lorentz transformations in a three-dimensional Minkowski space; that raises the question of whether a geometrical relation may exist between the Minkowski light cones and characteristics.

PACS numbers: 02.30.Jr, 02.20.Sv, 03.40.Gc

I. INTRODUCTION

Previous studies[1,2] have revealed the existence of three fundamental invariance transformations of the Euler equations of one-dimensional gas dynamics; in a later work,[3] it was shown that the existence of these symmetries entails SU(3) invariance of the equations, assuming a particular class of entropy distribution. These are hidden symmetries,[4] whose existence cannot be derived from general principles, and can only be brought to light through careful study of the Euler equations. Still, the determination of these symmetries is no less important than in the case of other equations occurring in mathematical physics; we show in the present paper that the fundamental, hidden symmetry is SU(3). Preliminary results indicate that the extension to N-dimensional gas flow may be feasible, at least in the case of spherical symmetry.

II. THE THREE BASIC INVARIANCE TRANSFORMATIONS

A. The system (S) of characteristic equations

We start with the fundamental characteristic relations

$$\partial_\alpha M = -\rho c\, \partial_\alpha t, \quad \partial_\beta M = +\rho c\, \partial_\beta t, \tag{2.1}$$

which may be viewed as defining the characteristic coordinates α, β vs the Lagrangian coordinates $M = \int \rho\, dr$ and t. Here, symbols ∂_α, ∂_β indicate partial derivatives $\partial/\partial\alpha$, $\partial/\partial\beta$, and M, t, ρ, c, have their usual meaning of mass, time, density, and sound velocity; they are, anyway, thoroughly defined through the present and the following equations.

The relations (2.1) are thus equivalent to the system

$$\frac{\partial_\alpha}{\partial_\alpha t} = \partial_t - \rho c\, \partial_M, \qquad \frac{\partial_\beta}{\partial_\beta t} = \partial_t + \rho c\, \partial_M, \tag{2.2}$$

obtained by solving for the partial derivatives ∂_α, ∂_β vs ∂_t, ∂_M.

Equations (2.2) clearly define α, β as characteristic coordinates, when ρ, c are interpreted as the density ($\rho = \partial M/\partial r|_t$) and sound velocity. There follows two more relations:

$$\partial_\alpha r = (v - c)\,\partial_\alpha t, \qquad \partial_\beta r = (v + c)\,\partial_\beta t, \tag{2.3}$$

where r is the position coordinate. Taking account of equations (2.2), that new system also reads

$$\partial_t r = v, \quad \partial_M r = 1/\rho. \tag{2.4}$$

We shall write the equation of motion in the form

$$\partial_\alpha P = (\gamma P/c)\partial_\alpha v, \quad \partial_\beta P = -(\gamma P/c)\partial_\beta v, \tag{2.5}$$

where P and γ are to be interpreted as the pressure and adiabatic index. Versus coordinates M, t, that is

$$\partial_t P + \gamma\rho P\, \partial_M v = 0, \qquad \partial_t v + (\rho c^2/\gamma P)\partial_M P = 0. \tag{2.6}$$

The first is the continuity equation, which, as is well known, ensures integrability of the space coordinate r, defined by the pair of equations (2.3); the second is to be interpreted as the Euler equation of motion.

[a)] On leave of absence from Centre d'Etudes Nucléaires de Saclay, Service d'Astrophysique, Gif-sur-Yvette, France.

Finally we introduce the equation of state:

$$PM^b/\rho^\gamma = \text{const}\,, \tag{2.7}$$

assuming a power-law entropy distribution of arbitrary index b. It turns out to be convenient to rewrite it in the form

$$\rho \approx P^{1/\gamma} M^{(1+1/\gamma')}$$

where $\gamma' = \gamma/(b-\gamma)$.

We thus obtain a system (S) of seven equations: (2.1), (2.3), (2.5) and (2.7), relating seven dependent variables: M, t, P, v, ρ, c, r. The system, however, is not complete: as already noted, the continuity equation, derived from Eqs. (2.5), ensures compatibility of the pair of equations (2.3), which are therefore not independent. A complete system is obtained by adjoining to (S) the subsidiary equation

$$\gamma P/\rho c^2 = 1\,. \tag{2.8}$$

The Euler equation (2.6) then assumes its usual form.

However, we need not introduce at first that additional constraint; instead, we directly proceed to the study of the incomplete system (S). Any results arising from that study will remain valid when the subsidiary equation is included.

B. Three invariance transformations of the system (S)

We introduce three transformations, defined by the following sets of transformation formulae for the seven dependent variables M, t, P, v, ρ, c, r:

1. Transformation (T')

$$\begin{aligned} &M' = P, \quad t' = v, \quad P' = M, \quad v' = t,\\ &\rho' = (\gamma/\gamma')(\rho P/M)\,,\\ &c' = -\gamma' M/\rho c, \quad r' = vt - r - \gamma' M/\rho\,. \end{aligned} \tag{2.9}$$

The two indices γ, γ' are exchanged by the transformation. It is readily checked that $(T')^2$ is the identity.

It will prove convenient to introduce two more notations:

$$v^* \equiv vt - r, \quad \psi \equiv r + \gamma' M/\rho,$$

in terms of which the following relations hold:

$$(v^*)' = \psi, \quad \psi' = v^*\,.$$

2. Transformation (T)*

$$\begin{aligned} &M^* = -M, \quad t^* = 1/t, \quad P^* = Pt^\gamma,\\ &(v)^* = v^*, \quad \rho^* = \rho t, \quad c^* = ct, \quad r^* = -r/t\,. \end{aligned} \tag{2.10}$$

The two indices remain invariant: $\gamma^* = \gamma$, $(\gamma')^* = \gamma'$.

Again, $(T^*)^2 = 1$. Hence, $(v^*)^* = v$; also note that $\psi^* = -\psi/t$.

3. The space translations ($\mathfrak{T}_r^h$)

They are characterized by the following trivial formulae:

$$\begin{aligned} &M' = M; \quad t' = t; \quad P' = P; \quad v' = v; \quad \rho' = \rho;\\ &c' = c; \quad r' = r + h\,, \end{aligned} \tag{2.11}$$

where h is the amount of the translation.[5] Again, γ and γ' remain invariant.

By direct substitution, it may be checked that all three are invariance transformations of the system (S). In fact, (T') and $(\mathfrak{T}_r)$ are invariance transformations for the complete system as well; but (T^*) is not, unless the gas is assumed to be monatomic [$\gamma = (N+2)/N = 3$, under the one-dimensional assumption]. Indeed, if we assume that relation (2.8) holds, it becomes, after transformation, $\gamma P^*/\rho^* c^{*2} = t^{(\gamma-3)}$, which is not unity unless $\gamma = 3$. [That also proves the fact that (2.8) is an independent equation and that, accordingly, the system obtained by adjoining it to (S) is a complete system.]

We should like to mention here that the invariance transformations that we consider (the trivial ones, such as $(\mathfrak{T}_r)$, excepted) bear no relation with those described in earlier literature, and whose main properties are summarized in Appendix B.

C. The SU(3) algebra of infinitesimal transformations

Viewed as operators acting on the six-dimensional manifold of (M, t, P, v, ρ, r), the three transformations are completely determined by the three sets of transformation formulae (2.9)–(2.11). Since one of them, the space translation $(\mathfrak{T}_r^h)$, includes an arbitrary parameter h, a continuous Lie group of symmetry is generated. We presently show that *the group structure is SU(3)*.

First, we construct a chain of six independent generators $G_1,\ldots,G_6$, starting from the generator G_1 of space translations. Defining G_1 by

$$\mathfrak{T}_r^h = 1 + hG_1\,,$$

the six generators are obtained according to the scheme

$$G_1 \overset{T^*}{\leftrightarrow} G_2 \overset{T'}{\leftrightarrow} G_3 \overset{T^*}{\leftrightarrow} G_4 \overset{T'}{\leftrightarrow} G_5 \overset{T^*}{\leftrightarrow} G_6\,,$$

which means that $G_2 = T^* G_1 T^*, \ldots, G_6 = T^* G_5 T^*$. The chain cannot be further extended, owing to the relations

$$T' G_1 T' = -G_1\,, \quad (T'T^*)^6 = 1\,,$$

which may be checked directly from the given sets of transformation formulae.

The transformation formulae defining these six generators read

G_1: $\delta M = \delta t = \delta P = \delta v = \delta\rho = 0$, $\delta r = -1$,

G_2: $\delta M = \delta t = \delta P = \delta\rho = 0$, $\delta v = +1$, $\delta r = +t$,

G_3: $\delta M = \delta P = \delta v = \delta\rho = \delta r = 0$, $\delta t = +1$,

G_4: $\delta M = 0$, $\delta t = -t^2$, $\delta P = \gamma Pt$, $\delta v = v^*$, $\delta\rho = \rho t$, $\delta r = -rt$,

G_5: $\delta M = \gamma' Mv$, $\delta t = r + \gamma' M/\rho$, $\delta P = 0$, $\delta v = -v^2$, $\delta\rho = (\gamma'+1)\rho v$, $\delta r = \gamma' Mv/\rho$,

G_6: $\delta M = \gamma' M v^*$, $\delta t = t(r + \gamma' M/\rho)$, $\delta P = -\gamma P \times (r + \gamma' M/\rho)$, $\delta r = r^2 + \gamma' Mvt/\rho$,

where δ symbolizes the variation (of a given variable) introduced by the generator.

Applying the general definition of commutators, two more generators can be constructed:

$$G_7 = [G_3, G_4] \quad \text{and} \quad G_8 = [G_2, G_5];$$

they are characterized by the following transformation formulae:

$$G_7:\quad \delta M = 0,\quad \delta t = -2t,\quad \delta P = \gamma P,\quad \delta v = v,$$
$$\delta\rho = \rho,\quad \delta r = -r;$$
$$G_8:\quad \delta M = \gamma' M,\quad \delta t = t,\quad \delta P = 0,\quad \delta v = -2v,$$
$$\delta\rho = (\gamma' + 1)\rho,\quad \delta r = -r.$$

G_7 and G_8 are readily interpreted as the generators of *scale transformations.*

Now, all commutators $[G_i, G_j]$ $(i,j = 1,\ldots,8)$ are found to be expressible in terms of linear combinations of the eight generators, so that they form *a Lie algebra of order 8* (Refs. 6 and 7). Explicit computation reveals that the commutation relations follow the SU(3) pattern; they are in fact the commutation relations of 3×3 matrices, identified with the eight generators as follows:

$$G_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\quad G_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$
$$G_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\quad G_4 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$G_5 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},\quad G_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$
$$G_7 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},\quad G_8 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

III. IRREDUCIBLE REPRESENTATIONS. CONSTRUCTION OF TWO VECTOR SPACES OF DIMENSION THREE: $X = (P^{1/\gamma}t,\ P^{1/\gamma}\psi,\ P^{1/\gamma})$ AND $U = (-M^{1/\gamma'}v,\ M^{1/\gamma'},\ M^{1/\gamma'}v^*)$

A. Identification of the linear, irreducible representations

The SU(3) group generated by the three fundamental symmetries (T'), (T^*) and $(\mathfrak{T}_r)$, operates—nonlinearly—on the six-dimensional manifold of parameters M, t, P, v, ρ, r.[8] It is known, however, that irreducible representations of SU(3) below dimension eight are either one- or three-dimensional; therefore, the representation derived in the preceding section is reducible, and, being of dimension six, it is natural to expect that it decomposes into a product of two three-dimensional representations. Thus we look for three nonlinear functions $X(p_k)$, $Y(p_k)$, $Z(p_k)$ of six independent variables p_k $(k = 1,\ldots,6)$, symbolizing the six parameters M, t, P, v, ρ, r; the three functions are required to constitute a linear, three-dimensional representation of SU(3), so that they are operated upon by each generator G_i according to the matrix law

$$\begin{pmatrix} \delta X \\ \delta Y \\ \delta Z \end{pmatrix} = H_i \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \tag{3.1}$$

where the H_i are eight 3×3 traceless matrices, linearly independent. Explicitly, we have $\delta X = \Sigma_k (\partial X/\partial p_k)\delta p_k$, etc., and since the transformation laws of the variables p_k are given, Eq. (3.1) constitutes a set of three linear first-order partial differential equations (p.d.e.) for the three unknown functions X, Y, Z.

There are eight such sets of equations—one for each generator—so that we have a system of 24 linear equations for three unknowns. There exist systematic methods for deriving the solutions of such overdetermined linear systems, when they are compatible.[9] In the present case, it is possible to arrive at a solution directly without actual calculation. First we observe that the three variables M, v, v^* do provide a three-dimensional, though still nonlinear, representation, as follows:

$$G_1:\quad \delta M = \delta v = 0,\quad \delta v^* = +1;$$
$$G_2:\quad \delta M = \delta v^* = 0,\quad \delta v = +1;$$
$$G_3:\quad \delta M = \delta v = 0,\quad \delta v^* = v;$$
$$G_4:\quad \delta M = \delta v^* = 0,\quad \delta v = v^*;$$
$$G_5:\quad \delta M = \gamma' M v,\quad \delta v = -v^2,\quad \delta v^* = -vv^*;$$
$$G_7:\quad \delta M = 0,\quad \delta v = v,\quad \delta v^* = -v^*;$$
$$G_8:\quad \delta M = \gamma' M,\quad \delta v = -2v,\quad \delta v^* = -v^*.$$

The representation becomes linear through the following choice of the three unknown functions:

$$\begin{aligned} U &= -M^{1/\gamma'}v, \\ V &= M^{1/\gamma'}, \\ W &= M^{1/\gamma'}v^*. \end{aligned} \tag{3.2}$$

It is easily seen that each generator G_i indeed operates linearly on U, V, W according to the law

$$\begin{pmatrix} \delta U \\ \delta V \\ \delta W \end{pmatrix} = K_i \begin{pmatrix} U \\ V \\ W \end{pmatrix},$$

where the K_i are traceless matrices, as required.

Another, complementary, three-dimensional representation may be constructed starting from the two variables P and t; in the process, a third variable

$$\psi = r + \gamma' M/\rho \tag{3.3}$$

comes into play, and it is found that the representation in terms of P, t, ψ is irreducible. It can also be made linear, through the following variable transformation:

$$\begin{aligned} X &= P^{1/\gamma}t, \\ Y &= P^{1/\gamma}\psi, \\ Z &= P^{1/\gamma}, \end{aligned} \tag{3.4}$$

so that the generators operate linearly on X, Y, Z according to the law

$$\begin{pmatrix} \delta X \\ \delta Y \\ \delta Z \end{pmatrix} = H_i \begin{pmatrix} X \\ Y \\ Z \end{pmatrix},$$

where the H_i are eight other traceless 3×3 matrices (see Sec. VD).

In conclusion, the original, nonlinear, six-dimensional SU(3) representation *has been reduced into a product of two* *linear, three-dimensional representations*. In the process of decomposition, *two complementary three-dimensional vector spaces have emerged*:

$$\mathbf{U} = (U,V,W) \quad \text{and} \quad \mathbf{X} = (X,Y,Z)\,.$$

We may now interpret the SU(3) symmetry of the subsystem (S) as *arising from the equivalence of all choices of base vectors* in the spaces $\{\mathbf{X}\}$ or $\{\mathbf{U}\}$.

B. The gasdynamical equations as vector equations in the spaces {X} and {U}

In order to obtain the new form of the equations, the six parameters M, t, P, v, ρ, r must now be expressed in terms of $\mathbf{X}$ and $\mathbf{U}$ as follows:

$$M = V^{\gamma'}, \quad v = -U/V, \quad v^* = W/V,$$
$$P = Z^{\gamma}, \quad t = X/Z, \quad \psi = Y/Z\,. \tag{3.5}$$

Taking account of the definitions of v^*, ψ, and of the equation of state,

$$r = (VY-1)/VZ, \quad \rho = \gamma' Z V^{(\gamma'+1)}\,.$$

Thus, the fundamental relation $\partial_\alpha M = -\rho c\, \partial_\alpha t$ becomes

$$\partial_\alpha V = -\Gamma^{-1}(Z\,\partial_\alpha X - X\,\partial_\alpha Z)\,, \tag{3.6}$$

where the coefficient Γ is related to the seventh parameter c (the sound velocity) by

$$\Gamma = Z/V^2 c\,. \tag{3.7}$$

Then, *as a result of the symmetry, the following vector equation holds*:

$$\partial_\alpha \mathbf{U} = -\Gamma^{-1}\mathbf{X}\wedge\partial_\alpha \mathbf{X}\,. \tag{3.8}$$

The corresponding equation vs variable β is obtained by changing the sign of Γ:

$$\partial_\beta \mathbf{U} = +\Gamma^{-1}\mathbf{X}\wedge\partial_\beta \mathbf{X}\,. \tag{3.9}$$

It may be checked directly that the above six equations are equivalent to the six characteristic equations of the system (S).

If we start from the other characteristic equation (2.5), $\partial_\alpha P = (\gamma P/c)\,\partial_\alpha v$, we obtain a symmetrical (conjugate) system of equations determining the differentials of $\mathbf{X}$:

$$\partial_\alpha \mathbf{X} = +\Gamma\,\mathbf{U}\wedge\partial_\alpha \mathbf{U}\,,$$
$$\partial_\beta \mathbf{X} = -\Gamma\,\mathbf{U}\wedge\partial_\beta \mathbf{U}\,, \tag{3.10}$$

where, as in Eqs. (3.8) and (3.9), the symbol $\wedge$ represents the exterior product in a three-dimensional space.

As a consequence of an elementary property of the exterior product, the following relations hold:

$$\mathbf{X}\cdot\partial_\alpha\mathbf{U} = \mathbf{X}\cdot\partial_\beta\mathbf{U} = \mathbf{U}\cdot\partial_\alpha\mathbf{X} = \mathbf{U}\cdot\partial_\beta\mathbf{X} = 0; \tag{3.11}$$

hence,

$$\mathbf{X}\cdot d\,\mathbf{U} = \mathbf{U}\cdot d\,\mathbf{X} = 0. \tag{3.12}$$

The above equation expresses the property that $\mathbf{U}$ *is the vector normal to the integral surface in the* $\mathbf{X}$*-space*, and reciprocally $\mathbf{X}$ is the direction normal to the corresponding surface in the U-space. As a result of Eq. (3.12), *the scalar product* $\mathbf{U}\cdot\mathbf{X}$ *is a constant*, which we can always set equal to unity without loss of generality, since the choice of units is arbitrary:

$$\mathbf{U}\cdot\mathbf{X} = 1\,. \tag{3.13}$$

The above expression constitutes the new form of the equation of state, which is the seventh and last equation in the system (S).

C. Some properties of the vector equations (3.8) and (3.10)

As stated, the set of equations (3.8), (3.9), (3.13) is equivalent to the system (S), and the same is true of the conjugate system of equations (3.10), (3.13). As these equations involve $\mathbf{U}$ and $\mathbf{X}$ simultaneously, it is of interest to show that $\mathbf{U}$, or $\mathbf{X}$, may be explicitly eliminated, yielding an equation involving one vector ($\mathbf{X}$ or $\mathbf{U}$) only. From the property of orthogonality of $\mathbf{U}$ and $\mathbf{X}$ to the integral surfaces, together with the normalizing constraint (3.13), there follows

$$\mathbf{U} = \partial_\alpha\mathbf{X}\wedge\partial_\beta\mathbf{X}/(\mathbf{X},\partial_\alpha\mathbf{X},\partial_\beta\mathbf{X})\,, \tag{3.14}$$

$$\mathbf{X} = \partial_\alpha\mathbf{U}\wedge\partial_\beta\mathbf{U}/(\mathbf{U},\partial_\alpha\mathbf{U},\partial_\beta\mathbf{U})\,, \tag{3.15}$$

where $(\mathbf{a},\mathbf{b},\mathbf{c})$ denotes the mixed product $\mathbf{a}\cdot\mathbf{b}\wedge\mathbf{c}$. In addition we derive from Eqs. (3.8)–(3.10) the relations

$$\partial_\alpha\mathbf{X}\cdot\partial_\beta\mathbf{U} = \partial_\beta\mathbf{X}\cdot\partial_\alpha\mathbf{U} = \Gamma(\mathbf{U},\partial_\alpha\mathbf{U},\partial_\beta\mathbf{U})$$
$$= -\Gamma^{-1}(\mathbf{X},\partial_\alpha\mathbf{X},\partial_\beta\mathbf{X}); \tag{3.16}$$

therefore, the two mixed products are proportional:

$$(\mathbf{U}-\partial_\alpha\mathbf{U},\partial_\beta\mathbf{U}) = -\Gamma^{-2}(\mathbf{X},\partial_\alpha\mathbf{X},\partial_\beta\mathbf{X}). \tag{3.17}$$

The simplest way to eliminate $\mathbf{X}$ is through differentiation of Eqs. (3.8) and (3.9), which yields

$$2\Gamma\partial^2_{\alpha\beta}\mathbf{U} + \partial_\alpha\Gamma\,\partial_\beta\mathbf{U} + \partial_\beta\Gamma\,\partial_\alpha\mathbf{U}$$
$$= 2\partial_\alpha\mathbf{X}\wedge\partial_\beta\mathbf{X} = 2(\mathbf{X},\partial_\alpha\mathbf{X},\partial_\beta\mathbf{X})\mathbf{U};$$

taking account of the proportionality of the mixed products, we obtain the eliminant

$$2\Gamma\,\partial^2_{\alpha\beta}\mathbf{U} + \partial_\alpha\Gamma\,\partial_\beta\mathbf{U} + \partial_\beta\Gamma\,\partial_\alpha\mathbf{U} + 2\Gamma^2$$
$$\times(\mathbf{U},\partial_\alpha\mathbf{U},\partial_\beta\mathbf{U})\mathbf{U} = 0\,. \tag{3.18}$$

The corresponding equation for $\mathbf{X}$ is

$$2\Gamma\,\partial^2_{\alpha\beta}\mathbf{X} - (\partial_\alpha\Gamma\,\partial_\beta\mathbf{X} + \partial_\beta\Gamma\,\partial_\alpha\mathbf{X}) - 2(\mathbf{X},\partial_\alpha\mathbf{X},\partial_\beta\mathbf{X})\mathbf{X} = 0; \tag{3.19}$$

we may rewrite it in the form

$$\partial^2_{\alpha\beta}(\mathbf{X}/\sqrt{\Gamma}) = Q\,\mathbf{X}\,,$$

with

$$Q = \partial^2_{\alpha\beta}\,\Gamma^{-1/2} + \Gamma^{-3/2}(\mathbf{X},\partial_\alpha\mathbf{X},\partial_\beta\mathbf{X})\,. \tag{3.20}$$

D. The explicit expression of factor Γ and the breaking of symmetry

The above discussion is independent of any assumption concerning the actual expression of factor Γ; but, in order to have a complete system, Γ has to be specified, and it must be realized that by so doing the SU(3) symmetry will not in general remain valid. At the same time, we should like to stress that the breaking of symmetry takes place through that multiplicative factor only and that the vector formalism developed in the preceding section always remains valid; the equations formally retain the SU(3) symmetry in spite of the symmetry breaking. There is no doubt that the equations of gas dynamics present many kinds of hidden regularities, such as the existence of the invariance transformations (T^*), (T'), $(\bar{T})$ discussed in Refs. 1 and 2, or the fact that the self-similar equations can be reduced to canonical forms of unexpectedly simple type (see Sec. IV); the formalism here developed provides a framework in which such regularities become predictable and are readily understood.

Now, concerning the explicit expression of Γ, we recall that it is related to the sound velocity c through Eq. (3.7), where c itself is determined by the subsidiary equation (2.8); in this way we obtain

$$\Gamma^2 = (\gamma'/\gamma)\, V^{(\gamma'-3)}/Z^{(\gamma-3)}. \tag{3.21}$$

The breaking of symmetry occurs as a result of the lack of invariance of Γ under linear transformations of the coordinates U_i or X_i.

For a monatomic gas, Γ is a function of V only, i.e., of the mass coordinate M; Eqs. (3.8) and (3.9) then formally retain the invariance under linear transformations in the **X**-space; but, even then, the breaking of symmetry does occur, since vectors **U** and **X** are not in fact independent.

Finally we point out that the explicit expression of Γ reintroduces **X** (resp. **U**)-dependent terms in the eliminant equations (3.18) and (3.19) derived in the preceding section; still, Eq. (3.18) is an equation for the vector **U** only, if the gas is monatomic; it determines the integral surfaces in the **U**-space.

E. A particular equation of state with exact SU(3) symmetry

When the equation of state assumes the form

$$P \approx \rho^3/M^4, \tag{3.22}$$

that is to say, when $\gamma = \gamma' = 3$, Γ is a mere constant, and the complete system of equations has an exact SU(3) symmetry. The equations assume the following form:

$$\partial_\alpha \mathbf{X} = \mathbf{U}\wedge\partial_\alpha \mathbf{U}, \quad \partial_\alpha \mathbf{U} = -\mathbf{X}\wedge\partial_\alpha \mathbf{X}, \tag{3.23}$$

together with the corresponding equations versus variable β. Elimination of **U** yields the equation

$$\partial^2_{\alpha\beta}\mathbf{X} = (\mathbf{X},\partial_\alpha \mathbf{X},\partial_\beta \mathbf{X})\mathbf{X}. \tag{3.24}$$

We have shown in an earlier work (Ref. 3) that *such an equation presents an infinite number of conservation laws, and thus is essentially integrable*, although we have not been able yet to reduce it to linear form. The present analysis shows that, even though the SU(3) symmetry is broken in the general case, the equations remain in form quite similar to the above Eq. (3.23), which should accordingly be viewed as the archetype of the general Euler equations of gas dynamics. It would be extremely interesting to have a more complete mathematical understanding of the above equation which, as stated, possesses an infinite number of conservation laws; a complete theory of it would probably provide insight on how to deal with the general case where the entropy distribution is arbitrary.

F. The case of non-power-law entropy distributions

It is possible to extend the SU(3) description to the case where the entropy distribution is truly arbitrary, i.e, is not power-law; the crucial step for that purpose is to generalize the definition (3.3) of variable ψ, which enters as a factor in the definition of the component Y: $Y \equiv P^{1/\gamma}\psi$. Assuming an entropy distribution $\sigma(M)$ of completely general form,

$$P/\rho^\gamma = \sigma(M), \tag{3.25}$$

we define ψ through the pair of characteristic equations:

$$\begin{gathered} c\,\partial_\alpha\psi + \psi\partial_\alpha v = v\,\partial_\alpha v^* - v^*\,\partial_\alpha v, \\ -c\,\partial_\beta\psi + \psi\,\partial_\beta v = v\partial_\beta v^* - v^*\,\partial_\beta v. \end{gathered} \tag{3.26}$$

Introducing $\theta \equiv (\psi - r)\rho$, we obtain the equations defining θ:

$$\partial_\alpha\theta + (\theta/\gamma)\partial_\alpha \log\sigma + \partial_\alpha M = 0,$$
$$\partial_\beta\theta + (\theta/\gamma)\partial_\beta \log\sigma + \partial_\beta M = 0,$$

which integrate as

$$\theta = -\Phi(M)/\Phi'(M), \tag{3.27}$$

where

$$\Phi(M) = \int \sigma^{1/\gamma}(M)\, dM. \tag{3.28}$$

As a result, the generalized expressions of ψ and Y read

$$\psi = r + \theta(M)/\rho, \tag{3.29}$$

$$Y = P^{1/\gamma} r - \Phi(M). \tag{3.30}$$

We note that Y is now, like $\Phi(M)$, only determined up to an arbitrary additive constant.

Keeping definitions (3.4) for X and Z, the vector $\mathbf{X} \equiv (X,Y,Z)$ is thus determined. We choose to generalize the V component of the conjugate vector **U** as

$$V = -1/\Phi(M), \tag{3.31}$$

and keep for the remaining components the definitions

$$U = -vV, \quad W = v^*V.$$

Equations (3.8) and (3.9) and their conjugate (3.10) relating vectors **U** and **X** then remain valid for general entropy distributions; the factor Γ is still expressed by Eq. (3.7), where the sound velocity c is, as before, determined by Eq. (2.8).

IV. A NEW COORDINATE SYSTEM CONSTRUCTED FROM VECTORS U AND X, AND ITS APPLICATIONS TO THE THEORY OF SELF-SIMILAR FLOW

A. General formalism

Let us introduce new coordinates ξ, η, ζ defined by

$$\xi = UX, \quad \eta = VY, \quad \zeta = WZ, \tag{4.1}$$

or, in compact form,

$$\xi_i = U_i X_i \quad (i = 1,2,3).$$

All integral surfaces reduce, in these coordinates, to the plane of equation

$$\xi + \eta + \zeta = 1, \tag{4.2}$$

representing the equation of state. We further define

$$H \equiv \Gamma^{-1}XYZ, \quad K \equiv \Gamma UVW \tag{4.3}$$

and note the identity

$$HK \equiv \xi\eta\zeta\,. \tag{4.4}$$

It is straightforward to show that the vector equations (3.8) and (3.10) become respectively

$$\begin{aligned}(H+K)\,\partial_\alpha \log U_i &= H\,\partial_\alpha \log \xi_i - (\xi\wedge\partial_\alpha\xi)_i\,,\\ (H+K)\partial_\alpha \log X_i &= K\,\partial_\alpha \log \xi_i + (\xi\wedge\partial_\alpha\xi)_i\,,\end{aligned} \tag{4.5}$$

where $(\xi\wedge\partial_\alpha\xi)_i$ denotes the ith component of the exterior product $\xi\wedge\partial_\alpha\xi$; thus, surprisingly, the coordinates ξ_i retain some of the properties of a three-dimensional vector.

Taking account of the definition (4.1) and of the properties

$$\begin{aligned}H &\approx XYZ^{(\gamma-1)/2}/V^{(\gamma'-3)/2}\,,\\ K &\approx UV^{(\gamma'-1)/2}W/Z^{(\gamma-3)/2}\,,\end{aligned} \tag{4.6}$$

we obtain an expression for the differentials of H and K,

$$\begin{aligned}(H+K)\,\partial_\alpha H &= HK\,(\partial_\alpha \log\xi + \partial_\alpha \log\eta + \tfrac{1}{2}(\gamma-1)\,\partial_\alpha\log\zeta)\\ &\quad - \tfrac{1}{2}(\gamma'-3)\,H^2\,\partial_\alpha\log\eta + H\Omega\,,\end{aligned} \tag{4.7}$$

$$\begin{aligned}(H+K)\,\partial_\alpha K &= HK\,(\partial_\alpha \log\xi + \tfrac{1}{2}(\gamma'-1)\,\partial_\alpha \log\eta + \partial_\alpha\log\zeta)\\ &\quad - \tfrac{1}{2}(\gamma-3)\,K^2\,\partial_\alpha\log\zeta - K\Omega\,,\end{aligned} \tag{4.8}$$

where $\Omega = \Omega_0 + \Omega_1$,

$$\begin{aligned}\Omega_0 &= \tfrac{1}{2}(\gamma-3)(\xi\,\partial_\alpha\eta - \eta\,\partial_\alpha\xi)\\ &\quad + \tfrac{1}{2}(\gamma'-3)(\zeta\,\partial_\alpha\xi - \xi\,\partial_\alpha\zeta)\,,\\ \Omega_1 &= (\xi\,\partial_\alpha\eta - \eta\,\partial_\alpha\xi) + (\eta\,\partial_\alpha\zeta - \zeta\,\partial_\alpha\eta)\\ &\quad + (\zeta\,\partial_\alpha\xi - \xi\,\partial_\alpha\zeta)\,.\end{aligned}$$

Owing to the relation (4.2), Ω_1 also reads

$$\Omega_1 = (3\xi-1)\,\partial_\alpha\eta - (3\eta-1)\,\partial_\alpha\xi\,.$$

Let us now introduce new coordinates u, v, w related to ξ, η, ζ by a fixed translation

$$\begin{aligned}u &\equiv \xi - 2/(\gamma+\gamma')\,,\\ v &\equiv \eta - (\gamma'-1)/(\gamma+\gamma')\,,\\ w &\equiv \zeta - (\gamma-1)/(\gamma+\gamma')\,,\end{aligned} \tag{4.9}$$

so that, by virtue of the equation of state,

$$u+v+w=0\,. \tag{4.10}$$

In terms of these, we have for the quantity Ω the very simple expression

$$\Omega = \tfrac{1}{2}(\gamma+\gamma')(u\,\partial_\alpha v - v\,\partial_\alpha u)\,. \tag{4.11}$$

Combining Eqs. (4.7) and (4.8) then yields the result

$$\begin{aligned}&\partial_\alpha(H-K) + \tfrac{1}{2}(\gamma'-3)H\ \partial_\alpha\log\eta - \tfrac{1}{2}(\gamma-3)\,K\,\partial_\alpha\log\zeta\\ &\quad = \tfrac{1}{2}(\gamma+\gamma')(u\,\partial_\alpha v - v\,\partial_\alpha u)\,,\end{aligned} \tag{4.12}$$

together with a corresponding equation versus variable β,

$$\begin{aligned}&\partial_\beta(H-K) + \tfrac{1}{2}(\gamma'-3)H\,\partial_\beta\log\eta - \tfrac{1}{2}(\gamma-3)\mathrm{K}\,\partial_\beta\log\zeta\\ &\quad = -\tfrac{1}{2}(\gamma+\gamma')(u\,\partial_\beta v - v\,\partial_\beta u)\,.\end{aligned} \tag{4.13}$$

Remembering the identity (4.4) relating H and K and Eqs. (4.2) and (4.9), the above Eqs. (4.12), (4.13) constitute an incomplete system for the determination of the three unknown functions u, v, and H; in order to have a complete system, it would be necessary to retain Eqs. (4.4) or (4.5) and at least some of the variables U_i or X_i as auxiliary unknowns.

But, in the special case of self-similar flow, the two equations (4.12), (4.13) are sufficient in order to completely determine the solution. We proceed along these lines, in the next subsection, to the treatment of self-similar flow; it will be seen that the above choice of variables provides a remarkably direct way of deriving the self-similar equations, and in addition yields the equations in simple and compact form.

B. The equations of self-similar flow

Owing to the fact that all quantities H, K, ξ, η, ζ or u, v, w involved in the system (4.12), (4.13) are *dimensionless*, the self-similar equations are obtained by merely replacing the partial differential symbols ∂_α, ∂_β by the total differential symbol d[10]; we thus obtain the two equations

$$u\,dv - v\,du = 0\,, \tag{4.14}$$

$$\begin{aligned}&2d\,(H-K) + (\gamma'-3)Hd\log\eta\\ &\quad - (\gamma-3)Kd\log\zeta = 0\,.\end{aligned} \tag{4.15}$$

The first one is immediately integrable as

$$v = mu\,,\quad w = -(m+1)u\,, \tag{4.16}$$

where m is a constant; hence the following result:

Self-similar flows are represented by straight lines in the plane of equation $\xi+\eta+\zeta=1$. *When the equation of state is given* (that is to say, when the indices γ and γ' are fixed), *the straight lines envelop a point, of coordinates*

$$\begin{aligned}&\xi_0 = 2/(\gamma+\gamma')\,,\quad \eta_0 = (\gamma'-1)/(\gamma+\gamma')\,,\\ &\zeta_0 = (\gamma-1)/(\gamma+\gamma')\,.\end{aligned}$$

As a result of Eqs. (4.9) and (4.16), the variables ξ and ζ are linear functions of η, so that the product $HK \equiv \xi\eta\zeta$ is a cubic polynomial g of variable η:

$$HK = g(\eta)\,. \tag{4.17}$$

Therefore, Eq. (4.15) is an ordinary differential equation for the function $H(\eta)$; in the case where the gas is monatomic ($\gamma=3$), the form of the equation becomes simpler:

$$\frac{d}{d\eta}\left[H - \frac{g(\eta)}{H}\right] = -\,\frac{(\gamma'-3)H}{2\eta}\quad(\gamma=3)\,. \tag{4.18}$$

Introducing a new variable Θ,

$$\Theta \equiv H - K\,, \tag{4.19}$$

in terms of which

$$\begin{aligned}H &\equiv (\Theta + \sqrt{\Theta^2+4g})/2\,,\\ K &\equiv (-\Theta + \sqrt{\Theta^2+4g})/2\,,\end{aligned} \tag{4.20}$$

the monatomic self-similar equation also reads

$$\frac{d\ln\Theta}{d\ln\eta} = \frac{-(\gamma'-3)}{4}\left[1 + \sqrt{1+\frac{4g(\eta)}{\Theta^2}}\right]. \tag{4.21}$$

In the particular case where the second index γ' is also equal to 3, the equation reduces to the form $d\Theta/d\eta = 0$; for such an equation of state, the problem of self-similar flow is readily integrable, and the general solution reads

$$\Theta = \text{const}\quad(\gamma=\gamma'=3)\,. \tag{4.22}$$

Going back to the original notations in terms of vectors **U** and **X**, that reads

$$XYZ - UVW = \text{const} \quad (\gamma = \gamma' = 3)\,, \tag{4.23}$$

a simple and symmetrical form.

The above equation of state was the subject of one of our earlier investigations (Ref. 3), in which we showed that it is characterized by the existence of an infinite number of conservation laws; we now see that the corresponding self-similar equations are integrable in closed form, and assume a very remarkable symmetrical form.

V. THE SU(3) PATTERN OF THE SET OF CONSERVATION LAWS

We first consider the case, presented in Sec. IIIE, where the SU(3) symmetry remains exact, which occurs when the equation of state assumes the form (3.22). There are then eight conservation laws forming a complete SU(3) octet that can be derived from the momentum conservation law. The manifestly SU(3)-symmetrical formalism developed in Sec. IIIB makes it possible to formulate these eight conservation laws in a simple and unified way; at the same time a geometrical interpretation emerges, in which the conserved quantities are viewed as "potentials," whose equipotentials are determined by the condition of orthogonality to given families of curves.

We then show (Sec. VC) that the formalism is susceptible to a straightforward generalization to the case of arbitrary polytropes with arbitrary entropy distributions. The number of conservation laws decreases from eight to five for a monatomic gas (four for other polytropes) as a result of the symmetry breaking, but the unified formalism and the geometrical interpretation remain.

A. The case with exact SU(3) symmetry

The physical meaning of the eight conservation laws has already been discussed in an earlier work (Ref. 3); let us recall that the six first, denoted Π, Π^*, E, E^*, σ_0, σ_1, are related, respectively, to momentum, center-of-mass motion, energy, the virial theorem, and the two independent scale transformations leaving the equation of state invariant, whereas the remaining two, here denoted τ_0, τ_1, have no straightforward interpretation. Their precise definition may be found in the same paper, but we rewrite it here for convenience, in characteristic form

$$\begin{aligned}
&\partial_\alpha \Pi = (v + c/3)\, \partial_\alpha M\,,\\
&\partial_\alpha \Pi^* = r\, \partial_\alpha M - t\, \partial_\alpha \Pi\,,\\
&\partial_\alpha E = \tfrac{1}{2}(v^2 + \tfrac{2}{3}vc + \tfrac{1}{3}c^2)\, \partial_\alpha M\,,\\
&\partial_\alpha E^* = -\tfrac{1}{2}r^2\, \partial_\alpha M + rt\, \partial_\alpha \Pi - t^2\, \partial_\alpha E\,,\\
&\partial_\alpha \sigma_0 = r\, \partial_\alpha \Pi - 2t\, \partial_\alpha E\,,\\
&\partial_\alpha \sigma_1 = t\, \partial_\alpha E + \tfrac{1}{2}M(v^2 - 2vc - c^2)\, \partial_\alpha t\,,\\
&\partial_\alpha \tau_0 = 2r\, \partial_\alpha E - 3M(c^3 + c^2 v + cv^2 - \tfrac{1}{3}v^3)\partial_\alpha t,\\
&\partial_\alpha \tau_1 = t\, \partial_\alpha \tau_0 - 6r\, \partial_\alpha \sigma_1 + 6rt\, \partial_\alpha E - 2r^2\, \partial_\alpha \Pi\,.
\end{aligned} \tag{5.1}$$

In the SU(3)-symmetrical notations of Sec. IIIB, the definition of the momentum Π becomes

$$\partial_\alpha \Pi = (Z - 3UV)\partial_\alpha V\,, \tag{5.2}$$

together with a similar equation versus variable β. A more interesting form, which leads to a unified formalism, is as follows:

$$\begin{aligned}
&\partial_\alpha(\Pi + UV^2) = Z\, \partial_\alpha V - V\, \partial_\alpha Z\,,\\
&\partial_\beta(\Pi + UV^2) = -Z\, \partial_\beta V + V \partial_\beta Z\,.
\end{aligned} \tag{5.3}$$

The above expression is of the general form

$$\begin{aligned}
&\partial_\alpha \epsilon_{ij} = X_i\, \partial_\alpha U_j - U_j\, \partial_\alpha X_i\,,\\
&\partial_\beta \epsilon_{ij} = -X_i\, \partial_\beta U_j + U_j\, \partial_\beta X_i\,,
\end{aligned} \tag{5.4}$$

where X_i ($i = 1,2,3$) and U_j ($j = 1,2,3$) stand for X,Y,Z and U,V,W, respectively. Equation (5.3) indicates that the integrability condition of ϵ_{32}, defined by the pair of equations (5.4), is satisfied, since

$$\epsilon_{32} = \Pi + UV^2\,.$$

Owing to the SU(3) symmetry, integrability of ϵ_{32} entails integrability of each of the ϵ_{ij}. As we presently show, only eight of the nine ϵ_{ij} are independent quantities, and these account for the existence of the eight conservation laws.

From the definition (5.4) one immediately deduces

$$\partial_\alpha(\mathrm{Tr}\, \epsilon_{ij}) = \mathbf{X}\cdot\partial_\alpha \mathbf{U} - \mathbf{U}\cdot\partial_\alpha \mathbf{X} = 0\,,$$

as a result of Eqs. (3.11), and similarly $\partial_\beta(\mathrm{Tr}\, \epsilon_{ij}) = 0$; thus the trace is a constant, which we may take to be zero:

$$\mathrm{Tr}\, \epsilon_{ij} = 0\,. \tag{5.5}$$

That is the announced relation constraining the three diagonal elements.

Detailed calculation leads to the following identifications:

$$\begin{aligned}
&\Pi = \epsilon_{32} - UV^2\,,\\
&\Pi^* = -\epsilon_{12} - WV^2\,,\\
&E = -\epsilon_{31} + (1/2V)(U^2V^2 - Z^2)\,,\\
&E^* = -\epsilon_{13} + (1/2V)(X^2 - V^2W^2)\,,\\
&\sigma_0 = (\epsilon_{11} - \epsilon_{33}) + XZ/V + UVW\,,\\
&\sigma_1 = -\epsilon_{11} + (X/2VZ)(U^2V^2 - Z^2)\,,\\
&\tau_0 = -2\epsilon_{21} + (U^2/Z)(VY - 1) - (Z/V^2)(VY + 1)\,.
\end{aligned} \tag{5.6}$$

In conclusion, the complicated set of conservation laws (5.1) has been shown to be equivalent to the matrix equation (5.4), *which thus provides a unified formulation of the octet of conservation laws*. It is sufficient, from a theoretical viewpoint, to deal with the ϵ_{ij} only, and forget about the more traditional Π, Π^*, E, etc.

B. A geometrical interpretation

The matrix equation (5.4) may be rewritten in the form

$$\begin{aligned}
&\partial_\alpha \epsilon_{ij} = X_i^2\, \partial_\alpha(U_j/X_i)\,,\\
&\partial_\beta \epsilon_{ij} = -X_i^2\, \partial_\beta(U_j/X_i)\,,
\end{aligned} \tag{5.7}$$

which is of the general type

$$\partial_\alpha A = +R\, \partial_\alpha B\,, \quad \partial_\beta A = -R\, \partial_\beta B\,. \tag{5.8}$$

The above equation constitutes the orthogonality condition of the systems of curves A = const and B = const, in a space with metric

$$ds^2 = dM^2 - \rho^2 c^2\, dt^2\,, \tag{5.9}$$

or with any other conformal metric. Obviously the metric is hyperbolic, with the null directions pointing along the characteristic curves. In terms of the M, t coordinates, Eq. (5.8) reads

$$\partial_t A = -\pi c R \partial_M B\,, \quad \partial_t B = -(\rho c/R)\partial_M A\,, \tag{5.10}$$

and, therefore, eliminating R, we obtain

$$\partial_t A\, \partial_t B - \rho^2 c^2\, \partial_M A\, \partial_M B = 0\,, \tag{5.11}$$

the announced orthogonality relation.

That notion of orthogonality turns out to be important owing to the actual form of the characteristic equations; thus, the fundamental characteristic equation (2.1) expresses orthogonality of particle trajectories (M = const) to spacelike sections (t = const), while the characteristic equation (2.5) expresses orthogonality of isobars (P = const) to isovelocity curves (v = const). More systematically, the SU(3)-symmetrical formulation (3.8), (3.9) indicates that the systems of curves U_k = const and X_i/X_j = const are mutually orthogonal (i,j,k being a substitution of 1,2,3) and, similarly, Eqs. (3.10) show that the curves X_k = const are orthogonal to the curves U_i/U_j = const.

Thus we arrive at *the geometrical interpretation of the eight independent conserved quantities* ϵ_{ij}, *as potentials whose equipotentials are orthogonal to the curves of equation*

$$U_j/X_i = \text{const}\,. \tag{5.12}$$

The latter may be viewed as "field lines," which serve to define the associated potentials ϵ_{ij}, up to a gauge transformation of the general form

$$\epsilon'_{ij} = f(\epsilon_{ij})\,.$$

In order to have a complete determination of ϵ_{ij}, account must be taken of the particular factor (X_i^2) occurring in the orthogonality relation (5.7).

C. The generalization to equations of state of arbitrary form

We will in fact consider for simplicity a power-law entropy distribution, but, again, that restriction is not essential as seen in Sec. III F.

The appropriate generalization of Eqs. (5.4), defining the conserved quantities ϵ_{ij}, turns out to be

$$\begin{aligned} \partial_\alpha \epsilon_{ij} &= +\Delta\,(\gamma' X_i\, \partial_\alpha U_j - \gamma U_j\, \partial_\alpha X_i)\,, \\ \partial_\beta \epsilon_{ij} &= -\Delta\,(\gamma' X_i\, \partial_\beta U_j - \gamma U_j\, \partial_\beta X_i)\,, \end{aligned} \tag{5.13}$$

where the factor Δ is defined as $\Delta \equiv (\Gamma/\gamma')Z^{(\gamma-3)} \equiv (\Gamma^{-1}/\gamma)V^{(\gamma'-3)}$. It must at once be pointed out that *not all of the* ϵ_{ij} *are integrable*, i.e., the Cauchy integrability condition ensuring compatibility of the pair of equations (5.13) is not automatically fulfilled. Thus *the SU(3) octet of conservation laws is in general incomplete*, a consequence of the symmetry breaking effect discussed in Sec. IIID. At the same time, it should be stressed that *the above SU(3) symmetrical formulation (5.13) is, nevertheless, completely general, in the sense that it represents the set of all surviving conservation laws which are actually known*. We show below that there are four conservation laws in the most general case, five if the gas is monatomic ($\gamma = 3$); and we discuss the particular conditions under which the remaining four (resp. three) conserved ϵ_{ij} may exist.

From a geometrical standpoint, it is interesting to note that equations (5.13) still can be written in the form of an orthogonality relation

$$\begin{aligned} \partial_\alpha \epsilon_{ij} &= +\Delta X_i U_j\, \partial_\alpha \log(U_j^{\gamma'}/X_i^{\gamma})\,, \\ \partial_\beta \epsilon_{ij} &= -\Delta X_i U_j\, \partial_\beta \log(U_j^{\gamma'}/X_i^{\gamma}) \end{aligned} \tag{5.14}$$

so that the equipotentials ϵ_{ij} = const are the curves orthogonal to the field lines of equation

$$U_j^{\gamma'}/X_i^{\gamma} = \text{const}. \tag{5.15}$$

We now proceed to establish the form of the Cauchy integrability conditions. Taking account of the general results derived in Sec. IIIC, we obtain, after some algebra,

$$\begin{aligned} \partial_\beta(\partial_\alpha \epsilon_{ij}) - \partial_\alpha(\partial_\beta \epsilon_{ij}) &= \Gamma^{-1}\Delta\,(\mathbf{X},\partial_\alpha \mathbf{X},\partial_\beta \mathbf{X}) \\ &\quad\times \left\{ \gamma'(\gamma-3)\left(\frac{\delta_{ij}}{3} - \frac{X_i}{X_j}\delta_{j3}\right) \right. \\ &\quad \left. - \gamma(\gamma'-3)\left(\frac{\delta_{ij}}{3} - \frac{U_j}{U_i}\delta_{i2}\right)\right\}, \end{aligned} \tag{5.16}$$

where δ_{ij} is the Kronecker symbol; therefore, the condition of integrability of ϵ_{ij} reads

$$\gamma'(\gamma-3)\left(\frac{\delta_{ij}}{3} - \frac{X_i}{X_j}\delta_{j3}\right) = \gamma(\gamma'-3)\left(\frac{\delta_{ij}}{3} - \frac{U_j}{U_i}\delta_{i2}\right). \tag{5.17}$$

That condition is automatically satisfied at least when the three Kronecker symbols become zero, which occurs in three cases: (i,j) = (1,2), (3,1), or (3,2). Thus the three conservation laws of ϵ_{12}, ϵ_{31}, and ϵ_{32} always exist, independently of the form of the equation of state; these are associated with the center-of-mass integral Π^*, energy E, and momentum Π, respectively, the correspondence being as follows:

$$\begin{aligned} \epsilon_{32} &= \Pi - Mv\,, \\ \epsilon_{12} &= -\Pi^* - Mv^*\,, \\ \epsilon_{31} &= -E + M\left(\tfrac{1}{2}v^2 - \gamma' c^2/\gamma(\gamma-1)\right). \end{aligned} \tag{5.18}$$

That generalizes the result [Eq. (5.6)] of Sec. VA.

Next, it is easily seen from Eq. (5.17) that ϵ_{13} exists if $\gamma = 3$, i.e, if the gas is monatomic, whereas ϵ_{21} exists when the index $\gamma' = 3$, that is to say, $b = 4\gamma/3$, according to Eq. (2.7). The remaining nondiagonal element, ϵ_{23}, can exist only when both γ and γ' equal 3, which is the equation of state discussed in Secs. IIIE and VA, possessing exact SU(3) symmetry. These three potentials express the laws of conservation of E^*, τ_0, and τ_1, respectively.

Turning now to the consideration of the diagonal elements, we similarly find that ϵ_{11} exists when $\gamma' = \gamma$, ϵ_{22} when $\gamma' = 2\gamma/(\gamma-1)$, and ϵ_{33} when $\gamma' = \gamma/(\gamma-2)$. But these are only particular cases of a generally valid conservation law, expressing *scale invariance*. Let us introduce two linear combinations λ and σ:

$$\lambda = \mathrm{Tr}\ \epsilon_{ij} \equiv \epsilon_{11} + \epsilon_{22} + \epsilon_{33}\,, \tag{5.19}$$

$$\sigma \equiv \epsilon_{11}/\gamma'' + \epsilon_{22}/\gamma + \epsilon_{33}/\gamma' \quad (1/\gamma'' + 1/\gamma + 1/\gamma' = 1)\,, \tag{5.20}$$

which are formally defined through the corresponding linear combinations of Eqs. (5.13), independently of any consideration of integrability of the ϵ_{ij}. In the same way as in Sec. VA it may then be seen that Tr ϵ_{ij} is indeed integrable, and in fact is an absolute constant, which again we take to be zero. The quantity σ is found to be always integrable as well, and constitutes our fourth generally valid conservation law. We observe that, in view of the way the two indices γ, γ' and the new index γ'' occur in Eq. (5.20), it would seem appropriate to rewrite γ_1, γ_2, and γ_3 in place of γ'', γ, and γ', from now on; then Eq. (5.20) would take the form

$$\sigma \equiv \sum_{i=1}^{3} \frac{\epsilon_{ii}}{\gamma_i}\,. \tag{5.21}$$

Since Tr $\epsilon_{ij} = 0$, equivalent expressions of the same conservation law are found in the general form

$$\sigma' = \sigma + \mu\ \mathrm{Tr}\ \epsilon_{ij}\,,$$

where μ is an arbitrary constant.

From Eq. (5.20) we easily recover the earlier particular results concerning the cases of integrability of ϵ_{11}, of ϵ_{22}, and of ϵ_{33}.

In conclusion, we have four generally valid conservation laws (five for a monatomic gas), and three others which may or may not exist, depending on the particular form of the equation of state. All together, *they form an SU(3) octet, and they all are expressed by the same universal formula* [*Eq. (5.13)*] *in matrix form, through which the formal symmetry of the octet is made manifest.* Thus the conservation laws of, e.g., energy and momentum constitute two aspects of a unique law of conservation of a quantity of more complex mathematical nature, a result which reminds us of the unified nature of energy and momentum in relativity.

D. The transformation formula for the conservation laws

We again consider the general case of arbitrary entropy distribution.

We have seen in Sec. IIIA that an SU(3) generator transforms the vectors $\mathbf{U}$ and $\mathbf{X}$ according to the formulae

$$\delta X = HX\,, \quad \delta U = KU\,, \tag{5.22}$$

where X, δX, U, δU denote three-dimensional column vectors, following standard matrix notation. The choice of a 3×3 traceless matrix H determines the infinitesimal transformation and, consequently, also determines the matrix K operating on the conjugate space $\{\mathbf{U}\}$. From the complete set of transformation formulae listed in Sec. IIC, it can be seen that the two matrices are in fact related by

$$K = -\tilde{H}\,, \tag{5.23}$$

where symbol $\sim$ denotes a transposition. For finite linear transformations, Eqs. (5.22) become

$$X' = \mathscr{M}X\,, \quad U' = \mathscr{N}U\,, \tag{5.24}$$

where $\mathscr{M}$ and $\mathscr{N}$ are arbitrary 3×3 matrices of unit determinant; and Eq. (5.23) implies between them the relation

$$\mathscr{N} = \tilde{\mathscr{M}}^{-1}\,. \tag{5.25}$$

Such a relation might have been expected without any calculation, in view of the fact that Eq. (3.13) reads, in matrix notation,

$$\tilde{U}X = 1\,, \tag{5.26}$$

which has to be satisfied by the transformed U', X' as well; and the simplest, probably unique way to ensure that result is obviously through Eq. (5.25).

Then, as a result of the SU(3) symmetry established in Secs. II and III, the transformed vectors $\mathbf{U}'$ and $\mathbf{X}'$ obey characteristic equations of unchanged form:

$$\partial_\alpha \mathbf{U}' = -\Gamma^{-1}\mathbf{X}' \wedge \partial_\alpha \mathbf{X}'\,, \qquad \partial_\beta \mathbf{U}' = +\Gamma^{-1}\mathbf{X}' \wedge \partial_\beta \mathbf{X}'\,, \tag{5.27}$$

$$\partial_\alpha \mathbf{X}' = +\Gamma\mathbf{U}' \wedge \partial_\alpha \mathbf{U}'\,, \qquad \partial_\beta \mathbf{X}' = -\Gamma\mathbf{U}' \wedge \partial_\beta \mathbf{U}'\,. \tag{5.28}$$

It must be noted that the above equations involve the original factor Γ rather than Γ', which in general differs from Γ; in this respect the transformed Eqs. (5.27) and (5.28) really differ from the original ones (3.8)–(3.10), as expected as a result of the symmetry breaking. But the following geometrical result (see Sec. VB) still holds:

The equations $X'_i/X'_j = \mathrm{const}$, and $U'_k = \mathrm{const}$ define mutually orthogonal systems of curves; and the equations $U'_i/U'_j = \mathrm{const}$, $X'_k = \mathrm{const}$ constitute orthogonal systems as well; here (i,j,k) stands for an arbitrary permutation of (1,2,3), and a prime denotes the operation of an arbitrary unimodular linear transformation. These two orthogonal systems of curves thus depend on eight continuous parameters. That result holds in the most general case, with arbitrary entropy distribution.

We next consider the effect of linear transformations on the conserved quantities ϵ_{ij}. Let us *define* the transformed ϵ'_{ij} by the formulae

$$\partial_\alpha \epsilon'_{ij} = +\Delta\,(\gamma' X'_i\,\partial_\alpha U'_j - \gamma U'_j\,\partial_\alpha X'_i)\,, \qquad \partial_\beta \epsilon'_{ij} = -\Delta\,(\gamma' X'_i\,\partial_\beta U'_j - \gamma U'_j\,\partial_\beta X'_i)\,, \tag{5.29}$$

where, in a similar manner as in Eqs. (5.27) and (5.28), the factor Δ is retained instead of its transformed Δ'. We certainly are free to choose the definitions of the ϵ and ϵ' as we please, in the most convenient way; it is only necessary to recognize that the definitions (5.13) of the ϵ_{ij} and (5.29) of the ϵ'_{ij} are not completely equivalent, owing to the symmetry breaking. The ϵ'_{ij} (provided their integrability condition is satisfied; see the discussion below) still have the character of conserved quantities just as well as the ϵ_{ij} do; but the two sets are, of course, not independent.

Starting from the definition (5.29) and Eqs. (5.22), one easily derives *the explicit transformation formula for infinitesimal transformations*:

$$\delta\epsilon = [H,\epsilon]\,, \tag{5.30}$$

where ϵ is the matrix whose matrix elements are the ϵ_{ij}, and the bracket denotes a commutator. It is interesting to note that Eq. (5.30) has the form of the quantum-mechanical evolution equation, in Heisenberg representation. In order that the transformed ϵ'_{ij} be integrable, the matrix H should be such that the corresponding matrix element of its commutator with the matrix ϵ contain only elements ϵ_{ij} belonging to the set of existing conservation laws.

Assume there exist n independent conservation laws and denote them with Greek indices, $\epsilon_{\alpha\beta}$. According to the Noether theorem, these conservation laws arise from n independent generators out of the SU(3) octet, forming a Lie subalgebra of order n (if the n generators did not constitute a subalgebra there would be more than n conservation laws, in contradiction with the hypothesis). For linear transformations H belonging to that subalgebra, the n transformed quantities $\epsilon'_{\alpha\beta}$ do exist; and they are expressible as n (independent) combinations of the original $\epsilon_{\alpha\beta}$, since the total number of independent conservation laws is only n. Thus we have a group of linear transformations, depending on n continuous parameters, transforming the n quantities $\epsilon_{\alpha\beta}$ into each other. Geometrically, there exist n families of curves $U'_{\beta}{}^{\gamma}/X'_{\alpha}{}^{\gamma} = \text{const}$ orthogonal to the equipotential curves $\epsilon'_{\alpha\beta} = \text{const}$, each family depending on n continuous parameters.

VI. CONCLUSION

Earlier works (Refs. 1–3) have shown the existence of special symmetries of the Euler equations of adiabatic gas flow; these symmetries are far from being immediately apparent when the equations are written in any of their usual forms, and they are not easily interpretable in terms of general physical principles either; it thus appeared necessary to construct a new formalism in which the hidden symmetries would become manifest.

In the one-dimensional case, which is the subject of the present work, our main result is that the gas dynamical equations are separable into an SU(3) invariant subsystem (S) and a subsidiary equation which generally breaks the symmetry; in that formulation the hidden symmetries are made manifest. One of its most striking features is *the emergence of two conjugate vector spaces* $\{\mathbf{X}\}$ *and* $\{\mathbf{U}\}$ *of dimension three, which may be constructed from the six-dimensional manifold of the physical variables* (M, t, P, v, ρ, r) in a manner determined by Eqs. (3.2) and (3.4). The consideration of these two vectors introduces a high degree of analytical order in what *a priori* looked like a completely noncovariant system. In terms of $\mathbf{X}$ and $\mathbf{U}$ the system assumes the compact vector form of Eqs. (3.8)–(3.10) and (3.13); the formal covariance is, however, but only to a minor extent, broken by the occurrence of a multiplicative factor Γ in the equations. The conservation laws are expressed by the Cauchy integrability conditions of the components ϵ_{ij} of a tensor ϵ, defined by Eqs. (5.13); there are at most eight independent conserved quantities, forming an SU(3) octet, since the tensor ϵ is traceless; but their number is generally less than eight (five for a monatomic gas; four otherwise) as a consequence of the symmetry breaking. A geometrical interpretation of the conservation laws in terms of systems of orthogonal curves is also presented (Secs. VB and VD).

A curious consequence of the present formulation is the invariance of monatomic gas flow under *Lorentz transformations* in a three-dimensional Minkowski space, as discussed in Appendix A; it would be very interesting to determine whether a connection can be established between the light cones of that Minkowski space and the characteristic curves.

One of the most intriguing points is the close analogy existing between the general form of the Euler equations and the more special form (3.23) discussed in Sec. III E, *which is characterized by an infinite number of conservation laws.* We in fact started the present study with that Eq. (3.23), in an earlier work (Ref. 3), and it turned up as a considerable surprise that so much of the formalism developed in Ref. 3 was of general validity, so that we may hope that further study of that equation would yield still more generalizable results.

It is of course tantalizing to attempt a generalization of the present formalism to spaces of higher dimensionality ($N = 2$ or 3). A preliminary study of the spherically symmetrical case indicates that part of the formalism is indeed generalizable, but the spherical assumption is almost certainly too restrictive to be very useful, as it will mask much of the underlying covariance that may be present.

The application of the above results to the theory of self-similar flow is discussed in Sec. IV.

ACKNOWLEDGMENTS

The author gratefully acknowledges Professor S. Kato's hospitality at the Department of Astronomy, Kyoto University, where the present work has been completed, and thanks him for his encouragement and interest. He also thanks a referee for pointing out useful references.

APPENDIX A: A LORENTZ-INVARIANT FORMULATION FOR MONATOMIC GAS FLOW

We have seen that a formal SU(3) symmetry is present, even for arbitrary entropy profiles, and that the symmetry generally breaks down owing to the lack of covariance of factor Γ. We now point out that, at least for monatomic gases, there still exists an exact SU(2) subgroup of symmetry, which may be constructed from the generator G_4 of the transformation (T^*), together with generators G_3, G_7 of time translation and scale transformation. One easily verifies that G_3, G_4, G_7 indeed constitute an SU(2) subalgebra of the SU(3) algebra that we constructed.

As is well known, SU(2) is isomorphic to the group SO(3) of three-dimensional rotations, and admits of irreducible representations of any integral dimension, from 1 to $+\infty$. The simplest one is the spinorial two-dimensional representation; we find that a possible choice for the spinor components is

$$\begin{pmatrix} v \\ v^* \end{pmatrix}. \tag{A1}$$

Another case of particular interest is that of vector representations of dimension three; we derive the following solution for the vector components:

$$\begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} vc^* + v^*c \\ v^*c^* - vc \\ v^*c^* + vc \end{pmatrix}, \tag{A2}$$

where $v^* \equiv vt - r$, $c^* \equiv ct$, as usual; we note that M is a *scalar*, i.e., an invariant of the group. Still another invariant is the quadratic form

$$A^2 + B^2 - C^2 = r^2c^2, \tag{A3}$$

so that the group in fact represents rotations in a vector space of complex coordinates (A, B, iC), that is to say, *Lorentz transformations* in the real space (A, B, C) where the metric has the hyperbolic signature $(+\ +\ -)$.

The above results make it possible to derive some interesting Lorentz-invariant forms of the partial differential equation of monatomic gas dynamics. Choosing the scalar M as unknown function z and the spinor components v, v^* as independent variables x, y, we obtain an equation of the Monge–Ampère general type:

$$s^2 - rt + a_1 r + a_2 s + a_3 t = a_0 \tag{A4}$$

(see Forsyth, Ref. 11), where, in standard notations,[11]

$$p = z'_x, \quad q = z'_y, \quad r = p'_x, \quad s = p'_y, \quad t = q'_y. \tag{A5}$$

The equation in fact assumes its simplest form when we choose $z \equiv 1/V(M)$ as unknown, and then reads

$$s^2 - rt = f(z), \tag{A6}$$

where the arbitrary function f is related to the entropy distribution σ by

$$f(z) = MV. \tag{A7}$$

[see Eqs. (3.28) and (3.31) for the general definition of V].

From the discussion of Sec. III E we conclude that Eq. (A6) has an exact SU(3) symmetry and an infinite number of conservation laws when $f(z)$ is proportional to z^{-4}. The isentropic case is characterized by $f=$ const, and is easily integrable. The case where $f(z)$ is proportional to $z^{-8/5}$ has been shown to be reducible to the isentropic case (Gaffet, Ref. 2).

APPENDIX B: ROGERS' CLASS OF INVARIANCE TRANSFORMATIONS

In their recent book, Rogers and Shadwick [Ref. 12, Sec. (3;9)] discuss a class of invariance transformations of the equations of unidimensional, nonsteady, adiabatic gas dynamics; these transformations were first introduced by Rogers (Ref. 13). The question naturally arises as to whether such transformations bear any relation with those discussed in the present paper; the answer is no, as we presently show.

Rogers' transformation may be analyzed in terms of a fundamental symmetry which we denote by (S_R), characterized by the essential properties that it conserves both the mass and position coordinate, up to a sign, and transforms the pressure into its inverse:

$$(S_R): r' = r, \quad M' = -M, \quad P' = 1/P.$$

The remaining variables then transform as follows:

$$t' = -\Pi, \quad v' = v/P, \quad \rho' = -\rho P/(P + \rho v^2),$$

in our system of notations. These transformation formulae are reciprocal, i.e., $(S_R)^2$ is the identity. Thus, the time coordinate and the momentum (Π) are exchanged by the symmetry.

It is interesting to note that our transformation $(T'\overline{T}T')$ shares with (S_R) the fundamental property of transforming the pressure into its inverse: $P' = 1/P$; we recall that the symmetry $(\overline{T})$, first introduced in (Ref. 2), is characterized by the properties $M' = 1/M$, $t' = -t$. Nevertheless, the $(T'\overline{T}T')$-transformation formulae are on the whole quite different from those of (S_R); they read

$$(T'\overline{T}T'): \quad P' = \frac{1}{P}, \quad r' = E - \frac{Mv^2}{2} + \frac{\gamma'(2-\gamma)}{\gamma(\gamma-1)} Mc^2,$$

$$t' = Mv - \Pi, \quad v' = -v, \quad M' = \frac{M}{P}, \quad \rho' = \frac{\gamma}{\gamma''} \frac{\rho}{P^2}.$$

There is in fact an essential difference between (S_R) and our own class of symmetries, which is that the former *does not leave the characteristic curves invariant*. The simplest way to see this is to start from the fundamental relation (2.1): $\partial_\alpha M = -\rho c\, \partial_\alpha t$, expressing orthogonality, in the sense defined in Sec. V E, of the systems of curves $M =$ const and $t =$ const. If the characteristic curves were invariant, the orthogonality condition would be preserved, and, since the curves $M' =$ const and $M =$ const coincide (as $M' = -M$), the curves $t' =$ const and $t =$ const would have to coincide as well. This is not the case, since $t' = \Pi$ is not a function of t only. Thus the characteristics are not invariant under (S_R). That property might, of course, have been derived directly, without appealing to the notion of orthogonality. (It can be shown, in fact, that a transformation that conserves r, M, *and* the characteristic coordinates α, β, would have to be the identity.)

[1]B. Gaffet, Preprint R.I.F.P. 442, May 1981.

[2]B. Gaffet, "Two hidden symmetries of the equations of ideal gas dynamics, and the general solution in cases of nonuniform entropy distribution," J. Fluid Mech. **134**, 179 (1983).

[3]B. Gaffet, "An infinite Lie group of symmetry of one-dimensional gas flow, for a class of entropy distributions," preprint, September 1982.

[4]These three transformations have been denoted (T^*), (T') and $(\overline{T})$, respectively.

[5]We will use the symbol $(\mathfrak{T}_r)$ rather than $(\mathfrak{T}_r^h)$ if the value of h is immaterial.

[6]C. Fronsdal, "Group theory and applications to particle physics," in *Brandeis Summer Institute, 1962, Vol. I* (Gordon and Breach, New York, 1962).

[7]M. E. Mayer, "Unitary symmetry of strong interactions," in *Brandeis Summer Institute, 1963, Vol. I* (Gordon and Breach, New York, 1963).

[8]The seventh parameter (the sound velocity c) does not enter the transformation formulae of the six M, t, P, v, ρ, r.

[9]A. R. Forsyth, *Theory of Differential Equations* (Dover, New York, 1959), Part IV, Vol. V, Chaps. III and IV.

[10]Dimensionless quantities depend upon a *single* variable, which we may take to be the ratio of characteristic coordinates β/α. Therefore, both ∂_α and ∂_β are proportional to the total differential d.

[11]A. R. Forsyth, *Theory of Differential Equations* (Dover, New York, 1959), Part IV, Vol. VI, Chaps. XVI and XVII.

[12]C. Rogers and W. F. Shadwick, *Bäcklund Transformations and Their Applications* (Academic, New York, 1982).

[13]C. Rogers, Z. Angew. Math. Phys. **19**, 58 (1968).

The continuous Heisenberg chain and constrained harmonic motion

Fernando Lund
Departamento de Física, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Casilla 5487, Santiago, Chile

(Received 1 August 1983; accepted for publication 16 September 1983)

It is shown that the equation describing the evolution of the classical continuous Heisenberg ferromagnet can be regarded as one aspect of quadratically constrained harmonic motion, as it is also the case in a number of other integrable systems. Two ingredients are used: a method to solve the inverse scattering problem for second-order operators using ordinary differential (rather than integral) equations and the equivalence of the Heisenberg chain with the nonlinear Schrödinger equation.

PACS numbers: 02.30.Jr, 75.10.Jm

INTRODUCTION

A number of nonlinear equations of mathematical physics have attracted wide attention in recent years.[1] This can be traced to the fact that their initial value problem can be solved, and the study of the solutions so obtained has revealed a number of surprising properties that were impossible to guess by doing perturbations on the linear approximation, as well as intriguing connections among rather disparate fields such as group representations, differential geometry, and scattering theory. Extension of these results to the quantum domain has permitted the construction of some nontrivial quantum field theories, unwrapping some more unexpected links with soluble models in statistical mechanics and isomonodromy deformations.[2]

The origin of the many relations among apparently very different—both mathematical and physical—systems is still far from being understood, and much effort is currently being spent trying to unify what seems unrelated. In this context, Deift, Lund, and Trubowitz[3–4] have shown that the Korteweg–de Vries, nonlinear Schrödinger, sine–Gordon, and Toda lattice equations on the line, as well as the Toda lattice on the circle, are different aspects of the same system: quadratic, complex, free oscillators constrained to an intersection of quadrics in phase space. This had been shown to be the case for the Korteweg–de Vries equation on the circle by Moser and Trubowitz.[5] In the present paper we shall extend these results to the Heisenberg ferromagnet equation. Also, we shall emphasize one aspect of the procedure that is of interest in itself: namely, along the way from the original equation to the constrained oscillators one finds a method to solve the inverse scattering problem for second-order operators via an ordinary, integrable, nonlinear differential equation. This was shown in the case of the Schrödinger operator by Deift and Trubowitz,[6] and it is in contrast to the previous approaches using linear integral equations.[7,8]

The continuous Heisenberg chain and its equivalence to the nonlinear Schrödinger equation are described in Sec. 1. Section 2 recalls the map from the space of solutions of the latter to a new space where inverse scattering for the second-order operator (2.1) can be regarded as the flow given by constraining an infinite number of quadratic oscillators to an intersection of two quadrics (Sec. 3) and the motion generated by a solution of the nonlinear Schrödinger equation corresponds to a related flow in the same subvariety of phase space (Sec. 4). These results are used in Sec. 5 to construct the announced extension to the Heisenberg chain. Concluding remarks are given in Sec. 6, and the appendix contains details relevant to Sec. 3, inverse scattering.

1. THE CONTINUOUS HEISENBERG CHAIN

The continuous Heisenberg spin chain is a one-dimensional system with dynamical variables $S_j(x,t), j = 1,2,3$, taking values on the unit sphere: $S_1^2 + S_2^2 + S_3^2 = 1$. The motion is governed by the Hamiltonian

$$H = \frac{1}{2}\int_{-\infty}^{\infty} \frac{\partial S^i}{\partial x}\frac{\partial S^i}{\partial x}\,dx\,, \tag{1.1}$$

with boundary conditions

$$\mathbf{S} \underset{|x|\to\infty}{\longrightarrow} (0,0,1) \tag{1.2}$$

and Poisson brackets

$$\{S_j(x,t), S_k(y,t)\} = \epsilon_{jkl} S_l(x,t)\delta(x-y)\,. \tag{1.3}$$

This gives the equation of motion

$$\frac{\partial \mathbf{S}}{\partial t} = \mathbf{S}\times\frac{\partial^2 \mathbf{S}}{\partial x^2}\,. \tag{1.4}$$

It is easy to check that this equation preserves the constraint $\mathbf{S}^2 = 1$. The initial value problem for this system was shown to be solvable using the inverse scattering method by Takhtajan.[9] Later, Zakharov and Takhtajan[10] proved that, in a sense, this system was equivalent to the Schrödinger equation with a cubic nonlinearity:

$$i\frac{\partial\psi}{\partial t} + \frac{\partial^2\psi}{\partial x^2} + 2|\psi|^2\psi = 0\,. \tag{1.5}$$

The sense of this equivalence is as follows: Both Eqs. (1.4) and (1.5) implement the vanishing of the curvature of certain connections in a principal fiber bundle with structure group SL(2,R),[11] and there is a gauge transformation that transforms one connection into the other. More explicitly, the nonlinear Schrödinger equation (1.5) is the compatibility condition for the linear system

$$\frac{\partial\Phi}{\partial x} = U_1\Phi\,, \quad \frac{\partial\Phi}{\partial t} = V_1\Phi\,, \tag{1.6}$$

with ($\psi_x \equiv \partial\psi/\partial x$)

$$U_1 = \begin{pmatrix} -ik & \psi \\ -\psi^* & ik \end{pmatrix},$$

$$V_1 = \begin{pmatrix} -2ik^2 + i|\psi|^2 & 2k\psi + i\psi_x \\ -2k\psi^* + i\psi_x^* & 2ik^2 - i|\psi|^2 \end{pmatrix}, \tag{1.7}$$

and the equation of the Heisenberg chain written in the form

$$\frac{\partial S}{\partial t} = -\frac{i}{2}\left[S, \frac{\partial^2 S}{\partial x^2}\right], \tag{1.8}$$

where $S \equiv S_i \sigma^i$ and σ^i are the Pauli spin matrices, is the compatibility condition for

$$\frac{\partial \Phi}{\partial x} = U_2 \Phi, \quad \frac{\partial \Phi}{\partial t} = V_2 \Phi, \tag{1.9}$$

with

$$U_2 = -ikS,$$

$$V_2 = kSS_x - 2ik^2 S.$$

The point is[10] that if $\psi(x,t)$ is a solution of (1.5) with $\psi \to 0$ as $|x| \to \infty$, then the solution of

$$\frac{\partial g}{\partial x} = \begin{pmatrix} 0 & \psi \\ -\psi^* & 0 \end{pmatrix} g,$$

$$\frac{\partial g}{\partial t} = -i \begin{pmatrix} |\psi|^2 & \psi_x^* \\ \psi_x & -|\psi|^2 \end{pmatrix} g, \tag{1.10}$$

with $g \to 1$ as $x \to +\infty$ is such that $S = g^{-1}\sigma^3 g$ is a solution of (1.8) with $S \to \sigma_3$ as $x \to +\infty$. Moreover, recall that

$$\frac{\partial \Phi}{\partial x} = \begin{pmatrix} -ik & \psi \\ -\psi^* & ik \end{pmatrix} \Phi$$

can be interpreted as a scattering problem when $\psi \to 0$ as $|x| \to \infty$. If the reflection coefficient corresponding to the given ψ vanishes at zero, i.e., $R(0) = 0$, then one has $S \to \sigma_3$ as $x \to -\infty$ also.

2. CONSTRAINED OSCILLATORS AND THE NONLINEAR SCHRÖDINGER EQUATION

It was shown by Deift, Lund, and Trubowitz[3,4] that the study of Eq. (1.5) could be reduced to the study of a system of harmonic oscillators with quadratic constraints. Below we shall show that this is also true for the Heisenberg chain, using the results of Zakharov and Takhtajan that we have just sketched. First, we recall some features of this equivalence of the nonlinear Schrödinger equation with a system of constrained oscillators.

The equation

$$\frac{\partial \Phi}{\partial x} = \begin{pmatrix} -ik & q(x) \\ r(x) & ik \end{pmatrix} \Phi \tag{2.1}$$

with $q, r \to 0$ as $|x| \to \infty$ has unique solutions $f(x,k)$ and $h(x,k)$ with asymptotics

$$f \underset{x \to +\infty}{\longrightarrow} \begin{pmatrix} e^{-ikx} & 0 \\ 0 & e^{ikx} \end{pmatrix},$$

$$h \underset{x \to -\infty}{\longrightarrow} \begin{pmatrix} 0 & e^{-ikx} \\ -e^{ikx} & 0 \end{pmatrix},$$

and they are related by $h = fT$ where T, the transition matrix

$$T = \begin{pmatrix} a_+(k) & b_-(k) \\ b_+(k) & -a_-(k) \end{pmatrix}, \tag{2.2}$$

is such that $a_+ a_- + b_+ b_- = 1$. The reflection coefficients $R_\pm(k)$ are defined by

$$R_\pm(k) = b_\pm(k) a_\pm^{-1}(k).$$

Notice that (2.1) seems to be more general than what we need, as we are interested in the case $-r^* = q = \psi$. However, this extra freedom is needed to get the map to the constrained oscillators, which is constructed as follows: given $q(x)$, $r(x)$, the solutions f and g as well as the reflection coefficients are uniquely determined. Define then new functions $X_\pm(k)$, $Y_\pm(k)$ through[4]

$$X_+(k) = (R_+(k)/\pi k)^{1/2} f_{12}(0,k),$$

$$Y_+(k) = (R_+(k)/\pi k)^{1/2} f_{22}(0,k),$$

$$X_-(k) = (R_-(k)/\pi k)^{1/2} f_{11}(0,k),$$

$$Y_-(k) = (R_-(k)/\pi k)^{1/2} f_{21}(0,k), \tag{2.3a}$$

for real $k \neq 0$, and

$$X_\pm(0) = (1/\sqrt{\pm 2i})(f_{12}(0,0) \pm f_{11}(0,0)),$$

$$Y_\pm(0) = (1/\sqrt{\pm 2i})(f_{22}(0,0) \pm f_{21}(0,0)). \tag{2.3b}$$

Here we are assuming $R_+(0) = R_-(0) = 0$, which is enough for our purpose, the study of the Heisenberg chain when $S \to \sigma_3$ as $x \to \pm\infty$. The more general case $R(0) \neq 0$ appears in the nonlinear Schrödinger case.[4] It is straightforward to consider it, although it is more cumbersome. Also, define[12]

$$X_{\pm j} = \left(\frac{2ic_{j\pm}}{k_{j\pm}}\right)^{1/2} f_{\substack{12 \\ 11}}(0, k_{j\pm}), \quad j = 1,\dots,n,$$

$$Y_{\pm j} = \left(\frac{2ic_{j\pm}}{k_{j\pm}}\right)^{1/2} f_{\substack{22 \\ 21}}(0, k_{j\pm}). \tag{2.3c}$$

Here, k_{j+} (resp. k_{j-}) are the eigenvalues of (2.1) given by the zeroes of a_+ (resp. a_-) in the upper (resp. lower) complex k plane. In general the number of eigenvalues whose imaginary part is positive need not be equal to the number of eigenvalues with negative imaginary part. For the case at hand, however, it is enough to consider this case. The c's are given by

$$c_{j\pm} = \tilde{c}_{j\pm} / a'_{j\pm},$$

where

$$a'_{j\pm} = \left.\frac{da}{dk}\right|_{k = k_{j\pm}},$$

and $\tilde{c}$ is given by the asymptotic behavior of the eigenfunctions:

$$\sim \begin{pmatrix} 0 & e^{-ik_{j+}x} \\ -e^{ik_{j-}x} & 0 \end{pmatrix} x \to -\infty,$$

$$\sim \begin{pmatrix} \tilde{c}_{j-} e^{-ik_{j-}x} & 0 \\ 0 & \tilde{c}_{j+} e^{ik_{j+}x} \end{pmatrix} x \to +\infty.$$

So, given two functions $q(x)$, $r(x)$ one inserts them in (2.1) and extracts all the information that can be given by the resulting scattering problem—Jost solutions, reflection coefficients, eigenvalues and eigenfunctions—and plugs it into (2.3). This gives a map

$$(q,r)\mapsto(X,Y).$$

The question we want to address is: How do X and Y change when $q=-r^*$ and q evolves according to the nonlinear Schrödinger equation (1.5)?

3. TRANSLATION FLOW AND INVERSE SCATTERING

Consider first the simpler evolution

$$q(\cdot),r(\cdot)\mapsto q(\cdot+t),r(\cdot+t). \tag{3.1}$$

That is, the potentials simply move by a uniform translation to the left with unit speed. What is the motion of the X's and Y's induced by the map (3.1)? The answer is[4]: X and Y evolve according to the differential equation

$$\begin{aligned}
\dot X_\pm(k) &= -ikX_\pm(k)-\Big(\sum lX_l^2\Big)Y_\pm(k),\\
\dot Y_\pm(k) &= ikY_\pm(k)-\Big(\sum lY_l^2\Big)X_\pm(k),\\
\dot X_{\pm j} &= -ik_{\pm j}X_{\pm j}-\Big(\sum lX_l^2\Big)Y_{\pm j},\\
\dot Y_{\pm j} &= ik_{\pm j}Y_{\pm j}+\Big(\sum lX_l^2\Big)X_{\pm j},
\end{aligned} \tag{3.2}$$

for $k\in R$ and $j=1,\dots,n$. We have written

$$\sum lX^2(l)\equiv\sum_{j=1}^{n}(k_{+j}X^2_{+j}+k_{-j}X^2_{-j}) + \int_{-\infty}^{+\infty} l(X^2_+(l)+X^2_-(l))\,dl$$

and similarly for ΣlY_l^2. Equations (3.2) are a system of ordinary, nonlinear differential equations. Notice that $X_\pm(0)$ and $Y_\pm(0)$ decouple from the other variables, that is, the motion of $X_\pm(k)$, $Y_\pm(k)$ $(k\neq 0)$, $X_{\pm j}$, $Y_{\pm j}$ is not affected by $X(0)$, $Y(0)$, and once the X,Y for $k\neq 0$ are known, substitution in (3.2) gives a linear equation for $X(0)$, $Y(0)$ with time-dependent coefficients.

The first important remark[4] is that (3.2) are the equations corresponding to the Hamiltonian

$$H_0=-\sum ikX_kY_k, \tag{3.3a}$$

with Poisson brackets

$$\begin{aligned}
&\{X_{ak},Y_{bl}\}=\delta_{ab}\delta(k-l),\quad a,b=+,-,\\
&\{X,X\}=\{Y,Y\}=0,
\end{aligned} \tag{3.3b}$$

and constraints

$$\begin{aligned}
\phi_1&=\sum X_k^2=0,\\
\phi_2&=\sum Y_k^2=0,
\end{aligned} \tag{3.3c}$$

with variables X,Y satisfying the initial condition (automatically preserved in time)

$$\sum X_kY_k=-i. \tag{3.4}$$

The dynamical system given by (3.3) is what we call a system of (complex) harmonic oscillators with quadratic constraints.

We have a map $(q,r)\mapsto(X,Y)$. Is it possible to go backwards? The answer is yes, due to the identities[4,12]

$$\begin{aligned}
q(0)&=-\int(X^2_+(k)+X^2_-(k))k\,dk+\sum_j j(X^2_{+j}+X^2_{-j}),\\
r(0)&=\int(Y^2_+(k)+Y^2_-(k))k\,dk+\sum_j j(Y^2_{+j}+Y^2_{-j}).
\end{aligned} \tag{3.5}$$

So, as we consider different potentials in the one-parameter family (labeled by t) given by (3.1), X,Y will change according to (3.2), and (3.5) will give the value of each potential at the origin in terms of the value of X,Y at "time" t. But the value of the potentials $q(\cdot+t)$, $r(\cdot+t)$ at the origin is nothing but the value of $q(\cdot)$, $r(\cdot)$ at a distance t from the origin! So, given a solution of (3.2) we can reconstruct the potentials q,r this way. Indeed, the (singular) initial value problem given by (3.2) with initial conditions, as $t\to+\infty$,

$$\begin{aligned}
&X_+(k)\to 0,\quad Y_+(k)\to(R_+(k)/\pi k)^{1/2},\\
&X_-(k)\to(R_-(k)/\pi k)^{1/2},\quad Y_-(k)\to 0,\\
&X_\pm(0)\to\pm 1/\sqrt{\pm 2i},\quad Y_\pm(0)\to 1/\sqrt{\pm 2i},\\
&X_{+j}\to 0,\quad Y_{+j}\to(2ic_{j+}/k_{j+})^{1/2}\tilde c_{j+},\\
&X_{-j}\to(2ic_{j-}/k_{j-})^{1/2}\tilde c_{j-},\quad Y_{-j}\to 0,
\end{aligned} \tag{3.6}$$

solves the inverse scattering problem for q,r. That is, given[13] $R_\pm$, $k_{j\pm}$, $c_{j\pm}$ one has initial data for the evolution equation that will give $X(t)$, $Y(t)$ and, with it, $q(t)$, $r(t)$ through (3.5). One still has to make sure that this initial value problem has global solutions. But this is assured because (3.3) is a completely integrable system! This is seen by explicit construction of a complete set of integrals of motion in involution[12]

$$I_{ak}=\sum_{b,l}(k-l)^{-1}(X_{ak}Y_{bl}-X_{bl}Y_{ak})^2,\quad a,b=+,-. \tag{3.7}$$

In this approach then, the inverse scattering problem for (2.1) has been solved through a set of coupled, nonlinear, ordinary, integrable differential equations. This is in contrast to previous methods of solution[7,8] using a linear integral equation. Notice that (3.4) is automatically satisfied by the data (3.6). The case $r=\mp q^*$, of interest for the nonlinear Schrödinger equation $iq_t+q_{xx}\pm 2|q|^2q=0$ is obtained by imposing

$$X_-=(i)Y^*_+,\quad Y_-=-(i)X^*_+, \tag{3.8}$$

on the initial data. It is immediate to check that this condition is preserved by the flow (3.2). In terms of scattering data this means having

$$\begin{aligned}
&R_-(k)=\pm R^*_+(k),\quad k_{j-}=k^*_{j+},\\
&c_{j-}=\pm c^*_{j+},\quad \tilde c_{j-}=\pm\tilde c^*_{j+},
\end{aligned} \tag{3.9}$$

where the star denotes complex conjugation.

We have seen then that when q,r change according to the translation flow XY evolve as a system of harmonic oscillators constrained to the intersection of the two quadratics

(3.3c). It is simple to check[4] that, in terms of the action variables (3.7), this motion is generated by the Hamiltonian

$$H_1 \equiv -\left(\sum kI_k\right)^{-1/2}\left(\sum k^2 I_k\right),$$

with constraints (3.3c).

4. NONLINEAR SCHRÖDINGER FLOW

The original question we had in mind was to find the motion X,Y when $q = -r^*$ evolves according to the nonlinear Schrödinger equation. The answer can now be given quite simply[4,12]; consider

$$H_2 = \sum k^2 I_k . \tag{4.1}$$

The flow generated by constraining this second Hamiltonian to the *same* subvariety (3.3c) obviously commutes with the one generated by H_1 and is also integrable. The point is that when $q(= -r^*)$ obeys the nonlinear Schrödinger equation, XY change according to this second flow,[4] with initial data (preserved by H_2) satisfying (3.4) and (3.8). It is in this sense that the study of the nonlinear Schrödinger equation is reduced to the study of an integrable system of constrained harmonic oscillators.

5. HEISENBERG FLOWS

We are now in a position to go back to the Heisenberg chain. We have already mentioned that given a solution $\psi(x,t)$ of the nonlinear Schrödinger equation, one finds a solution $S(x,t)$ of the continuous Heisenberg chain $S(x,t) = g^{-1}\sigma^3 g$ by solving the system (1.10). The converse is also true[10]: take $S(x,t)$ a solution of (1.8) with boundary conditions $S \to \sigma_3$ as $|x| \to \infty$, with $S^2 = I$, $S = S^+$, $\operatorname{tr} S = 0$; it can be diagonalized by a unitary $g(x,t)$, $S = g^{-1}\sigma_3 g$. Since g is unitary, $(\partial g/\partial x)g^{-1}$ is anti-Hermitian. Moreover, it is always possible to choose g such that $(\partial g/\partial x)g^{-1}$ is off-diagonal:

$$\frac{\partial g}{\partial x} g^{-1} = \begin{pmatrix} 0 & \psi \\ -\psi^* & 0 \end{pmatrix}, \tag{5.1}$$

with a unique $\psi(x,t)$ that automatically goes to zero when $S \to \sigma_3$ with g going to a constant unitary diagonal matrix. Then, it can be shown[10] that ψ is a solution of (1.5) and that

$$\frac{\partial g}{\partial t} g^{-1} = -i\begin{pmatrix} |\psi|^2 & \psi_x^* \\ \psi_x & -|\psi|^2 \end{pmatrix}. \tag{5.2}$$

From this we see that, with the boundary condition $g \to I$ as $x \to +\infty$, g is nothing but the solution f of (2.1) evaluated at $k = 0$:

$$g(x,t) = f(x,0,t) .$$

In other words,

$$g(x,0) = \frac{i}{2}\begin{pmatrix} X_+(0) - iY_+^*(0) & X_+(0) + iY_+^*(0) \\ Y_+(0) + iX_+^*(0) & Y_+(0) - iX_+^*(0) \end{pmatrix}. \tag{5.3}$$

So, if we consider the translation flow for S: $S(\cdot) \to S(\cdot + t)$, we have induced a flow $\psi(\cdot) \to \psi(\cdot + t)$ through (5.1), which in turn induces a flow for X,Y given by (3.2) which is generated by the Hamiltonian H_0 (3.3a) with constraints (3.3c). Now to go back, that is to recover $S(\cdot + t)$ from a solution to this set (i.e., the oscillator set) of differential equations, one simply uses (5.3) and $S = g^{-1}\sigma_3 g$. That is, of the infinitely many oscillators that are evolving, only one $(X(0), Y(0))$ is needed to know how S is changing. Of course, the evolution of $X(0), Y(0)$ is not autonomous but coupled to that of all the rest. This is to be contrasted to the situation in the nonlinear Schrödinger case (3.5) where a combination of all the oscillators gives the behavior of q,r. Exactly the same situation is true for the flow $S(\cdot) \to S(\cdot,t)$ given by (1.8), the Heisenberg equation, which corresponds to $\psi(\cdot) \to \psi(\cdot,t)$, the nonlinear Schrödinger flow, which in turn corresponds to the flow generated by H_2 (4.1). Given a solution to this new differential equation, $S(\cdot) \to S(\cdot,t)$ is recovered through (5.3). Indeed, it is possible to check directly that (5.2) is the differential equation obeyed by $X(0)$, $Y(0)$ with Hamiltonian H_2 and constraints (3.3c). This is achieved by deriving formulas analogous to (3.5) for ψ_x and $|\psi|^2$.[12]

6. CONCLUDING REMARKS

We have constructed a map from the space of solutions of the continuous Heisenberg chain equation to a space of solutions of an equation that is obtained as a flow in an integrable system of free quadratic oscillators with quadratic constraints, using the nonlinear Schrödinger equation as a crutch. As mentioned in the introduction, a similar map can be constructed for a number of partial differential equations, thereby unifying them in this sense.

In the case of linear equations the Fourier transform can also be regarded as a map to a space of quadratic uncoupled oscillators. The nonlinear equations we have considered are then characterized by the fact that the nonlinearities are introduced into a linear system by a finite number of quadratic constraints. This is similar to what happens in the nonlinear sigma models,[14] in which case the number of constraints is infinite: one per space point.

In conclusion, we may say that KdV, nonlinear Schrödinger, sine–Gordon, Toda lattice, *and* Heisenberg chain are but different aspects of the same system, harmonic motion quadratically constrained to a subvariety of phase space.

ACKNOWLEDGMENT

Enlightening discussions with P. Deift and E. Trubowitz are gratefully acknowledged.

APPENDIX

It was mentioned in the text that the inverse scattering problem for the operator

$$L = \begin{pmatrix} -\dfrac{\partial}{\partial x} & q \\ -r & \dfrac{\partial}{\partial x} \end{pmatrix}, \tag{A1}$$

where q,r are complex-valued functions of a real variable, can be reduced to the problem of solving a nonlinear, integrable, ordinary differential equation. This was done for the Schrödinger operator on the line by Deift and Trubowitz,[6] and for the operator L by Deift, Lund, and Trubowitz.[4] Here we give some details that were omitted in the last reference.

The spectral theory for (A1) was given by Zakharov and

Shabat[7] and by Ablowitz *et al.*[8] The eigenvalue problem

$$L\Phi = ik\Phi \tag{A2}$$

is equivalent to Eq. (2.1). The spectrum of L consists of a continuous part, the real line, and a discrete part given by the zeroes of a_+ and a_- [cf. (2.1)] in the upper and lower halves of the complex k plane, respectively. If q,r decay faster than any power of z at infinity, the number of zeroes of a_+ and a_- is finite. Their multiplicity is not restricted, and, to the best of our knowledge, a necessary and sufficient condition on the potentials q,r ensuring that the zeroes are simple has not been found. All treatments *assume* a_+,a_- to have only simple zeroes, and we shall do so here as well. Also, in principle, they could lie on the real axis in which case they would not correspond to bound states but to resonances.

This would not add any difficulties but would make the formulas more cumbersome, so we shall assume there are no real zeroes.

Next, take $r = -q^*$. In this case, using (3.9), the following is true[12]:

$$q(x) = 2i \sum_{j=1}^{N} c_j f_{12}^2(x,k_j) - 2i \sum_{j=1}^{N} c^* f_{22}^{*2}(x,k) - \frac{1}{\pi}\int_{-\infty}^{\infty} dk\,(R(k) f_{12}^2(x,k) + R^*(k) f_{22}^{*2}(x,k)), \tag{A3}$$

where we have dropped the " + " subscript. Then, substitution of (A3) into

$$\frac{\partial}{\partial x}\begin{pmatrix} f_{12}(x,k) \\ f_{22}(x,k)\end{pmatrix} = \begin{pmatrix} -ik & q(x) \\ -q^*(x) & ik \end{pmatrix}\begin{pmatrix} f_{12}(x,k) \\ f_{22}(x,k)\end{pmatrix}, \tag{A4}$$

$$k \in R, \quad k = k_1,\ldots,k_N,$$

gives a nonlinear ordinary differential equation for f_{12}, f_{22}. Given $R(k)$, k_j, and c_j together with the boundary condition

$$\begin{pmatrix} f_{12} \\ f_{22}\end{pmatrix} \underset{x\to+\infty}{\longrightarrow} \begin{pmatrix} 0 \\ e^{ikx}\end{pmatrix},$$

the corresponding initial value problem can be solved, and substituting the solution back in (A3) gives $q(x)$, solving the inverse scattering problem. Equation (A4) has global solutions because with the definitions (2.3) it turns into (3.2) which, as we saw in the text, is an integrable system.

[1] For reviews, see A. C. Scott, F. Y. F. Chu, and D. W. McLaughlin, Proc. IEEE **61**, 1449 (1973); F. Lund, "Solitary Waves and Solitons," in *Proceedings of the First Chilean Symposium of Theoretical Physics*, edited by A. Seguel *et al.* (Universidad Técnica, Santiago, 1978); S. V. Manakov and V. E. Zakharov, eds., *Proceedings of the Soviet-American Symposium on Soliton Theory* (North-Holland, Amsterdam, 1981).

[2] L. D. Faddeev, "Quantum Completely Integrable Models of Field Theory," Leningrand preprint, 1979; H. B. Thacker, Rev. Mod. Phys. **53**, 253 (1981); M. Jimbo, T. Miwa, Y. Mori, and M. Sato, Physica D **1**, 80 (1980).

[3] P. Deift, F. Lund, and E. Trubowitz, Proc. Natl. Acad. Sci. USA **77**, 716 (1980).

[4] P. Deift, F. Lund, and E. Trubowitz, Commun. Math. Phys. **74**, 141 (1980).

[5] J. Moser, "Various Aspects of Integrable Hamiltonian Systems," preprint, Courant Institute, 1979.

[6] P. Deift and E. Trubowitz, Commun. Pure Appl. Math. **32**, 121 (1979).

[7] V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].

[8] M. Ablowitz, D. Kaup, A. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).

[9] L. A. Takhtajan, Phys. Lett. A **64**, 235 (1977).

[10] V. E. Zakharov and L. A. Takhtajan, Teor. Mat. Fiz. **38**, 26 (1979).

[11] M. Crampin, F. A. E. Pirani, and D. C. Robinson, Lett. Math. Phys. **2**, 15 (1977).

[12] F. Lund, Physica D **3**, 350 (1981).

[13] The residues of $a_\pm^{-1}$ at its poles are also needed to get c_j from $\tilde{c}_j$. They can be obtained because $a_\pm$ are given in terms of R_+, k_{j+} via a dispersion relation. See Ref. 8.

[14] K. Pohlmeyer, Commun. Math. Phys. **46**, 207 (1976).

An exact recursion for the composite nearest-neighbor degeneracy for a $2 \times N$ lattice space

R. B. McQuistan
Department of Physics and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

J. L. Hock
Department of Electrical Engineering and Computer Science, Marquette University, Milwaukee, Wisconsin 53233

(Received 25 January 1983; accepted for publication 9 September 1983)

A set theoretic argument is utilized to develop a recursion relation that yields exactly the composite nearest-neighbor degeneracy for simple, indistinguishable particles distributed on a $2 \times N$ lattice space. The associated generating functions, as well as the expectation of the resulting statistics are also treated.

PACS numbers: 02.50. + s, 05.50. + q

I. INTRODUCTION

The present paper considers the problem of determining the composite nearest-neighbor degeneracy for a $2 \times N$ rectangular lattice space. Specifically, we develop a recursion relation that yields exactly the multiplicity of those arrangements of simple, indistinguishable particles on a $2 \times N$ rectangular lattice that create prescribed numbers of the three possible types of nearest-neighbor pairs [occupied pairs (particle-particle), mixed pairs (vacancy-particle), and vacant pairs (vacancy-vacancy)]. We represent these numbers by n_{11}, n_{01}, and n_{00}, respectively.

A number of previous papers have dealt with the nearest-neighbor degeneracy for one-dimensional, $1 \times N$ lattice spaces when the number of one type of nearest-neighbor pair is specified,[1-3] as well as the composite degeneracy when the numbers of *all* three types of nearest-neighbor pairs are prescribed.[4] Kedem[5] has also discussed one-dimensional nearest-neighbor degeneracies and their relationship to sufficient statistics associated with binary stationary mth-order Markov chains. He applies his results to a determination of the asymptotic distribution of rare events. Yan[6] has treated several interesting one-dimensional nearest-neighbor degeneracy problems involving particles that occupy integral numbers of linearly contiguous lattice sites.

We first note that the quantities n_{11}, n_{01}, and n_{00} are not independent when N is specified because their sum is the total number of nearest neighbors, regardless of the number of particles present, i.e.,

$$3N - 2 = n_{11} + n_{01} + n_{00}\,. \tag{1}$$

[Here, as elsewhere, we do not distinguish between 0–1 and 1–0 nearest neighbor pairs.]

If, from each occupied lattice site, we draw three lines connecting the occupied site with its nearest-neighbor sites, there will be a total of $3q$ lines where q is the number of particles (see Fig. 1). Accounting for these lines yields

$$3q = 2n_{11} + n_{01} + \delta_1\,, \tag{2}$$

where δ_1 is the number of lines that go off the ends of the lattice space when one or more of the four end sites are occupied. A similar construction for the vacant sites results in

$$3[2N - q] = 2n_{00} + n_{01} + \delta_0\,. \tag{3}$$

Adding Eqs. (2) and (3), while considering Eq. (1) yields the expected result that

$$\delta_1 + \delta_0 = 4\,. \tag{4}$$

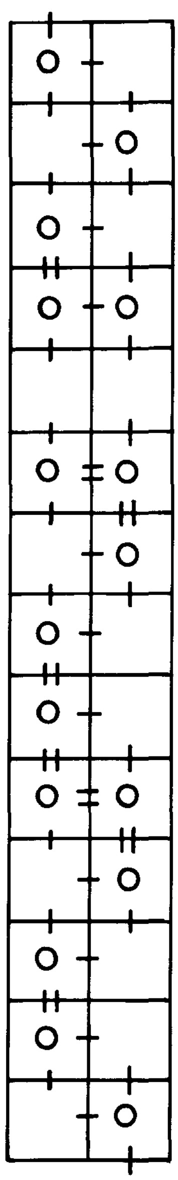

FIG. 1. On this 2×14 space there are 16 particles, so there are 48 lines. From each particle are drawn three lines, each line to a nearest-neighbor site. There are eight occupied nearest-neighbor pairs, with each of which is associated two lines, and 30 mixed nearest-neighbor pairs; $\delta_1 = 2$.

0022-2488/84/020261-05$02.50

When calculating the nearest-neighbor degeneracies, one must take account of the end-to-end as well as the top-to-bottom reflections of the lattice space. This additional degeneracy factor depends on the state of occupation of the four end sites, i.e., it depends on the δ's. As previously discussed,[4] the degeneracy factor for a one-dimensional lattice space [which is either 1 or 2 (for an end-to-end reflection)] is completely determined by whether n_{01} is even or odd. As can be seen from Eqs. (2) and (3), whether n_{01} is even or not does *not* specify the state of occupation of the four end compartments and hence does not dictate the value of the degeneracy factor for a $2\times N$ lattice (which may be 1, 2, or 4).

Note that if N, q, n_{11}, and n_{00} are specified, n_{01} is fixed and the values of the individual δ's are determined uniquely.

We designate $A\,[N,q,n_{11},n_{00}]$ to be the total number of all possible arrangements when q particles are distributed on a $2\times N$ lattice in such a way as to create n_{11} occupied nearest-neighbor pairs *and* n_{00} vacant nearest-neighbor pairs.

II. RECURSION RELATION FOR $A[N,q,n_{11},n_{00}]$

To establish a recursion for $A\,[N,q,n_{11},n_{00}]$, we must first differentiate between an $\alpha(N)$-space (which consists of two aligned rows of N equivalent rectangular sites) and a $\beta(N)$-space [an $\alpha(N)$-space to which an additional lattice site is affixed at the lower left-hand corner] (see Fig. 2).

Next we define $a[N,q,n_{11},n_{00}]$ to be the set of all arrangements of q particles on an $\alpha(N)$-space that contain n_{11} occupied *and* n_{00} vacant, nearest-neighbor pairs. Thus $\#a[N,q,n_{11},n_{00}]$, the number of elements of $a[N,q,n_{11},n_{00}]$ is $A\,[N,q,n_{11},n_{00}]$.

Let $a_j\,[N,q,n_{11},n_{00}]$ $(j=1,...,7)$ be subsets of $a[N,q,n_{11},n_{00}]$. Each of the a_j's is characterized by the state of occupation of the four end sites (see Fig. 3). Every arrangement in a_j differs from every arrangement in a_k $(j\neq k)$, i.e.,

$$a_j\,[N,q,n_{11},n_{00}]\cap a_k\,[N,q,n_{11},n_{00}]=\varnothing\,, \tag{5}$$

a null set.

In addition every element of $a[N,q,n_{11},n_{00}]$ will be found in one of the $a_j\,[N,q,n_{11},n_{00}]$, i.e.,

$$a[N,q,n_{11},n_{00}]=\bigcup_{j=1}^{7}a_j\,[N,q,n_{11},n_{00}]\,. \tag{6}$$

We conclude then that $\#a[N,q,n_{11},n_{00}]$, the number of elements of the set $a[N,q,n_{11},n_{00}]$ is given by

$$\#a[N,q,n_{11},n_{00}]=\sum_{j=1}^{7}\#a_j\,[N,q,n_{11},n_{00}]$$

or

$$A\,[N,q,n_{11},n_{00}]=\sum_{j=1}^{7}A_j\,[N,q,n_{11},n_{00}]\,, \tag{7}$$

where, four example, $\#a_2[N,q,n_{11},n_{00}]\equiv A_2[N,q,n_{11},n_{00}]$ is

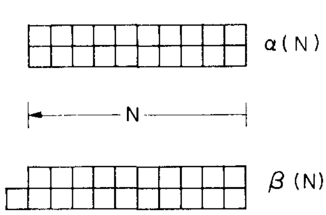

FIG. 2. (a) Definition of an $\alpha(N)$-space. (b) Definition of a $\beta(N)$-space.

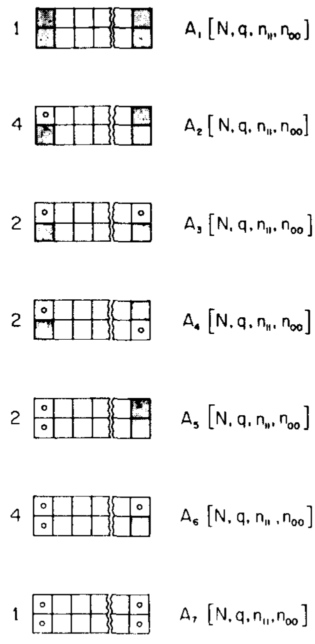

FIG. 3. Seven α-spaces are characterized by the state of occupation of the end sites. The end-to-end and top-to-bottom degeneracy factors are indicated numerically proximate to the appropriate space.

the number of arrangements when one and only one of the four end sites is occupied. The end-to-end and top-to-bottom reflections of the arrangements contained in a_2 result in a degeneracy factor of four (4). The degeneracy factor associated with each of the seven subsets, a_j $(j=1,...,7)$, shown in Fig. 3 is indicated next to the corresponding figure.

Similarly, Fig. 4 serves to define sixteen subsets of $b\,[N,q,n_{11},n_{00}]$, the set of all arrangements of q particles on a $\beta(N)$-space that create prescribed numbers of occupied and

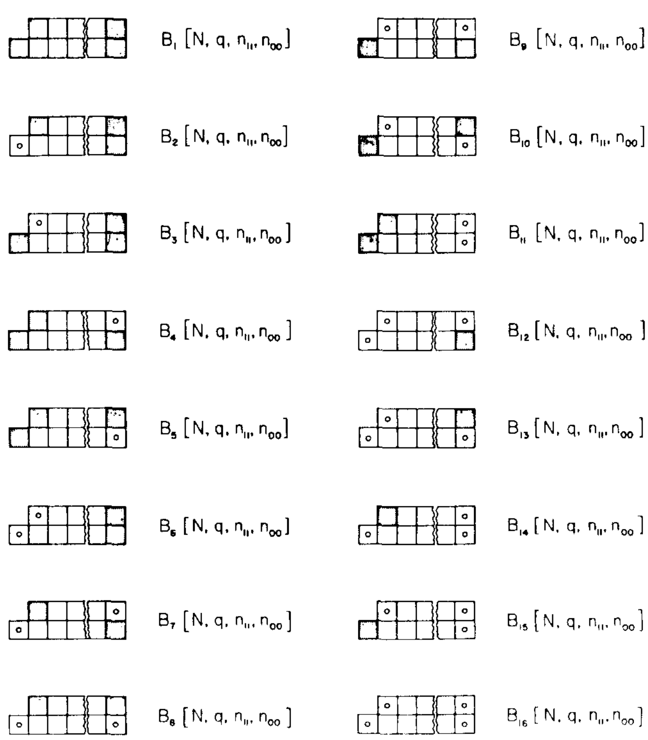

FIG. 4. Sixteen β-spaces are characterized by the state of occupation of the end sites.

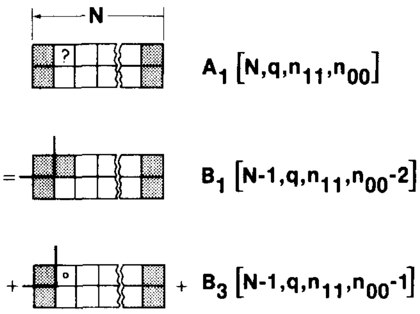

FIG. 5. This figure shows the decomposition of the set of $a_1[N,q,n_{11},n_{00}]$ into two mutually exclusive subsets on the basis of the state of occupation of the top site in the $N-1$ column.

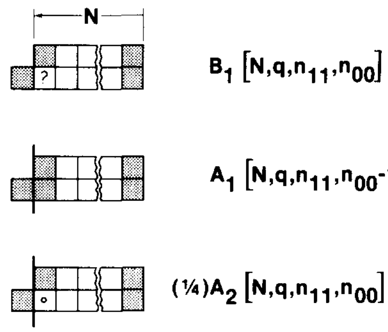

FIG. 6. This figure shows the decomposition of the set of $b_1[N,q,n_{11},n_{00}]$ into two mutually exclusive subsets on the basis of the state of occupation of the bottom site of the $N-1$ column.

vacant nearest-neighbor pairs. The degeneracy factor for all $b_j[N,q,n_{11},n_{00}]$ $(j=1,...,16)$ is unity. Thus, $b_4[N,q,n_{11},n_{00}]$ is the set of all arrangements on a $\beta(N)$-space when q, n_{11}, and n_{00} are specified and when the top right-hand site is occupied (and all the other end sites are vacant). Because

$$b\,[N,q,n_{11},n_{00}] = \bigcup_{j=1}^{16} b_j\,[N,q,n_{11},n_{00}] \tag{8}$$

we conclude that

$$\begin{aligned} B\,[N,q,n_{11},n_{00}] &\equiv \#b\,[N,q,n_{11},n_{00}] \\ &= \sum_{j=1}^{16} \#b_j\,[N,q,n_{11},n_{00}] \\ &= \sum_{j=1}^{16} B_j\,[N,q,n_{11},n_{00}]\ . \end{aligned} \tag{9}$$

Next we decompose each of the seven a_j's into two subsets, one in which the top site in the $(N-1)$th column is vacant and one in which that site is occupied. Similarly we decompose each of sixteen b_j's into two subsets, i.e., on the basis of the state of occupation of the top site of the $(N-1)$th column. There is closure in the sense that the decomposition of any of these twenty-three sets results in two subsets, both of which are members of the twenty-three sets.

As an example, consider $a_1[N,q,n_{11},n_{00}]$ (see Fig. 5). If the top site of the $(N-1)$th column is vacant, a $\beta(N-1)$-space is created on which all four end sites are vacant. The set of all arrangements of q particles on such a space is $b_1[N-1,q,n_{11},n_{00}-2]$ because n_{11} occupied nearest-neighbor pairs and $n_{00}-2$ vacant nearest-neighbor pairs must be created [two vacant nearest-neighbor pairs have been excluded from the $\beta(N-1)$-space].

If, on the other hand, the top site of the $(N-1)$th column is occupied, a $\beta(N-1)$-space is created on which the top left-hand site is occupied and the remaining three end sites are vacant. The set of all arrangements of q particles on such a space is $b_3[N-1,q,n_{11},n_{00}-1]$ (see Fig. 5).

Every arrangement in $b_1[N-1,q,n_{11},n_{00}-2]$ differs from every arrangement in $b_3[N-1,q,n_{11},n_{00}-1]$ by the state of occupation of the top site of the $(N-1)$th column, i.e.,

$$b_1[N-1,q,n_{11},n_{00}-2] \cap b_3[N-1,q,n_{11},n_{00}-1] = \varnothing\ . \tag{10}$$

In addition, every element of $a_1[N,q,n_{11},n_{00}]$ will be found either in $b_1[N-1,q,n_{11},n_{00}-2]$ or in $b_3[N-1,q,n_{11},n_{00}-1]$, i.e.,

$$\begin{aligned} a_1\,&[N,q,n_{11},n_{00}] \\ &= b_1[N-1,q,n_{11},n_{00}-2] \\ &\quad \cup b_3[N-1,q,n_{11},n_{00}-1]\ . \end{aligned} \tag{11}$$

We conclude that

$$\begin{aligned} \#a_1[N,q,n_{11},n_{00}] &= \#b_1[N-1,q,n_{11},n_{00}-2] \\ &\quad + \#b_3[N-1,q,n_{11},n_{00}-1]\ , \end{aligned}$$

or reindexing we obtain

$$\begin{aligned} A_1[N,q,n_{11},n_{00}+2] &= B_1[N,q,n_{11},n_{00}] \\ &\quad + B_3[N,q,n_{11},n_{00}+1]\ . \end{aligned} \tag{12a}$$

By decomposing the remaining a_j sets in a similar manner we arrive at the following equations

$$\tfrac{1}{4}A_2[N,q+1,n_{11}+1,n_{00}] = B_1[N-1,q,n_{11}+1,n_{00}] + B_3[N-1,q,n_{11},n_{00}]\ , \tag{12b}$$

$$\tfrac{1}{2}A_3[N,q+1,n_{11}+1,n_{00}] = B_4[N-1,q,n_{11}+1,n_{00}] + B_9[N-1,q,n_{11},n_{00}]\ , \tag{12c}$$

$$\tfrac{1}{2}A_4[N,q+1,n_{11}+1,n_{00}] = B_5[N-1,q,n_{11}+1,n_{00}] + B_{10}[N-1,q,n_{11},n_{00}]\ , \tag{12d}$$

$$\tfrac{1}{2}A_5[N,q+1,n_{11}+2,n_{00}] = B_2[N-1,q,n_{11}+1,n_{00}] + B_6[N-1,q,n_{11},n_{00}]\ , \tag{12e}$$

$$\tfrac{1}{4}A_6[N,q+1,n_{11}+2,n_{00}] = B_7[N-1,q,n_{11}+1,n_{00}] + B_{12}[N-1,q,n_{11}+n_{00}]\ , \tag{12f}$$

$$A_7[N,q+1,n_{11}+2,n_{00}] = B_{14}[N-1,q,n_{11}+1,n_{00}] + B_{16}[N-1,q,n_{11},n_{00}]\ . \tag{12g}$$

Similarly, we can decompose each of the sixteen b sets into two subsets on the basis of the state of occupation of the bottom site in the $(N-1)$th column. This gives rise to the following equations (see, e.g., Fig. 6):

$$B_1[N,q,n_{11},n_{00}+1] = A_1[N,q,n_{11},n_{00}] + \tfrac{1}{4}A_2[N,q,n_{11},n_{00}+1]\ , \tag{13a}$$

$$B_2[N,q+1,n_{11}+1,n_{00}] = A_1[N,q,n_{11}+1,n_{00}] + \tfrac{1}{4}A_2[N,q,n_{11},n_{00}]\ , \tag{13b}$$

$$B_3[N,q,n_{11},n_{00}+1] = \tfrac{1}{4}A_2[N,q,n_{11},n_{00}] + \tfrac{1}{2}A_5[N,q,n_{11},n_{00}+1]\,, \tag{13c}$$
$$B_4[N,q,n_{11},n_{00}+1] = \tfrac{1}{4}A_2[N,q,n_{11},n_{00}] + \tfrac{1}{2}A_4[N,q,n_{11},n_{00}+1]\,, \tag{13d}$$
$$B_5[N,q,n_{11},n_{00}+1] = \tfrac{1}{4}A_2[N,q,n_{11},n_{00}] + \tfrac{1}{2}A_3[N,q,n_{11},n_{00}+1]\,, \tag{13e}$$
$$B_6[N,q+1,n_{11}+1,n_{00}] = \tfrac{1}{4}A_2[N,q,n_{11}+1,n_{00}] + \tfrac{1}{2}A_5[N,q,n_{11},n_{00}]\,, \tag{13f}$$
$$B_7[N,q+1,n_{11}+1,n_{00}] = \tfrac{1}{4}A_2[N,q,n_{11}+1,n_{00}] + \tfrac{1}{2}A_4[N,q,n_{11},n_{00}]\,, \tag{13g}$$
$$B_8[N,q+1,n_{11}+1,n_{00}] = \tfrac{1}{4}A_2[N,q,n_{11}+1,n_{00}] + \tfrac{1}{2}A_3[N,q,n_{11},n_{00}]\,, \tag{13h}$$
$$B_9[N,q,n_{11},n_{00}+1] = \tfrac{1}{2}A_3[N,q,n_{11},n_{00}] + \tfrac{1}{4}A_6[N,q,n_{11},n_{00}+1]\,, \tag{13i}$$
$$B_{10}[N,q,n_{11},n_{00}+1] = \tfrac{1}{2}A_4[N,q,n_{11},n_{00}] + \tfrac{1}{4}A_6[N,q,n_{11},n_{00}+1]\,, \tag{13j}$$
$$B_{11}[N,q,n_{11},n_{00}+1] = \tfrac{1}{2}A_5[N,q,n_{11},n_{00}] + \tfrac{1}{4}A_6[N,q,n_{11},n_{00}+1]\,, \tag{13k}$$
$$B_{12}[N,q+1,n_{11}+1,n_{00}] = \tfrac{1}{2}A_3[N,q,n_{11}+1,n_{00}] + \tfrac{1}{4}A_6[N,q,n_{11},n_{00}]\,, \tag{13l}$$
$$B_{13}[N,q+1,n_{11}+1,n_{00}] = \tfrac{1}{2}A_4[N,q,n_{11}+1,n_{00}] + \tfrac{1}{4}A_6[N,q,n_{11},n_{00}]\,, \tag{13m}$$
$$B_{14}[N,q+1,n_{11}+1,n_{00}] = \tfrac{1}{2}A_5[N,q,n_{11}+1,n_{00}] + \tfrac{1}{4}A_6[N,q,n_{11},n_{00}]\,, \tag{13n}$$
$$B_{15}[N,q,n_{11},n_{00}+1] = \tfrac{1}{4}A_6[N,q,n_{11},n_{00}] + A_7[N,q,n_{11},n_{00}+1]\,, \tag{13o}$$
$$B_{16}[N,q+1,n_{11}+1,n_{00}] = \tfrac{1}{4}A_6[N,q,n_{11}+1,n_{00}] + A_7[N,q,n_{11},n_{00}]\,. \tag{13p}$$

Equations (12) and (13) may be solved for any one of these quantities $A_j[N,q,n_{11},n_{00}]$. For example solving for $A_1[N,q,n_{11},n_{00}]$ results in

$$\begin{aligned}
A_1&[N+3,q+3,n_{11}+4,n_{00}+4] \\
&= A_1[N+2,q+3,n_{11}+4,n_{00}+1] \\
&\quad + A_1[N+2,q+2,n_{11}+4,n_{00}+4] \\
&\quad + A_1[N+2,q+2,n_{11}+3,n_{00}+3] \\
&\quad + A_1[N+2,q+1,n_{11}+1,n_{00}+4] \\
&\quad + A_1[N+1,q+2,n_{11}+4,n_{00}+1] \\
&\quad - A_1[N+1,q+2,n_{11}+3,n_{00}] \\
&\quad + A_1[N+1,q+1,n_{11}+3,n_{00}+3] \\
&\quad - A\,[N+1,q+1,n_{11}+1,n_{00}+1] \\
&\quad + A_1[N+1,q,n_{11}+1,n_{00}+4] \\
&\quad - A_1[N+1,q,n_{11},n_{00}+3] \\
&\quad - A_1[N,q,n_{11}+3,n_{00}+3] \\
&\quad + 3A_1[N,q,n_{11}+2,n_{00}+2] \\
&\quad - 3A_1[N,q,n_{11}+1,n_{00}+1] \\
&\quad + A_1[N,q,n_{11},n_{00}]\,.
\end{aligned} \tag{14}$$

It is important to note that a recursion of the same form is obtained for all the $A_j[N,q,n_{11},n_{00}]$ as well as for all the $B_j[N,q,n_{11},n_{00}]$. The only consideration that differentiates these sets is the differences in the initial conditions. We conclude that the structure of the recursion given in Eq. (14) is valid for any and all *a* and *b* sets and represents a property of the $2\times N$ lattice.

In this regard it is interesting to note that if we subtract Eq. (2) from Eq. (3) we obtain

$$3[N-q] = n_{00} - n_{11} + \tfrac{1}{2}(\delta_0 - \delta_1)\,,$$

so that any change in the values of N, q, n_{11}, or n_{00} must conform to

$$3[\Delta N - \Delta q] = \Delta n_{00} - \Delta n_{11}\,. \tag{15}$$

An examination of the arguments of A_1 in Eq. (14) indicates that Eq. (15) is always satisfied.

III. GENERATING FUNCTIONS AND EXPECTATION

We first form the polynomials

$$f_{N,q}(x,y) = \sum_{n_{11}}\sum_{n_{00}} A\,[N,q,n_{11},n_{00}]x^{n_{11}}y^{n_{00}}\,. \tag{16}$$

If the value of n_{11} is specified then the sum over n_{00} contains five terms as can be seen from the elimination of n_{01} using Eqs. (2) and (3), i.e.,

$$3[N-q] + n_{11} + [(\delta_1 - \delta_0)/2] = n_{00}\,, \tag{17}$$

where $[(\delta_1 - \delta_0)/2]$ can take on the values $-2, -1, 0, +1, +2$.

Upon substituting into Eq. (16) a recursion for $A\,[N,q,n_{11},n_{00}]$ of the form given in Eq. (15) we obtain the following relationship for $f_{N,q}(x,y)$:

$$\begin{aligned}
f_{N+3,q+3}(x,y) &= y^3 f_{N+2,q+3}(x,y) \\
&\quad + (1+xy)f_{N+2,q+2}(x,y) \\
&\quad + y^3(1-xy)f_{N+1,q+2}(x,y) \\
&\quad + x^3 f_{N+2,q+1}(x,y) \\
&\quad + xy(1-x^2y^2)f_{N+1,q+1}(x,y) \\
&\quad + x^3(1-xy)f_{N+1,q}(x,y) \\
&\quad - xy(1-xy)^3 f_{N,q}(x,y)\,.
\end{aligned} \tag{18}$$

Equation (18), combined with the initial conditions

$$\begin{aligned}
&f_{1,0}(x,y) = y\,, \quad f_{1,1}(x,y) = 2\,, \quad f_{1,2}(x,y) = x\,, \\
&f_{2,0}(x,y) = y^4\,, \quad f_{2,1}(x,y) = 4y^2\,, \\
&f_{2,2}(x,y) = 2 + 4xy\,, \quad f_{2,3}(x,y) = 4x^2\,, \\
&f_{2,4}(x,y) = x^4\,, \\
&f_{3,0}(x,y) = y^7\,, \quad f_{3,1}(x,y) = 4y^5 + 2y^4\,, \\
&f_{3,2}(x,y) = 4y^3 + 4y^2 + 2xy^4 + 4xy^3 + xy^2\,, \\
&f_{3,3}(x,y) = 2 + 4xy^2 + 4xy + 6x^2y^2 + 4x^2y, \\
&f_{3,4}(x,y) = x^2y + 4x^2 + 4x^3y + 4x^3 + 2x^4y\,, \\
&f_{3,5}(x,y) = 2x^4 + 4x^5\,, \quad f_{3,6}(x,y) = x^7
\end{aligned} \tag{19}$$

will generate, as coefficients of the various powers of x and y, the required degeneracies.

When $x = y = 1$, Eq. (18) reduces to[7]

$$f_{N+3,q+3} = f_{N+2,q+3} + 2f_{N+2,q+2} + f_{N+2,q+1}, \quad (20)$$

which, as expected, is the recursion for $\binom{2(N+3)}{q+3}$.

If we next form the polynomials

$$h(x, y, z, \eta) = \sum_{N=1}^{\infty} \sum_{q} f_{N,q}(x, y) z^q \eta^N \quad (21)$$

and substitute Eq. (18) for $f_{N,q}(x, y)$, we obtain

$$\begin{aligned} h(x, y, z, \eta) = \eta\{&g_1(x, y, z) + \eta[g_2(x, y, z) - g_1(x, y, z)[y^3 + z(1 + xy) + x^3z^2] \\ &+ \eta^2[g_3(x, y, z) - g_2(x, y, z)[y^3 + z(1 + xy) + x^3z^2] \\ &- g_1(x, y, z)[y^3(1 - xy)z + xy(1 - x^2y^2)z^2 + x^3(1 - xy)z^3]]\}/\{1 - \eta[y^3 + (1 + xy)z + x^3z^2] \\ &- \eta^2[y^3(1 - xy)z + xy(1 - x^2y^2)z^2 + x^3(1 - xy)z^3] + \eta^3[xy(1 - xy)^2z^3]\}, \end{aligned} \quad (22)$$

where

$$g_N(x, y, z) \equiv \sum_q f_{N,q}(x, y) z^q .$$

For $x = y = z = 1$, Eq. (22) becomes

$$h(1,1,1, \eta) = 4\eta/(1 - 4\eta) . \quad (23)$$

To determine, from the recursion given in Eq. (14), $\langle n_{11} \rangle_{N,q}$, the expectation of n_{11}, we define

$$\langle \theta_{11} \rangle_{N,q} \equiv [3N - 2]^{-1} \langle n_{11} \rangle_{N,q} , \quad (24)$$

where

$$\langle n_{11} \rangle_{N,q} \equiv \binom{2N}{q}^{-1} \sum_{n_{11},n_{00}} n_{11} A[N,q,n_{11},n_{00}] \quad (25)$$

in which

$$\sum_{n_{11},n_{00}} A[N,q,n_{11},n_{00}] = \binom{2N}{q} . \quad (26)$$

Utilizing Eq. (14) in Eq. (25) we obtain a recursion for $\langle \theta_{11} \rangle_{N,q}$:

$$\begin{aligned} \binom{2N}{q}&(3N - 2)\langle \theta_{11} \rangle_{N,q} \\ &= \binom{2N-2}{q}(3N - 5)\langle \theta_{11} \rangle_{N-1,q} \\ &\quad + 2\binom{2N-2}{q-1}(3N - 5)\langle \theta_{11} \rangle_{N-1,q-1} \\ &\quad + \binom{2N-2}{q-2}(3N - 5)\langle \theta_{11} \rangle_{N-1,q-2} + 3\binom{2N-2}{q-2} . \end{aligned} \quad (27)$$

Assuming

$$\langle \theta_{11} \rangle_{N,q} = \langle \theta_{11} \rangle_{N-i,q-j} = \langle \theta_{11} \rangle$$

(i, j fixed and finite when N and $q \to \infty$ in such a way that $q/2N \equiv \theta$ remains fixed), we obtain

$$\begin{aligned} \langle \theta_{11} \rangle &= \lim_{\substack{N\to\infty \\ q\to\infty}} \langle \theta_{11} \rangle_{N,q} = \lim_{\substack{N\to\infty \\ q\to\infty}} \binom{2N-2}{q-2} \Big/ \binom{2N}{q} \\ &= \lim_{\substack{N\to\infty \\ q\to\infty}} \frac{q(q-1)}{2N(2N-1)} = \theta^2 . \end{aligned} \quad (28)$$

An examination of the arguments of the fundamental recursion, Eq. (14), reveals that there is a symmetry in the index subtracted from n_{11} and n_{00}. For example, $n_{11} - 3$ and $n_{00} - 3$ both appear four times. Thus we may conclude from Eq. (28) that $\langle \theta_{00} \rangle = (1 - \theta)^2$ and that $\langle \theta_{01} \rangle = 2\theta(1 - \theta)$.

IV. CONCLUSION

A recursion relation is developed that permits an exact determination of the number of arrangements of q simple, indistinguishable particles on a $2 \times N$ lattice that exhibits a prescribed number of occupied and vacant (and mixed) nearest-neighbor pairs. Utilizing this recursion we have calculated the associated generating functions and the expectation of n_{11}.

ACKNOWLEDGMENT

The research reported here has been supported in part by the Air Force Office of Scientific Research under Grant AFOSR 81-0192.

[1] R. B. McQuistan, J. Math. Phys. **13**, 1317 (1972).
[2] R. B. McQuistan, J. Math. Phys. **15**, 1845 (1974).
[3] R. B. McQuistan, Fibonacci Quarterly **14**, 353 (1976).
[4] R. B. McQuistan, J. Math. Phys. **22**, 1260 (1981).
[5] B. Kedem, J. Math. Phys. **22**, 456 (1981).
[6] C. C. Yan, J. Math. Phys. **17**, 69 (1976); J. Math. Phys. **17**, 1684 (1976).
[7] J. Riordan, *Combinatorial Identities* (Wiley, New York, 1968).

On the construction of state spaces for classical dynamical systems with a time-dependent Hamilton function[a)]

Günter Braunss
Mathematisches Institut der Justus-Liebig-Universität, D 6300 Giessen, Federal Republic of Germany

(Received 28 January 1982; accepted for publication 30 December 1982)

The damped linear and the forced harmonic oscillator are used as standard examples for a dynamical system with a time-dependent Hamilton function to investigate the problem of constructing a Hilbert state space and evolution operators in this space.

PACS numbers: 03.20. + i

INTRODUCTION

It has been shown[1–3] that if the spectrum of the one-parameter group $\alpha = \{\alpha_t \mid t\in\mathbb{R}\}$ of time evolution operators of a dynamical system in a Hilbert state space $\mathbb{H}$ has a nonvoid absolutely continuous part then the system can have a nonvanishing microscopic entropy production; i.e., there might exist a positive linear operator M with domain and range in $\mathbb{H}$ so that

$$\langle f_t, Mf_t\rangle > 0 \quad (f_t \equiv \alpha_t f), \tag{L1}$$

$$\frac{d}{dt}\langle f_t, Mf_t\rangle \leqslant 0 \tag{L2}$$

for all $t\geqslant 0$ and all f from a core D_M of M; M is called a *Lyapunov operator* for α and the operator Γ defined by $-\alpha_t^*\Gamma\alpha_t f = d(\alpha_t^* M\alpha_t)f/dt$ is called the corresponding *microscopic entropy production operator*. For a conservative Hamiltonian system the group α is determined by the corresponding Liouville operator and the state space is constructed in the usual fashion as introduced by Koopman.[4] For a system with a time-dependent Hamiltonian this method cannot be applied since there is no constant energy submanifold. This has caused doubt[5] whether a statistical description which goes along with the introduction of a state space makes sense at all. We do not share this opinion since, even if the energy is not a constant of motion, there are other constants of motion which might specify suitable invariant manifolds Σ with an appropriate measure μ so that $L^2(\Sigma,\mu)$ is a reasonable state space—at least for our purpose, meaning that we can use its states and evolution operators with regard to microscopic entropy and asymptotic stability properties. We shall consider here two standard examples—the damped linear oscillator and the forced harmonic oscillator (both in one dimension)—to demonstrate that there are at least three possible ways to construct a state space: The first one uses the method of suspension[4] (i.e., the extended cotangent bundle). In the second method the system is imbedded into a larger conservative system. It turns out that both methods (at least for our examples) are mathematically equivalent, delivering the same state space and evolution operators and differing only in an interpretation of parameters. In the third and last method the system is mapped by a canonical transformation on a conservative Hamiltonian system for which the state space is constructed in the Koopman fashion. Since this canonical transformation is unitary in the state space, the time evolution operators are unitary. However, although the result is perhaps more satisfactory from a physical point of view since no artificial parameters are used it has a handicap: The evolution operators do not constitute a group (whereas they do so in the first two methods); hence the spectral criterium for the existence of an entropy production mentioned above cannot be applied.

I. USING THE EXTENDED PHASE SPACE

Let $T^*(M)\subset\mathbb{R}^2$ be a cotangent bundle and let $H{:}T^*(M)\times\mathbb{R}\to\mathbb{R}$ be a time-dependent Hamilton function. Let further $T^*(M_e)$ be the cotangent bundle of the extended phase space and let[4,6,7]

$$\widetilde{H}(q,p,E,t) = H(p,q,t) + E \quad (=\text{const}) \tag{1}$$

be the suspension of $H(p,q,t)$ so that the dynamical equations of (1) are

$$\frac{dp}{ds} = -\frac{\partial\widetilde{H}}{\partial q},\quad \frac{dE}{ds} = -\frac{d\widetilde{H}}{\partial t},$$
$$\frac{dq}{ds} = \frac{\partial\widetilde{H}}{\partial t},\quad \frac{dt}{ds} = \frac{\partial\widetilde{H}}{\partial E}. \tag{2}$$

From the last equation we obtain $t = \tau + s$. Assume now that x and K are constants of motion so that

$$q(t) = a(t;x,K),\quad p(t) = b(t;x,K). \tag{3}$$

Since the volume on $T^*(M_e)$ is $dp\wedge dE\wedge dq\wedge dt = -dp\wedge dq\wedge dt\wedge d\widetilde{H}$, it follows[4] that

$$dp\wedge dq\wedge dt = D\,dx\wedge d\tau\wedge dK,$$
$$D = \frac{\partial a}{\partial x}\frac{\partial b}{\partial K} - \frac{\partial a}{\partial K}\frac{\partial b}{\partial x}, \tag{4}$$

is an invariant measure of the flow of (1). As a first example we consider the damped linear oscillator with a Hamilton function[8,9]

$$H(p,q,t) = (e^{-kt}p^2 + e^{kt}q^2)/2 \tag{5}$$

and solutions

$$q(t) = (2K)^{1/2}e^{-kt/2}\sin(\omega t + t_0)$$
$$= (2K)^{1/2}e^{-k(s+\tau)/2}\sin(\omega s + x) \equiv \tilde{q}(s), \tag{6a}$$

$$p(t) = (2K)^{1/2}e^{kt/2}\cos(\omega t + t_0 + \theta)$$
$$= (2K)^{1/2}e^{k(s+\tau)/2}\cos(\omega s + x + \theta) \equiv \tilde{p}(s), \tag{6b}$$

[a)] Dedicated to Bryce DeWitt on his sixtieth birthday.

where $\omega^2 = 1 - k^2/4 > 0$ and $\theta = \arccos \omega$. Thus $D = \omega$ in (4). As an invariant submanifold we choose now the one specified by $K = \text{const}$ so that our state space becomes $\mathbb{H} = L^2(T^1 \times \mathbb{R}, dx \times d\tau)$ (with T^1 denoting the one-dimensional torus and dx and $d\tau$ denoting the Lebesgue measures on T^1 and $\mathbb{R}$, respectively). The corresponding Liouville operator is then $L = -i(\omega\partial/\partial x + \partial/\partial\tau)$. Since $i\partial/\partial\tau$ has an absolutely continuous spectrum, we can expect a nonvanishing microscopic entropy. Indeed, it has been shown[3] that the shift on $\mathbb{R}$ determines a class of nontrivial Lyapunov operators (the point is that these Lyapunov operators are not classical observables meaning that they do not commute with all classical observables[1,2]). The constant $(2K)^{1/2}$ which specifies our invariant submanifold has an obvious physical meaning namely that of an amplitude at a fixed time and phase x (thus it can be related to the number of particles or quanta of an ensemble). K appears also with an ergodic property of $H(p, q, t)$ for one shows easily

$$\lim_{T\to\infty}\left[(1/T)\int_0^T H\,dt\right] = \int_0^{2\pi} H\,dx = K. \tag{7}$$

In Sec. III we shall show that K is, in addition, the energy of a harmonic oscillator which is uniquely related to the damped linear oscillator via a canonical map.

As a second example, we consider now the forced harmonic oscillator with a Hamilton function[6]

$$H(p,q,t) = (p^2+q^2)/2 + c(t)q \tag{8}$$

and solutions

$$\begin{aligned} q(t) &= (2K)^{1/2}\sin(t+x) + q_*(t), \\ p(t) &= (2K)^{1/2}\cos(t+x) + p_*(t), \quad p_* = \dot{q}_*, \end{aligned} \tag{9}$$

where q_* is a particular solution depending on c. It follows $D = -1$ in (4). Assume first that c is (smooth) periodic, say mod $\omega 2\pi \neq 2\pi$. Then the (relevant) state space will be $\mathbb{H} = L^2(T^1 \times \omega T^1, dx \times d\tau)$ and the evolution is given by corresponding shifts on T^1 and ωT^1. The Liouville operator has in this case a pure point spectrum so that we cannot expect entropy production but rather stability (which seems obvious from the assumed properties of c). Assume now that c is not periodic but otherwise reasonable and defined on $\overline{\mathbb{R}}$ or $\overline{\mathbb{R}}_+$ [say, $c(t) = e^{-t}$]. Then we get a state space $\mathbb{H} = L^2(T^1 \times \mathbb{R}, dx \times d\tau)$ and the Liouville operator has due to the shift of τ on $\mathbb{R}$ a nonvoid absolutely continuous spectral part. Hence we can expect entropy production [which again seems obvious from the assumed properties of $c(t)$].

II. EMBEDDING INTO A CONSERVATIVE SYSTEM

We can write the solution of the damped linear oscillator in the following way: $q = \overline{Q}_1\overline{Q}_2$, $p = \overline{P}_1\overline{P}_2$, where $\overline{Q}_1(t) = e^{-kt/2}$, $\overline{Q}_2(t) = (2K)^{1/2}\sin(\omega t + x)$, $\overline{P}_1(t) = e^{kt/2}$, $\overline{P}_2(t) = (2K)^{1/2}\cos(\omega t + x + \theta)$ belong to Hamilton functions

$$\overline{H}_1(\overline{P}_1, \overline{Q}_1) = -k\overline{P}_1\overline{Q}_1/2 = \text{const} = k/2, \tag{10a}$$

$$\overline{H}_2(\overline{P}_2, \overline{Q}_2) = (\overline{P}_2^2 + \overline{Q}_2^2 + k\overline{P}_2\overline{Q}_2)/2 = \text{const} = K. \tag{10b}$$

The transformation $(\overline{Q}_1, \overline{Q}_2) \to (Q_1 = \overline{Q}_1, Q_2 = q = F(\overline{Q}_1, Q_2))$ is canonical if the momenta are transformed according to $\overline{P}_1 = P_1 + p\overline{Q}_2$, $\overline{P}_2 = p\overline{Q}_1 = pQ_1$. We get then

$$\begin{aligned} H(P_1, Q_1, p, q) &= -kP_1Q_1/2 + (p^2Q_1^2 + q^2Q_1^{-2})/2 \\ &= \overline{H}_1(\overline{P}_1, \overline{Q}_1) + \overline{H}_2(\overline{P}_2, \overline{Q}_2) \end{aligned} \tag{11}$$

as the Hamilton function of a conservative system in which the damped linear oscillator is a subsystem. Replace for the sake of numerical simplicity in (10b) $\overline{Q}_1^2$ by $\omega_0^2\overline{Q}_1^2$ and let $\omega = (\omega_0^2 - k^2/4)^{1/2} = k = 1$. Letting now $\overline{Q}_1(t) = e^{-(t+x_1)}$, $\overline{Q}_2(t) = (2\overline{E}_2)^{1/2}\sin(t + x_2)$, where $\overline{H}_1 = \overline{E}_1$, $\overline{H}_2 = \overline{E}_2$, $H = E = \overline{E}_1 + \overline{E}_2$, we obtain the Koopman state space $\overline{\mathbb{H}} = L^2(\mathbb{R} \times T^1 \times \mathbb{R}_+, dx_1 \times dx_2 \times dx_3)$, $x_3 = \overline{E}_2$. Thus our relevant state space is $\mathbb{H} = L^2(\mathbb{R} \times T^1, dx_1 \times dx_2)$ with evolution operators which act as shifts on T^1 and $\mathbb{R}$. Hence we have a nonvoid absolutely continuous spectral part for the Liouville operator, so we can expect entropy production. Looking back to Sec. I, we see that our state space and the evolution operators are exactly those we obtained there, although the parameters τ and x_2 have a different meaning.

As to the forced harmonic oscillator, consider the Hamilton function

$$\begin{aligned} H(p_1, q_1, p_2, q_2) = (p_1^2 + a_1q_1^2)/2 + (p_2^2 + a_2q_2^2)/2 \\ - b_1p_1p_2 - b_2q_1q_2, \end{aligned} \tag{12}$$

where a_1, a_2, b_1, b_2 are real constant. If we choose these constants such that $a_1b_1 + b_2 = 0$ and $a_1(1 - b_1^2) = 1$, then we get the following equations of motion:

$$\ddot{q}_1 + q_1 = b_1(a_2 - a_1)q_2, \tag{13a}$$

$$\ddot{q}_2 + bq_2 = 0, \quad b \equiv a_2 - a_1b_1^2. \tag{13b}$$

That is, (13a) is a forced harmonic oscillator and a subsystem of the conservative system which has a Hamilton function (11). The special choices for the exterior force of this subsystem which are provided by q_2 will be sufficient for our purpose; they are: (i) $b < 0$, (ii) $1 \neq b > 0$. For (i) we may choose then $q_2(t) = e^{-(t+x_2)}$, and for (ii) we can let $q_2(t) = \sin(\sqrt{b}\,t + x_2)$. This leads via the Koopman construction with respect to (11) for (i) to the relevant state space $\mathbb{H} = L^2(T^1 \times \mathbb{R}, dx_1 \times dx_2)$ and for (ii) to $\mathbb{H} = L^2(T^1 \times T^1, dx_1 \times dx_2)$, and the evolution operators are in both cases the respective shifts. For (i) this means a nonvoid and for (ii) an empty absolutely continuous spectral part for the corresponding Liouville operator. Thus we have obtained again the same state space and evolution operators as in Sec. I.

Note that the matrix associated with the bilinear form (11) is Hermitian so that we can find $\bar{p}_1, \bar{q}_1, \bar{p}_2, \bar{q}_2$ such that $H(P_1, Q_1, p, q) = \overline{H}_1(\bar{p}_1, \bar{q}_1) + \overline{H}_2(\bar{p}, \bar{q}_2)$.

In Appendix A we consider the field equations for a spatially homogeneous scalar field with nonzero rest mass in de Sitter space, the scalar field playing the part of a damped linear oscillator as a subsystem of a closed Hamiltonian system (scalar + gravitational field).

In Appendix B we give another example of a conservative Hamiltonian system not equivalent to (11), of which the damped linear oscillator is a subsystem; the resulting Hilbert state space, however, coincides with those of Secs. I and II.

III. MAPPING ON A FREE HARMONIC OSCILLATOR

Let (p, q) belong to the damped linear oscillator and consider the transformation

$$(p, q) \to (\bar p, \bar q): q(t) = e^{-kt/2}\bar q(t), \quad p(t) = e^{kt/2}\bar p(t). \quad (14)$$

This is a canonical map provided

$$p\,dq - H\,dt = \bar p\,d\bar q - \bar H\,dt, \qquad \bar H(\bar p, \bar q) = (\bar p^2 + \bar q^2 + k\bar p\bar q)/2. \quad (15)$$

Evidently, (15) maps the damped linear oscillator (5) on a harmonic oscillator with a Hamilton function which is given in (15). Note that this transformation is generated by a Poisson bracket: If

$$\beta_t = e^{it\delta}, \quad \delta = (ik/2)\{\bar p\bar q, \cdot\}_{\text{Poisson}}, \quad (16)$$

then one easily verifies

$$p(t) = \beta_t \bar p(t), \quad q(t) = \beta_t \bar q(t). \quad (17)$$

So if $\bar\alpha_t$ is the shift $\bar p(0) \to \bar p(t) = (\bar\alpha_t \bar p)(0)$, $\bar q(0) \to \bar q(t) = (\bar\alpha_t \bar q)(0)$, it follows that, with an initial condition $p(0) = \bar p(0)$, $q(0) = \bar q(0)$, the time evolution operators for the damped linear oscillator are $\alpha_t = \bar\alpha_t \beta_t$, $t \in \mathbb{R}$. Since $\bar\alpha_t$ and β_t do not commute (unless $t = 0$), the α_t do not form a group. However, they are unitary in the state space which we can construct as the Koopman space $\mathbb{H}$ for the Hamilton function $\bar H(\bar p, \bar q) = (\bar p^2 + \bar q^2 + k\bar p\bar q)/2 = \text{const} = E$. Proceeding in the usual fashion, we obtain $\mathbb{H} = L^2(T^1, dx)$, where x refers to

$$\bar q(t) = (2E)^{1/2}\sin(\omega t + x), \qquad \bar p(t) = (2E)^{1/2}\cos(\omega t + x + \theta), \quad \theta = \arccos\omega. \quad (18)$$

To show that the α_t are unitary, it suffices to prove that δ as given by (16) is s. a. in $\mathbb{H}$. Now, a short calculation yields

$$(-i\delta f)(x) = [\omega - \lambda(x)]f'(x), \quad f \in C^1(0, 2\pi), \quad (19)$$

where

$$\lambda(x) = \omega[1 - (k/2)\sin(2x + \theta)][1 + (k/2)^2\sin^2 x]^{-1}.$$

Since λ is evidently real bounded and periodic (mod2π), all[10] maximal extensions of δ are s. a. with domain $D_\delta = \{f \in C^1(0, 2\pi) | f(0) = zf(2\pi)\}, |z| = 1$ (we let $z = 1$ so that D_δ is a set of single-valued functions). This proves the unitarity of the α_t in $\mathbb{H}$. The strong derivative of α_t with respect to t is, by the way, the time-dependent Liouville operator L associated with the Hamilton function (5). In the state space $\mathbb{H}$ it is given by

$$(-iLf)(x) = \lambda(x)f'(x). \quad (20)$$

By the same arguments that were used to prove that δ is s. a., it can be shown that L is s. a., having the same domain as δ. It is easy to show (confirming thus a general theorem proven in Ref. 2) that there is no Lyapunov operator M which is a multiplication function (that is, M is not a classical observable). For assume the contrary to be true; i.e., let $(Mf)(x) = m(x)f(x)$. Then it follows from the Lyapunov conditions (L1) and (L2) above that

$$-i(Lm)(x) = \lambda(x)m'(x) = \gamma(x) \geqslant 0, \quad (21)$$

where $m > 0$ must be single-valued. Consequently,

$$0 = m(2\pi) - m(0) = \int_0^{2\pi} \lambda(x)^{-1}\gamma(x)\,dx, \quad (22)$$

from which it follows $\gamma = 0$ a.e.; hence there is no entropy production.

The forced harmonic oscillator can be treated in a similar way. Let its Hamilton function be given by (8), and let $\bar q$ and $\bar p$ be solutions of the corresponding free harmonic oscillator. Consider the map

$$\bar q \to q = \bar q + q_*, \quad \bar p \to p = \bar p + p_*, \quad p_* = \dot q_*, \quad (23)$$

where q and p are the solutions given in (9). It is easy to show that (23) is, too, generated by a Poisson bracket, namely $\{-p_*\bar q + q_*\bar p, \cdot\}_{\text{Poisson}}$. Proceeding similarly as above, we find that [for suitable functions c in (8)] the time evolution operators are unitary in the Koopman state space associated with the harmonic oscillator but do not constitute a group.

ACKNOWLEDGMENTS

I am indebted to the referee for pointing out a number of papers which had escaped my attention as well as for elucidating some physical aspects. For discussions I am further indebted to P. Edelmann and H. Ruppenthal.

APPENDIX A

As a physically relevant example in which a damped linear oscillator appears as a subsystem of a conservative Hamiltonian system, we consider a scalar field with a nonzero rest mass coupled to a gravitational field, the scalar field playing the part of the oscillator. Assume that we have (in a four-dimensional space–time manifold) a Robertson–Walker metric $ds^2 = -dt^2 + S^2\,d\sigma^2$, $S \equiv S(t)$. Let ϕ be a scalar field depending on time only which is coupled to the gravitational field represented by S. The Einstein field equations are then[11] (the dot denotes differentation with respect to t)

(a) $E = \text{const} = -\dot S^2 + (4\pi/3)S^2(\dot\phi^2 + \omega^2\phi^2)$,

(b) $\ddot\phi + 3(\dot S/S)\dot\phi + \omega^2\phi = 0$.

If we substitute $dt \to S^{-1}dt$ and introduce

$$q_1 = S^2, \quad p_1 = S\dot S/4, \quad q_2 = (8\pi/3)^{-1/2}\phi, \quad p_2 = (8\pi/3)^{1/2}S^4\dot\phi,$$

then Eqs. (a) and (b) are equivalent to a Hamiltonian system with a Hamilton function

$$(*) \quad H(p_1, q_1, p_2, q_2) = -p_1^2 + (q_1^{-2}p_2^2 + \omega^2 q_1^2 q_2)/2,$$

that is, q_2 is a scalar field with a rest mass ω coupled to a gravitational field q_1 in a metric $ds^2 = -q_1^{-1}\,dt^2 + q_1\,d\sigma^2$, where $d\sigma^2$ represents a spatial 3-sphere whose curvature E is given by the values of the level set of H, i.e., $H = \text{const} = E$. We let it suffice to remark that from the structure of the Hamilton function (*) one can conjecture the associated Liouville operator to have absolutely continuous parts in its spectrum; hence one can expect microscopic entropy production and thus asymptotic stability.

APPENDIX B

The Hamiltonian (11) in which the damped linear oscillator appeared as a subsystem is not the only one which has this property. For example, the Hamilton function

$$H(p_1, q_1, p_2, q_2) = p_1 p_2 + \tfrac{1}{2}k(p_1 q_1 - p_2 q_2) + \omega^2 q_1 q_2, \quad \omega^2 = 1 - k^2/4,$$

which stems from the Morse–Feshbach Lagrangian (see Ref. 8) and where q_2 is the coordinate function of the damped linear oscillator (5) (q_1 is the coordinate function of the same linear oscillator but with a negative damping $-k$), serves the same purpose. Its solutions are

$$\text{(a)}\quad \begin{aligned} q_2(t) &= Ce^{-k(t+\tau)/2}\sin(\omega t + x_2),\\ p_2(t) &= Ce^{k(t+\tau)+2}\cos(\omega t + x_1),\end{aligned}$$

$$\text{(b)}\quad \begin{aligned} q_1(t) &= Ce^{k(t+\tau)/2}\sin(\omega t + x_1),\\ p_1(t) &= Ce^{-k(t+\tau)}\cos(\omega t + x_2),\end{aligned}$$

$$\text{(c)}\quad \omega C^2\cos(x_2 - x_1 - \theta) = H = \text{const},\quad \theta = \arccos\omega.$$

In order that $H \geqslant 0$ and C be real, we let $x_2 = x_1$ and $C = \text{const}$ real. It is no difficulty to see from (a) and (b) that one obtains then the same state space as in Secs. I and II, namely $L^2(T^1 \times \mathbb{R}, dx_1 \times d\tau)$ and the same evolution operators (which are shifts on T^1 and $\mathbb{R}$). The nonequivalence of the Morse–Feshbach Hamiltonian and (11) follows from the different spectra of the matrices corresponding to their bilinear forms.

APPENDIX C

We have considered in detail so far only two examples of linear systems. Therefore, we would like to outline briefly a nonlinear example, namely that of a damped anharmonic oscillator $\ddot{q} + k\dot{q} + V'(q) = 0$, which has a (time-dependent) Hamilton function $H(p,q,t) = e^{-kt}p^2/2 + e^{kt}V(q)$. If we let $V(q) = aq^2/2 + q^{n+1}(n+1)$, where $a = 2k^2(n+1)(n+3)^{-2}$, then $q(t) = e^{-bt}v(e^{-\beta t})$ where v is a solution of $\beta^2 v'' + v^n = 0$ and $\beta = b(n-1)/2$, $b = 2k(n+3)^{-1}$. Let, in particular, $k = n = 3$ so that $\beta = b = 1 = 2a$. The substitution $e^{-t} \to t, p \to p, q \to q$ is a canonical transformation if $H \to -t^2p^2/2 - t^{-4}V(q)$ [where $V(q) = q^2 + q^4/4$]. Our solution becomes then $q(t) = Ct\,\mathrm{cn}(Ct + x)$, where $C = $ real const and cn$\equiv$cosinus amplitudinis (mod $1/\sqrt{2}$). We can repeat now all the constructions of Secs. I and II. As to Sec. I, this is obvious from the general scheme we have provided there. As to Sec. II it suffices to note that we have a factorization $q = \bar{q}_1\bar{q}_2$, where $\bar{q}_1$ and $\bar{q}_2$ belong to systems with Hamilton functions $\bar{H}_1 = \bar{p}_1^2/2$ and $\bar{H}_2 = \bar{p}_2^2/2 + \bar{q}_2^4/4$, respectively. Hence we can copy the above given construction. Again, we get as the (relevant) state space $L^2(\omega T^1 \times \mathbb{R})$ and evolution operators which are the respective shifts. The Morse–Feshbach Hamiltonian in Appendix B can, by the way, be generalized so as to include the damped anharmonic oscillator, for if q_2 denotes its coordinate function, then one checks easily that $H(p_1,q_1,q_2) = p_1p_2 + \alpha p_1q_1 - \beta p_2q_2 + q_1F(q_2)$ yields for q_2 the following equation of motion:

$$\ddot{q}_2 + (\alpha + \beta)\dot{q}_2 + \alpha\beta q_2 + F(q_2) = 0,$$

that is,

$$V'(q_2) = \alpha\beta q_2 + F(q_2).$$

APPENDIX D

Although we are not concerned here with quantization problems, we would like to point out a somewhat strange result concerning these problems. Let us both in the (properly symmetrized) quantum mechanical Hamiltonian corresponding to (11) (case A) and the Morse–Feshbach type (case B) for the damped linear oscillator assume, as usual, $i[P_j, Q_k] = \delta_{jk}1$ and $[P_j, P_k] = [Q_j, Q_k] = 0$, and let $\dot{Q}_j \equiv dQ_j/dt$ (strongly). Omitting all finesse concerning domains of operators, it follows then that

$$\text{(A)}\quad \begin{aligned} i[\dot{Q}_1(t), Q_1(t)] &= i[\dot{Q}_1(t), \dot{Q}_2(t)] = 0,\\ i[\dot{Q}_2(t), Q_2(t)] &= Q_1^2(t) = e^{-kt}Q_1^2(0);\end{aligned}$$

$$\text{(B)}\quad \begin{aligned} [\dot{Q}_1(t), Q_1(t)] &= [\dot{Q}_2(t), Q_2(t)] = 0,\\ i[\dot{Q}_1(t), \dot{Q}_2(t)] &= -k\,1.\end{aligned}$$

If we use the construction of Sec. III, that is, if we let $P(t) = \exp(it\delta)\bar{P}(t)$, $Q(t) = \exp(it\delta)\bar{Q}(t)$, where $\delta = \frac{1}{4}ikad(\bar{P}\bar{Q} + \bar{Q}\bar{P})$ [cf. (16)] and the couple $(\bar{P}, \bar{Q})$ belongs to the harmonic oscillator in (15), then it follows that $i[P, Q] = i[\bar{P}, \bar{Q}] = 1$ and $i[\dot{Q}(t), Q(t)] = e^{-kt}1$. The last equation falls with respect to the commutator of $\dot{Q}_2(t)$ and $Q_2(t)$ at least qualitatively in line with case A, although both results contradict arguments raised in Ref. 12. By the way, the example considered in Appendix A (a scalar field q_2 coupled to a gravitational field with a metric $ds^2 = q_1^{-1}dt^2 + q_1^2\,d\sigma^2$) would (properly quantized) lead to $i[\dot{Q}_1,Q_1]/2 = -1$, $i[\dot{Q}_2,Q_2] = Q_1^{-2}$, and $i[\dot{Q}_1,\dot{Q}_2] = 4P_2Q_1^{-3}$ so that, in particular, the second commutator equation agrees due to the expected asymptotic properties of Q_1 (expanding universe) qualitatively with the corresponding one in case (A).

APPENDIX E

So far we have considered the example of the damped linear oscillator, $\ddot{q} + k\dot{q} + q = 0$, either as derived from a time-dependent Hamiltonian or as part of a larger conservative system. However, this equation can also be derived from a time-independent (one-dimensional) Hamiltonian (cf. Ref. 13):

$$H = -kpq/2 - \ln[\cos(\omega pq)] + \ln q,\quad \omega^2 = 1 - k^2/4.$$

To see what the state space corresponding to $H = $ const looks like let $q(t) = Ce^{-k(t+\tau)/2}\cos[\omega(t+\tau)]$ be the general solution with constants of motion $\tau \in \mathbb{R}$, $C \in \mathbb{R}_+$. Then a short calculation yields $H = \ln C$. The volume element in phase space is

$$dp \wedge dq = dH \wedge dq/[-(kq/2) + \omega q\tan(\omega pq)] = dH \wedge dq/\dot{q}.$$

Hence the invariant measure on the submanifold specified by $H = \ln C = $ const is $dq/\dot{q} = -\omega\,d\tau$, where $d\tau$ is the Lebesque measure on $\mathbb{R}$. Thus our state space is $\tilde{\mathbb{H}} = L^2(\mathbb{R})$ and the Liouville operator is $L = -i\partial/\partial\tau$. Consequently, L has an absolutely continuous spectrum, so we can expect asymptotic Lyapunov stability in $\tilde{\mathbb{H}}$. Comparing this with the above result, where the state space was $\mathbb{H} = L^2(T^1 \times \mathbb{R}) \simeq L^2(T^1) \otimes L^2(\mathbb{R})$, we see that $\tilde{\mathbb{H}}$ is actually the restriction of $\mathbb{H}$ to that factor (subsystem), namely $L^2(\mathbb{R})$, on which dynamics (i.e., the shift $t \to t + \tau$) provides us already with the premises which guarantee asymptotic stability. The reason why $\tilde{\mathbb{H}}$ and $\mathbb{H}$ do not coincide is due to our choice of the corresponding constants of motion: For $\mathbb{H}$ we had let $q(t) = (2K)^{1/2}e^{-kt/2}\cdot\cos[\omega(t+\tau)]$, $\omega\tau = t_0 + \pi/2$ (cf. Sec. I), where the invariant submanifold Σ was specified by $K = $ const. In the here considered example we had let $q(t) = Ce^{-k(t+\tau)/2}\cos[\omega(t+\tau)]$ and Σ specified by $H = \ln C = $ const, thus $d(\ln C) = dK + K\,d\tau$.

[1]B. Misra, Proc. Natl. Acad. Sci. USA **75** (4), 1627–1631 (1978).
[2]G. Braunss, "Asymptotic stability and microscopic entropy," (to be published).
[3]G. Braunss, "On a class of Lyapunov operators," Mitt. Math. Sem. Giessen, 149 (1981).
[4]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin-Cummings, Reading, MA, 1978).
[5]P. Havas, Acta Phys. Austr., 38 (1973).
[6]W. Thirring, *Lehrbuch der Mathematischen Physik I* (Springer, Vienna, 1977).
[7]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vols. I, II.
[8]R. M. Santilli, *Foundations of Theoretical Mechanics I* (Springer, New York, 1978).
[9]H. H. Denman and L. H. Buch, J. Math. Phys. **14**, 326 (1973).
[10]N. Dunford and J. T. Schwartz, *Linear Operators II* (Interscience, New York, 1963).
[11]G. F. R. Ellis and S. W. Hawking, *The Large Scale Structure of Space–Time* (Cambridge U.P., Cambridge, 1973).
[12]W. E. Brittin, Phys. Rev. **77**, 3 (1950).
[13]P. Havas, Nuovo Cimento Suppl. **5**, 363 (1957).

Symplectic approach to nonconservative mechanics

F. Cantrijn[a)]
Instituut voor Theoretische Mechanica, Rijksuniversiteit Gent, Krijgslaan 281-S9, B-9000 Gent, Belgium

(Received 14 December 1982; accepted for publication 11 February 1983)

The dynamics of autonomous nonconservative systems is studied in terms of Lagrangian submanifolds of a special symplectic manifold. Both the Hamiltonian and Lagrangian description are taken into consideration and the transition between the two descriptions is established by means of the generating function of a symplectic relation.

PACS numbers: 03.20. + i, 02.40.Sf, 02.30.Jr

1. INTRODUCTION

In several papers,[1] Tulczyjew has developed a new symplectic formulation of particle dynamics which also has an interesting extension to field dynamics.[2] His approach, which differs from the standard symplectic treatments of mechanics,[3] applies to nonrelativistic as well as relativistic systems and allows a more uniform treatment of Hamiltonian and Lagrangian dynamics. Considering, for instance, the case of a conservative nonrelativistic particle system, we can sketch the main idea as follows. Let M be a C^∞-differentiable manifold with cotangent bundle T^*M and tangent bundle TM (the phase space and state space of the system). The canonical symplectic form on T^*M induces a symplectic form on TT^*M which corresponds to two different special symplectic structures,[1] associated with the fibrations $TT^*M \rightarrow T^*M$ and $TT^*M \rightarrow TM$, respectively. It then turns out that the tangent vectors to the phase space trajectories of the system constitute a Lagrangian submanifold of TT^*M (this is sometimes called the "reciprocity property of particle dynamics"[1]). According to whether one regards this Lagrangian submanifold as being (locally) generated with respect to the first special symplectic structure by a function H defined on T^*M, or with respect to the second one by a function L defined on TM, one recovers the Hamilton or Lagrange equations of motion, respectively. The transition between the two descriptions is effectuated by a Legendre transformation which can be characterized by means of the generating function of a symplectic relation, namely, the graph of the identity transformation of TT^*M (see, e.g., Ref. 4).

The purpose of the present paper is to construct a similar symplectic scheme for autonomous systems which can be described by equations of the form

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q^i} + Q_i, \tag{1}$$

for some smooth functions H and Q_i on T^*M. In particular, this includes the phase space description of classical mechanical systems with forces not derivable from a potential[5] (friction forces, gyroscopic forces, ···). We will therefore refer to systems of type (1) as nonconservative systems.

Since our analysis will primarily be based on Refs. 1 and 4, we first recall in the next section some definitions and properties from these papers. Starting from a geometric characterization of a nonconservative system, it will be seen in Sec. 3 that the dynamics of such a system defines a Lagrangian submanifold of TT^*M with respect to a particular symplectic form. In Sec. 4 we then show that this symplectic form corresponds to two special symplectic structures in such a way that the description of a Lagrangian submanifold of TT^*M in terms of one of these structures produces the phase space equations (1) of a nonconservative system, whereas the description in terms of the other structure yields a second-order system

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = \widetilde{Q}_i \tag{2}$$

for some smooth functions L and $\widetilde{Q}_i$ on TM. The latter equations agree with the general form of the equations of motion in state space of a nonconservative mechanical system.[5] The transition between the two descriptions by means of a Legendre transformation will be briefly discussed in Sec. 5. Finally, in Sec. 6 we conclude with some general remarks and indicate, as an illustration, how any dynamical system on T^*M can be lifted to a Hamiltonian system on TT^*M within the present symplectic framework.

For definitions and properties from symplectic geometry we refer, e.g., to Abraham and Marsden,[3] and Weinstein.[6] The notations we adopt are mainly those of Ref. 3. For any manifold P, the natural projections from the cotangent bundle and the tangent bundle onto P will be denoted by π_P: $T^*P \rightarrow P$ and τ_P: $TP \rightarrow P$, respectively. The canonical 1-form on T^*P will be denoted by θ_P and the corresponding symplectic form by ω_P (i.e., $\theta_P = p_i dq^i$ and $\omega_P = d\theta_P$). All mappings, vector fields, and differential forms are assumed to be of class C^∞.

2. PRELIMINARIES

For a detailed description of the concepts and for a proof of the properties mentioned in this section, we refer to the papers of Refs. 1 and 4.

Definition 2.1: A *special symplectic manifold* is a collection $(P, Q, \pi, \theta, \chi)$ where P, Q are differentiable manifolds, π: $P \rightarrow Q$ a differential fibration, θ a 1-form on P, and χ: $P \rightarrow T^*Q$ a diffeomorphism such that $\pi_Q \circ \chi = \pi$ and $\chi^*\theta_Q = \theta$.

In particular, it follows from this definition that $(P, d\theta)$ is a symplectic manifold: the *underlying symplectic manifold* of the special symplectic structure $(P, Q, \pi, \theta, \chi)$.

Property 2.1: If $(P, Q, \pi, \theta, \chi)$ is a special symplectic

[a)] Research associate at the National Fund for Scientific Research, Belgium.

manifold, K a submanifold of Q, and f a smooth function on K, then[7]

$$N = \{ p\in P:\pi(p)\in K,\ \langle u,\theta(p)\rangle = \langle T\pi(u),df(\pi(p))\rangle \text{ for all } u\in T_pP \text{ with } T\pi(u)\in T_{\pi(p)}K \}$$

is a Lagrangian submanifold of $(P, d\theta)$. (See, e.g., Ref. 4 or also, for a more general discussion, Ref. 8.) N is called the Lagrangian submanifold generated by f.

Property 2.2: If $(P_i, Q_i, \pi_i, \theta_i, \chi_i)$ for $i = 1, 2$ are two special symplectic manifolds, then $(P_2\times P_1, Q_2\times Q_1, \pi_2\times\pi_1, \theta_2\ominus\theta_1, \chi_{21})$ is a special symplectic manifold, with $\theta_2\ominus\theta_1 = pr_2^*\theta_2 - pr_1^*\theta_1$; $pr_i: P_2\times P_1\to P_i$ the natural projections; $\chi_{21}: P_2\times P_1\to T^*(Q_2\times Q_1)$, $(p_2, p_1)\mapsto(\chi_2(p_2), -\chi_1(p_1))$.

Let P be a differentiable manifold and denote the graded algebras of differential forms on P and on TP by $\Omega(P)$ and $\Omega(TP)$, respectively. One can then introduce two particular derivation operators I_T and D_T from $\Omega(P)$ into $\Omega(TP)$ as follows.[1,4]

Let f be a smooth function and α a 1-form, both defined on P. Put $I_Tf = 0$ and define $I_T\alpha$ as a function on TP by $I_T\alpha(x) = \langle x,\alpha(\tau_P(x))\rangle$ for each $x\in TP$. Since derivations are completely determined by their action on functions and 1-forms, it follows that I_T can be extended to a derivation of degree -1 from $\Omega(P)$ into $\Omega(TP)$. A derivation of degree 0 is obtained by putting $D_T = I_Td + dI_T$.

If $P = T^*M$ for some differentiable manifold M, denote the natural bundle coordinates on P and TP $(= TT^*M)$ by (q^i, p_i) and $(q^i, p_i, \dot q^i, \dot p_i)$, respectively. We then have

$$I_T\omega_M = \dot p_i dq^i - \dot q^i dp_i, \tag{3}$$

$$D_T\theta_M = \dot p_i dq^i + p_i d\dot q^i. \tag{4}$$

It can be verified that the 2-form $dI_T\omega_M$ $(= dD_T\theta_M)$ is a symplectic form on TT^*M.

As is well known, on the second tangent bundle TTP there exists a canonical involution s_P which, in terms of natural bundle coordinates $(q^i, \dot q^i, u^i, v^i)$ is given by[9]

$$s_P(q^i,\dot q^i,u^i,v^i) = (q^i,u^i,\dot q^i,v^i). \tag{5}$$

The following properties of s_p are immediately verified:

$$s_P\circ s_P = 1_{TTP},\quad \tau_{TP}\circ s_P = T\tau_P,\quad T\tau_P\circ s_P = \tau_{TP}. \tag{6}$$

We finally mention a property which, when properly extended, leads to an equivalent characterization of the derivation operator D_T (cf. the second paper of Ref. 1). Let $\alpha\in\Omega^1(P)$ and $y\in TTP$ be given. Suppose ϕ: $\mathbb{R}\to TP$ is an integral curve (or representative) of $s_P(y)$, i.e., $\phi(0) = \tau_{TP}(s_P(y))$ and $T\phi(0,1) = s_P(y)$. Then

$$\langle y,D_T\alpha\rangle = \frac{d}{dt}\langle\phi,\alpha\rangle(0), \tag{7}$$

where $\langle\phi,\alpha\rangle(t)\equiv\langle\phi(t),\alpha(\tau_P(\phi(t)))\rangle$.

3. NONCONSERVATIVE SYSTEMS AND LAGRANGIAN SUBMANIFOLDS

Let M be a C^∞-differentiable manifold (dimension n) and μ a nonclosed horizontal 1-form on T^*M, i.e., $d\mu\neq 0$ and in local canonical coordinates

$$\mu = Q_i dq^i \tag{8}$$

for some smooth functions $Q_i = Q_i(q,p)$.

Given any function $H\in C^\infty(T^*M)$, we adopt the following definition:

Definition 3.1: The nonconservative system associated with the pair (μ,H) is the unique vector field $\Delta\in\mathscr{X}(T^*M)$ for which

$$i_\Delta\omega_M = -dH + \mu. \tag{9}$$

Expressing (9) in terms of local coordinates, using (8) and the expression $\omega_M = dp_i\wedge dq^i$, it is seen that the differential equations associated with Δ are of the form (1). Clearly, any other pair (μ', H') consisting of a horizontal 1-form μ' and a smooth function H' will define the same nonconservative system iff there exists a function $f\in C^\infty(M)$ such that $\mu' = \mu + d(f\circ\pi_M)$ and $H' = H + f\circ\pi_M$. This freedom in the characterization of a nonconservative system in terms of a pair (μ, H) will have no influence on the subsequent analysis. It will therefore always be tacitly assumed that a fixed choice for μ has been made. The function H will then be called the Hamiltonian of the nonconservative system. Before proceeding, a few remarks are in order here.

It is clear from the previous considerations that we restrict ourselves to systems having a globally defined Hamiltonian. Alternatively, we could have defined a nonconservative system as any vector field Δ for which

$$L_\Delta\omega_M = d\mu \tag{10}$$

for some horizontal 1-form μ (where L_Δ denotes the Lie derivative with respect to Δ). Obviously, (9) implies (10) whereas, in general, the converse only holds locally. The requirement of μ being nonclosed is added to exclude the possibility that Δ trivially becomes a (local) Hamiltonian vector field.

Finally, the reason for taking μ to be horizontal is mainly based on the following two arguments. First of all, as mentioned above, this condition naturally leads to the phase space form of the equations of motion of a nonconservative mechanical system. The functions Q_i in (8) can then be interpreted as the phase space components of the (generalized) forces which are not derivable from a potential. Secondly, as will be seen in the next section, horizontality of μ is an important assumption with a view on the transition to the state space (or Lagrangian) description of nonconservative systems. Nevertheless, as far as the phase space portrait is concerned, much of the sequel holds equally well when starting from a general 1-form μ on T^*M. In particular, this would give rise to a more symmetric form of the equations of motion, with an additional term appearing on the right-hand side of the $\dot q^i$-equations in (1). The passage from (9), with μ horizontal, to this more general situation can be established by means of an arbitrary symplectic transformation.[10]

We now turn to the characterizaiton of a nonconservative system, associated with a given μ, in terms of a Lagrangian submanifold of TT^*M with respect to a particular symplectic structure. By means of μ and the canonical symplectic form ω_M, we define the following 1-form on TT^*M:

$$\theta_0 = I_T\omega_M - \tau^*_{T^*M}\mu. \tag{11}$$

In terms of the natural coordinates $(q^i, p_i, \dot q^i, \dot p_i)$ on TT^*M and using (3), the local expression for θ_0 becomes

$$\theta_0 = (\dot p_i - Q_i)dq^i - \dot q^i dp_i. \quad (12)$$

In the first paper of Ref. 1 it has been shown that $I_T\omega_M = \beta^*\theta_{T^*M}$, where $\beta: TT^*M \to T^*T^*M$ is the bundle isomorphism defined by

$$\beta(x) = i_x\omega_M(\tau_{T^*M}(x)) \quad (13)$$

for every $x \in TT^*M$. Hence, θ_0 can also be written as

$$\theta_0 = \beta^*\theta_{T^*M} - \tau^*_{T^*M}\mu. \quad (14)$$

One immediately verifies that the 2-form $\omega_0 = d\theta_0$ is a symplectic form on TT^*M. Regarding a vector field Δ on T^*M as a smooth section of TT^*M, i.e., $\Delta: T^*M \to TT^*M$ and $\tau_{T^*M} \circ \Delta = \mathbb{1}_{T^*M}$, we get the following result [which is intuitively trivial on the local expression (12)].

Proposition 3.1: A vector field $\Delta \in \mathscr{X}(T^*M)$ defines a nonconservative system, associated with μ, iff

$$\Delta^*\theta_0 = -dH \quad (15)$$

for some function $H \in C^\infty(T^*M)$.

Proof: Using (14) we find

$$\Delta^*\theta_0 = (\beta \circ \Delta)^*\theta_{T^*M} - (\tau_{T^*M} \circ \Delta)^*\mu$$
$$= (\beta \circ \Delta)^*\theta_{T^*M} - \mu.$$

Taking account of the property[11] $(\beta \circ \Delta)^*\theta_{T^*M} = \beta \circ \Delta$ and the definition (13) of β, this can be rewritten as

$$\Delta^*\theta_0 = i_\Delta\omega_M - \mu, \quad (16)$$

from which the result follows in view of Definition 3.1. ■

Denoting the image set of Δ in TT^*M by Im Δ, the previous proposition immediately yields the following corollary.

Corollary 3.2: If Δ is a nonconservative system, defined by (9) for some smooth function H, then Im Δ is a Lagrangian submanifold of (TT^*M, ω_0).

Proof: Im Δ is a submanifold of TT^*M with dim(Im Δ) $= \frac{1}{2}$ dim(TT^*M) and, moreover, from (15) we get $\Delta^*\omega_0 = 0$, i.e., $\omega_0|_{\mathrm{Im}\,\Delta} = 0$. ■

The converse of this corollary holds if, instead of (9), we consider the relation (10). More precisely we have:

Proposition 3.3: Given $\Delta \in \mathscr{X}(T^*M)$, then Im Δ is a Lagrangian submanifold of (TT^*M, ω_0) if and only if $L_\Delta\omega_M = d\mu$ [i.e., if and only if (9) holds locally].

Proof: The proof follows immediately from the observation that (16) implies $\Delta^*\omega_0 = L_\Delta\omega_M - d\mu$. ■

In the next section, it will be shown that (TT^*M, ω_0) is the underlying symplectic manifold of two special symplectic structures, associated with the fibrations $\tau_{T^*M}: TT^*M \to T^*M$ and $T\pi_M: TT^*M \to TM$, respectively. This will enable us to recover both the phase space and the state space description of a nonconservative system from the identification of a suitable Lagrangian submanifold of (TT^*M, ω_0).

4. SPECIAL SYMPLECTIC STRUCTURES

We first introduce the following mapping:

$$\chi: TT^*M \to T^*T^*M$$
$$x \to \chi(x) = i_x\omega_M(\tau_{T^*M}(x)) - \mu(\tau_{T^*M}(x)).$$

Since ω_M is nondegenerate, χ is clearly a bundle isomorphism and so, in particular, a diffeomorphism. In local coordinates, using the expression (8) for μ, we have

$$\chi(q^i, p_i, \dot q^i, \dot p_i) = (q^i, p_i, \dot p_i - Q_i, -\dot q_i). \quad (17)$$

Comparing the definition of χ with (13), it is seen that $\chi = \beta - \mu \circ \tau_{T^*M}$. Consequently, it follows that

$$\chi^*\theta_{T^*M} = \beta^*\theta_{T^*M} - \tau^*_{T^*M}(\mu^*\theta_{T^*M}).$$

Using the fact that[11] $\mu^*\theta_{T^*M} = \mu$, and taking into account (14), we get

$$\chi^*\theta_{T^*M} = \theta_0.$$

Moreover, the definition of χ immediately implies that $\pi_{T^*M} \circ \chi = \tau_{T^*M}$ [which can also be seen from (17)]. We may therefore conclude that the collection $(TT^*M, T^*M, \tau_{T^*M}, \theta_0, \chi)$ satisfies all the requirements of Definition 2.1 and, hence, we have:

Proposition 4.1: $(TT^*M, T^*M, \tau_{T^*M}, \theta_0, \chi)$ is a special symplectic manifold with underlying symplectic manifold (TT^*M, ω_0). ■

Given $H \in C^\infty(T^*M)$, the Lagrangian submanifold of (TT^*M, ω_0) generated with respect to the special symplectic structure $(TT^*M, T^*M, \tau_{T^*M}, \theta_0, \chi)$ by $-H$, is locally described by

$$(\dot p_i - Q_i)dq^i - \dot q^i dp_i = -dH(q, p). \quad (18)$$

[See Property 2.1 and the expression (12) for θ_0.]

From (18) one derives the phase space equations of motion (1) of a nonconservative system.

We now proceed towards the construction of a second special symplectic manifold corresponding to the same underlying symplectic manifold (TT^*M, ω_0). The whole argument is inspired by and is an immediate extension of Tulczyjew's symplectic approach of Lagrangian dynamics.[1]

For $x \in TT^*M$ and $y \in TTM$, with

$$\tau_{TM}(y) = T\pi_M(x), \quad (19)$$

let $\gamma: \mathbb{R} \to TM$ and $\kappa: \mathbb{R} \to T^*M$ be integral curves of $s_M(y)$ and x, respectively, with s_M the involution operator on TTM (see Sec. 2), and where

$$\tau_M \circ \gamma = \pi_M \circ \kappa. \quad (20)$$

In particular, one may then consider the function $\langle \gamma, \kappa \rangle(t) \equiv \langle \gamma(t), \kappa(t) \rangle$. We now construct a mapping Ψ: $TT^*M \to T^*TM$ by the following prescription:

$$\langle y, \Psi(x) \rangle = \frac{d}{dt} \langle \gamma, \kappa \rangle(0) - \langle z, \mu(\tau_{T^*M}(x)) \rangle \quad (21)$$

for all $y \in TTM$ satisfying (19) and where $z \in TT^*M$ is such that

$$\tau_{T^*M}(z) = \tau_{T^*M}(x) \quad \text{and} \quad T\pi_M(z) = \gamma(0). \quad (22)$$

In order to see that (21) makes sense and, as such, defines $\Psi(x)$ unambiguously, we first notice that the first term on the right-hand side is independent of the chosen representatives (i.e., integral curves) γ and κ of $s_M(y)$ and x, for which (20) holds. Secondly, we must prove that the right-hand side of (21) is also independent of the choice made for z, provided (22) is satisfied. For that purpose it suffices to show that $\langle \tilde z, \mu(\tau_{T^*M}(x)) \rangle = 0$ whenever $T\pi_M(\tilde z) = 0$. For brevity, put $\tau_{T^*M}(x) = \alpha$. Choose $\tilde z \in T_\alpha(T^*M)$ with $T\pi_M(\tilde z) = 0$. Since μ is horizontal, one can always find a suitable 1-form ν on M

such that $\mu(\alpha) = (\pi_M^* \nu)(\alpha)$, at the given point α. Hence,

$$\langle \tilde z,\mu(\alpha)\rangle = \langle \tilde z,(\pi_M^*\nu)(\alpha)\rangle$$
$$= \langle T\pi_M(\tilde z),\nu(\pi_M(\alpha))\rangle = 0,$$

which gives the desired result.

Expressing (21) in local coordinates, using (8) and taking account of (19), (20), and (22), one can verify that Ψ is locally given by

$$\Psi(q^i,p_i,\dot q^i,\dot p_i) = (q^i,\dot q^i,\dot p_i - Q_i, p_i). \tag{23}$$

Next, we introduce the following 1-form on TT^*M:

$$\tilde\theta_0 = D_T\theta_M - \tau_{T^*M}^*\mu. \tag{24}$$

From the definition of the derivation operator D_T and (11), this can be rewritten as

$$\tilde\theta_0 = \theta_0 + dI_T\theta_M. \tag{25}$$

We now have the following result.

Proposition 4.2: $(TT^*M, TM, T\pi_M, \tilde\theta_0, \Psi)$ is a special symplectic manifold with underlying symplectic manifold (TT^*M, ω_0).

Proof: By means of (21) and (23) one readily verifies that Ψ is a bijective local diffeomorphism and, hence, a diffeomorphism. From the construction of Ψ it also follows that

$$\pi_{TM}\circ\Psi = T\pi_M. \tag{26}$$

Let $w\in TTT^*M$ and recall that θ_{TM} denotes the canonical 1-form on T^*TM.

For simplicity, we omit in the subsequent computations the indication of the base point at which the inner products are taken. We then have

$$\langle w,\Psi^*\theta_{TM}\rangle = \langle T\Psi(w),\theta_{TM}\rangle,$$

or, using the definition of the canonical 1-form,[11]

$$\langle w,\Psi^*\theta_{TM}\rangle = \langle (T\pi_{TM}\circ T\Psi)(w),\tau_{T^*TM}(T\Psi(w))\rangle$$
$$= \langle T(\pi_{TM}\circ\Psi)(w),\Psi(\tau_{TT^*M}(w))\rangle.$$

Taking account of (26) we finally obtain

$$\langle w,\Psi^*\theta_{TM}\rangle = \langle TT\pi_M(w),\Psi(\tau_{TT^*M}(w))\rangle \tag{27}$$

for all $w\in TTT^*M$.

Before proceeding we first mention the following relations:

$$\tau_{TM}\circ TT\pi_M = T\pi_M\circ\tau_{TT^*M}, \tag{28a}$$
$$TT\pi_M\circ s_{T^*M} = s_M\circ TT\pi_M, \tag{28b}$$
$$T\tau_{T^*M}\circ s_{T^*M} = \tau_{TT^*M}. \tag{28c}$$

These relations are most easily verified in local coordinates. For instance, denoting the natural coordinates on TTT^*M by $(q,p,\dot q,\dot p,u,\bar u,v,\bar v)$ we have $\tau_{TT^*M}(q,p,\dot q,\dot p,u,\bar u,v,\bar v) = (q, p, \dot q, \dot p)$. On the other hand, taking account of (5), we get $(T\tau_{T^*M}\circ s_{T^*M})(q,p,\dot q,\dot p,u,\bar u,v,\bar v) = T\tau_{T^*M}(q,p,u,\bar u,\dot q,\dot p,v,\bar v) = (q,p,\dot q,\dot p)$, by which (28c) is verified. The proof of (28a) and (28b) is completely similar.

Given $w\in TTT^*M$, put $x = \tau_{TT^*M}(w)\in TT^*M$ and $y = TT\pi_M(w)\in TTM$. Using (28a) it is seen that

$$\tau_{TM}(y) = (T\pi_M\circ\tau_{TT^*M})(w)$$

and, hence, (19) is satisfied. By (28b) we also have

$$s_M(y) = TT\pi_M(s_{T^*M}(w)). \tag{29}$$

Now, let ζ: $\mathbb{R}\to TT^*M$ be an integral curve of $s_{T^*M}(w)$, i.e., $T\zeta(0,1) = s_{T^*M}(w)$. Putting $\gamma = T\pi_M\circ\zeta$, we find

$$T\gamma(0,1) = (TT\pi_M\circ T\zeta)(0,1),$$
$$= TT\pi_M(s_{T^*M}(w)),$$

and so, by (29),

$$T\gamma(0,1) = s_M(y).$$

This shows that γ is an integral curve of $s_M(y)$. Analogously, putting $\kappa = \tau_{T^*M}\circ\zeta$ and using (28c), one can verify that κ is an integral curve of x, and, moreover, γ and κ satisfy (20). Returning to (27) and taking account of (21), we then get

$$\langle w,\Psi^*\theta_{TM}\rangle = \frac{d}{dt}\langle\gamma,\kappa\rangle(0) - \langle z,\mu\rangle \tag{30}$$

for any $z\in TT^*M$ satisfying (22). Since $\gamma(0) = (\tau_{TM}\circ s_M)(y)$, successive application of (29), (28a), and (28c) gives

$$\gamma(0) = (T\pi_M\circ T\tau_{T^*M}\circ s_{T^*M}\circ s_{T^*M})(w).$$

Recalling that s_{T^*M} is an involution operator [see the first relation of (6)], we find

$$\gamma(0) = (T\pi_M\circ T\tau_{T^*M})(w),$$

from which it follows that $z = T\tau_{T^*M}(w)$ satisfies (22). With this choice for z and replacing γ and κ by $T\pi_M\circ\zeta$ and $\tau_{T^*M}\circ\zeta$, respectively, (30) becomes

$$\langle w,\Psi^*\theta_{TM}\rangle = \frac{d}{dt}\langle T\pi_M\circ\zeta,\tau_{T^*M}\circ\zeta\rangle(0) - \langle T\tau_{T^*M}(w),\mu\rangle,$$
$$= \frac{d}{dt}\langle\zeta,\theta_M\rangle(0) - \langle w,\tau_{T^*M}^*\mu\rangle,$$

where again use has been made of the definition of θ_M.[11] By (7) we finally obtain

$$\langle w,\Psi^*\theta_{TM}\rangle = \langle w,D_T\theta_M\rangle - \langle w,\tau_{T^*M}^*\mu\rangle$$

or, with (24),

$$\langle w,\Psi^*\theta_{TM}\rangle = \langle w,\tilde\theta_0\rangle.$$

Since this relation holds for all $w\in TTT^*M$, it follows that

$$\Psi^*\theta_{TM} = \tilde\theta_0.$$

Equation (25) finally shows that $d\tilde\theta_0 = d\theta_0 = \omega_0$, which completes the proof. ■

With (4) and (8), the local expression for $\tilde\theta_0$ becomes

$$\tilde\theta_0 = (\dot p_i - Q_i)dq^i + p_i d\dot q^i. \tag{31}$$

Now let N be a Lagrangian submanifold of (TT^*M, ω_0) which, in the sense of Property 2.1, is generated by a function $L\in C^\infty(TM)$ with respect to the special symplectic structure $(TT^*M, TM, T\pi_M, \tilde\theta_0, \Psi)$. Using (31), N is then locally described by

$$(\dot p_i - Q_i)dq^i + p_i d\dot q^i = dL,$$

which is equivalent to

$$p_i = \frac{\partial L}{\partial\dot q^i},\quad \dot p_i = \frac{\partial L}{\partial q^i} + Q_i.$$

This clearly leads to the state space equations (2) of a nonconservative system, with $\tilde Q_i(q,\dot q)\equiv Q_i(q,\partial L/\partial\dot q)$.

Summarizing, the dynamics of a nonconservative system defined by (9) is characterized by a Lagrangian submanifold of (TT^*M, ω_0), which is generated with respect to $(TT^*M, T^*M, \tau_{T^*M}, \theta_0, \chi)$ by the function $-H$. If this Lagrangian submanifold can moreover be generated with respect to $(TT^*M, TM, T\pi_M, \tilde{\theta}_0, \Psi)$ by a function $L \in C^\infty(TM)$, we find an equivalent state space description for the given system. Of course, the existence of such a (globally defined) Lagrangian depends on certain regularity conditions, but it is not our intention to go further into this matter here.

In the next section we will see how the transition between the two descriptions of a nonconservative system can be formulated in terms of a Legendre transformation.

5. THE LEGENDRE TRANSFORMATION

This section is a straightforward application of the general theory of Legendre transformations as developed by Tulczyjew.[4] We therefore omit the details.

For convenience, we first recall some definitions from Ref. 4. Suppose (P,ω) is the underlying symplectic manifold of two special symplectic manifolds $(P, Q_1, \pi_1, \theta_1, \chi_1)$ and $(P, Q_2, \pi_2, \theta_2, \chi_2)$.

Definition 5.1: The transition from the description of a Lagrangian submanifold of (P, ω) in terms of a generating function with respect to $(P, Q_1, \pi_1, \theta_1, \chi_1)$ to the description in terms of a generating function with respect to $(P, Q_2, \pi_2, \theta_2, \chi_2)$ is called the *Legendre transformation* from $(P, Q_1, \pi_1, \theta_1, \chi_1)$ to $(P, Q_2, \pi_2, \theta_2, \chi_2)$. The identity transformation 1_P of P being a symplectic transformation, its graph is a Lagrangian submanifold of $(P \times P, \omega \ominus \omega)$, with $\omega\ominus\omega = pr_2^*\omega - pr_1^*\omega$.[3,6] (For a mapping ϕ: $P \rightarrow P$, graph ϕ is here defined as $\{(\phi(p), p): p \in P\}$.) According to Property 2.2, we also have that $(P \times P, \omega\ominus\omega)$ is the underlying symplectic manifold of the special symplectic manifold $(P \times P, Q_2 \times Q_1, \pi_2 \times \pi_1, \theta_2 \ominus \theta_1, \chi_{21})$.

Definition 5.2: The generating function E_{21} of graph 1_P with respect to $(P \times P, Q_2 \times Q_1, \pi_2 \times \pi_1, \theta_2\ominus\theta_1, \chi_{21})$ is called the generating function of the Legendre transformation from $(P, Q_1, \pi_1, \theta_1, \chi_1)$ to $(P, Q_2, \pi_2, \theta_2, \chi_2)$.

We now apply this to the situation described in the previous section. As we have shown, (TT^*M, ω_0) is the underlying symplectic manifold of the two special symplectic manifolds $(TT^*M, T^*M, \tau_{T^*M}, \theta_0, \chi)$ and $(TT^*M, TM, T\pi_M, \tilde{\theta}_0, \Psi)$. Again, using Property 2.2, it follows that $(TT^*M \times TT^*M, TM \times T^*M, T\pi_M \times \tau_{T^*M}, \tilde{\theta}_0 \ominus \theta_0, \Phi)$ is a special symplectic manifold with $\Phi(x, x') = (\Psi(x), -\chi(x'))$. Let E' denote the generating function of graph 1_{TT^*M} with respect to this special symplectic structure. By Definition 5.2, E' is then the generating function of the Legendre transformation from $(TT^*M, T^*M, \tau_{T^*M}, \theta_0, \chi)$ to $(TT^*M, TM, T\pi_M, \tilde{\theta}_0, \Psi)$, i.e., from the phase space to the state space description of a nonconservative system associated with μ. The next proposition shows that E' is independent of μ and, in fact, coincides with the generating function of the inverse Legendre transformation of particle dynamics.[4]

Proposition 5.1: E' is defined on the Whitney sum[12] $TM \times_M T^*M$ by

$$E'(v,\alpha) = \langle v,\alpha \rangle. \tag{32}$$

Proof: Let δ: $TT^*M \rightarrow TT^*M \times TT^*M$ denote the diagonal mapping, i.e., $\delta(x) = (x,x)$. We then have, using (11) and (24),

$$\begin{aligned}\delta^*(\tilde{\theta}_0 \ominus \theta_0) &= D_T\theta_M - \tau^*_{T^*M}\mu - I_T\omega_M + \tau^*_{T^*M}\mu \\ &= dI_T\theta_M.\end{aligned}$$

The remainder of the proof is now completely similar to the proof of Proposition 6.1 in Ref. 4. ■

Let H be a smooth function on T^*M and let N be the Lagrangian submanifold of (TT^*M, ω_0), which is generated with respect to $(TT^*M, T^*M, \tau_{T^*M}, \theta_0, \chi)$ by $-H$.

Under suitable assumptions[4] it can be shown that a generating function of N with respect to $(TT^*M, TM, T\pi_M, \tilde{\theta}_0, \Psi)$ will be given by

$$L(v) = \mathrm{Stat}_{(\alpha)}[E'(v,\alpha) - H(\alpha)], \tag{33}$$

where, following the notation of Ref. 4, $\mathrm{Stat}_{(\alpha)}$ indicates that for each v the function $E'(v,\cdot) - H(\cdot)$ is evaluated at a critical point α. It is clear that, in general, the function L will only be defined on a submanifold of TM.

In local coordinates, with $v = (q,\dot{q})$ and $\alpha = (q, p)$, we recover from (33) and (32) the well-known relationship between the Hamiltonian and the Lagrangian, namely,

$$L(q,\dot{q}) = \dot{q}^i p_i(q,\dot{q}) - H(q, p(q,\dot{q})),$$

where the functions $p_i(q,\dot{q})$ are obtained by solving the relations

$$\dot{q}^j = \frac{\partial H}{\partial p_j}(q, p)$$

with respect to the momenta.

Similarly, it can be shown that the Legendre transformation from $(TT^*M, TM, T\pi_M, \tilde{\theta}_0, \Psi)$ to $(TT^*M, T^*M, \tau_{T^*M}, \theta_0, \chi)$ is generated by the function E which is defined on the Whitney sum $T^*M \times_M TM$ by

$$E(\alpha,v) = -\langle v,\alpha \rangle,$$

and the transition from a given Lagrangian L to a corresponding Hamiltonian H then reads

$$H(\alpha) = -\mathrm{Stat}_{(v)}[E(\alpha,v) + L(v)].$$

6. CONCLUSIONS

Inspired by Tulczyjew's symplectic treatment of particle dynamics, we have constructed a symplectic framework for the description of nonconservative dynamical systems defined by (9). Some additional remarks are in order.

First, it should be emphasized that the symplectic structure we have introduced on TT^*M depends on the given horizontal 1-form μ (or, more precisely, on $d\mu$). Hence, to the extent that μ can be interpreted in practical applications as the representative of forces which are not derivable from a potential, the symplectic form ω_0 will also depend on these forces.

Secondly, although we have always confined ourselves to systems having a globally defined Hamiltonian, it is clear that the above treatment immediately extends to systems whose dynamics can be characterized by a Lagrangian submanifold of (TT^*M, ω_0), which is generated by a function defined on a submanifold of T^*M.

It is a well-known property that by doubling the degrees of freedom any dynamical system can be cast into Hamiltonian form. We finally want to illustrate that an analogous property can be formulated within the present framework. More precisely, we will show how the special symplectic structure $(TT^*M, T^*M, \tau_{T^*M}, \theta_0, \chi)$ can be used to lift any vector field on T^*M, and thus also, in particular, the vector field Δ defined by (9), to a global Hamiltonian vector field on TT^*M.

Recall that any diffeomorphism $\phi: P \to Q$ between two differentiable manifolds, lifts to a symplectic diffeomorphism $T^*\phi: T^*Q \to T^*P$ for which

$$(T^*\phi)^*\theta_P = \theta_Q \tag{34}$$

(see, e.g., Ref. 3, p. 180). Moreover, if $\phi_i: P_i \to P_{i+1}$ $(i = 1, 2)$ are diffeomorphisms, then

$$T^*(\phi_2 \circ \phi_1) = (T^*\phi_1) \circ (T^*\phi_2). \tag{35}$$

By means of the mapping χ, associated with the special symplectic structure $(TT^*M, T^*M, \tau_{T^*M}, \theta_0, \chi)$ one can now lift any diffeomorphism $\phi: T^*M \to T^*M$ to a diffeomorphism $\hat{\phi}: TT^*M \to TT^*M$, which is defined by

$$\hat{\phi} = \chi^{-1} \circ (T^*\phi)^{-1} \circ \chi.$$

One immediately verifies that

$$\tau_{T^*M} \circ \hat{\phi} = \phi \circ \tau_{T^*M}. \tag{36}$$

Since $\chi^*\theta_{T^*M} = \theta_0$ and taking account of (34), it follows that

$$\hat{\phi}^*\theta_0 = \theta_0, \tag{37}$$

and thus, in particular, $\hat{\phi}^*\omega_0 = \omega_0$.

Consequently, $\hat{\phi}$ is a symplectic diffeomorphism of (TT^*M, ω_0). For any two diffeomorphism ϕ_1, ϕ_2: $T^*M \to T^*M$, one finds with (35),

$$(\phi_2 \circ \phi_1)\hat{} = \hat{\phi}_2 \circ \hat{\phi}_1. \tag{38}$$

Let $X \in \mathscr{X}(T^*M)$ be an arbitrary vector field with (local) flow[3] consisting of the one-parameter group $\{\phi_s : s \in I\}$, where $I \subseteq \mathbb{R}$ is some open interval. Then, using (38), it can be seen that $\{\hat{\phi}_s : s \in I\}$ is a (local) one-parameter group of diffeomorphisms on TT^*M and let $\hat{X}$ denote the vector field which generates it. In view of (36), it readily follows that $T\tau_{T^*M} \circ \hat{X} = X \circ \tau_{T^*M}$ (i.e., $\hat{X}$ and X are τ_{T^*M}-related). Moreover, since each $\hat{\phi}_s$ satisfies (37), we also derive that $L_{\hat{X}}\theta_0 = 0$ or, equivalently,

$$i_{\hat{X}}\omega_0 = -d\langle \hat{X}, \theta_0 \rangle.$$

Hence, each vector field $X \in \mathscr{X}(T^*M)$ lifts to a Hamiltonian vector field $\hat{X}$ on TT^*M with Hamiltonian $F_X = \langle \hat{X}, \theta_0 \rangle$. In local coordinates, representing X by

$$X = \xi^i(q,p)\frac{\partial}{\partial q^i} + \eta_i(q,p)\frac{\partial}{\partial p_i}$$

and using the expression (12) for θ_0, we get

$$F_X = (\dot{p}_i - Q_i)\xi^i - \dot{q}^i\eta_i.$$

As a last property, we mention that for any two vector fields X and Y on T^*M, one has $[X,Y]\hat{} = [\hat{X}, \hat{Y}]$.

ACKNOWLEDGMENTS

We would like to thank Professor R. Mertens and Dr. W. Sarlet for reading the manuscript and for their interest in our work.

[1] W. M. Tulczyjew, "Les sous-variétés lagrangiennes et la dynamique hamiltonienne," C. R. Acad. Sci. Paris Ser. A **283**, 15–18 (1976); W. M. Tulczyjew, "Les sous-variétés lagrangiennes et la dynamique lagrangienne," C. R. Acad. Sci. Paris Ser. A **283**, 675–678 (1976); W. M. Tulczyjew, "A symplectic formulation of particle dynamics," in *Differential Geometrical Methods in Mathematical Physics, Lecture Notes in Mathematics*, Vol. 570 (Springer-Verlag, Berlin, 1977), pp. 457–463.

[2] W. M. Tulczyjew, "A symplectic formulation of field dynamics," in *Differential Geometrical Methods in Mathematical Physics, Lecture Notes in Mathematics*, Vol. 570 (Springer-Verlag, Berlin, 1977), pp. 464–468.

[3] R. Abraham and J. E. Marsden, *Foundations of Mechanics*, 2nd ed. (Benjamin/Cummings, Reading, MA, 1978).

[4] W. M. Tulczyjew, "The Legendre transformation," Ann. Inst. H. Poincaré Sect. A **27**, 101–114 (1977).

[5] G. Hamel, *Theoretische Mechanik*, Grundlehren. Math. Wiss., Vol. 57 (Springer-Verlag, Berlin, 1967); L. A. Pars, A *Treatise on Analytical Dynamics* (Heinemann, London, 1965).

[6] A. Weinstein, *Lectures on Symplectic Manifolds*, CBMS Regional Conf. Ser. in Math., Vol. 29 (Am. Math. Soc., Providence, RI, 1977).

[7] The inner product of a vector field X and a differential form α will be denoted by $i_X\alpha$, whereas in case α is a 1-form, we also use the notation $\langle X, \alpha \rangle$.

[8] J. Sniatycki and W. M. Tulczyjew, "Generating Forms of Lagrangian Submanifolds," Indiana Univ. Math. J. **22**, 267–275 (1972).

[9] C. Godbillon, *Géométrie Différentielle et Mécanique Analytique* (Hermann, Paris, 1969).

[10] E. L. Stiefel and G. Scheifele, *Linear and Regular Celestial Mechanics*, Grundlehren. Math. Wiss., Vol. 174 (Springer-Verlag, Berlin, 1971); R. W. Weber, *Kanonische Theorie Nichtholonomer Systeme*, Diss. ETH, Vol. 6876 (Peter Lang, Bern, 1981).

[11] By definition of θ_P, one has for any $\alpha \in T^*P$ and $x \in T_\alpha(T^*P)$: $\langle x, \theta_P(\alpha) \rangle = \langle T\pi_P(x), \alpha \rangle = \langle T\pi_P(x), \tau_{T^*P}(x) \rangle$. θ_P can also be characterized by the property that for any $v \in \Omega^1(P)$, $v^*\theta_P = v$. (See, e.g., Ref. 3, pp. 178–179.)

[12] The Whitney sum $TM \times_M T^*M$ (also denoted by $TM \oplus T^*M$) is a vector bundle over M, which consists of the pairs $(v, \alpha) \in TM \times T^*M$ for which $\tau_M(v) = \pi_M(\alpha)$. For a precise definition of the Whitney sum of two vector bundles, see, e.g., Ref. 9, p. 54.

The forced Toda lattice: An example of an almost integrable system

D. J. Kaup
Institute for Nonlinear Studies, Clarkson College of Technology, Potsdam, New York 13676

(Received 26 July 1983; accepted for publication 9 September 1983)

A method for solving forced integrable systems is presented. The method requires the knowledge of at least one piece of information about the solution. Once this is known, one may then construct the remainder of the solution. In this sense these systems are "almost integrable." The forced semi-infinite Toda lattice is used as an example and to illustrate the method.

PACS numbers: 03.40.Kf

I. INTRODUCTION

Although the inverse scattering transform[1] (IST) is well established as a method for solving free integrable systems, little work has been done on forced integrable systems. By "free" we mean those systems without some type of forcing term. Typical examples of free integrable systems would be the sine–Gordon equation[2]

$$\phi_{tt} - \phi_{xx} + \sin\phi = 0, \tag{1}$$

where the boundary conditions are $\phi(x \to \pm\infty, t) = 2\pi n$, or the nonlinear Schrödinger equation

$$i\psi_t = \psi_{xx} \pm 2(\psi^*\psi)\psi, \tag{2}$$

with the boundary conditions of $\psi(x \to \pm\infty, t) = \alpha$, where α is an arbitrary complex constant.[3,4] On the other hand, a "forced" system would have some forcing terms which determine much of the motion. As an example of a forced integrable system, the driven sine–Gordon chain is where Eq. (1) is valid for $x > 0$, while the value of $\phi(0,t)$ is externally controlled. If one drove this system such that $\phi(0,t) = 2\pi t$, then for every one unit of time a new kink would have been injected into this sine–Gordon chain. Other examples are easily imagined.

One will note that the above-mentioned "free integrable" systems are all completely solved by the IST. And this method of solution is well known. But in general the "forced integrable" systems are not solvable, except in special cases wherein one may utilize some symmetry.[5] Otherwise most of what we know of such forced systems has been obtained by numerical methods.

If one reflects on what happens to the scattering data in a forced integrable system, one can appreciate some of the complexity of such systems. For example, in the above-mentioned driven sine–Gordon chain, the scattering data must vary as some complicated function of time, simply because in every new unit of time, an additional kink must appear, which means that a new pole in the reflection coefficient has to move across the real axis up into the upper half of the complex ζ-plane (ζ is the eigenvalue of the scattering problem). On the other hand, the time dependence for free integrable systems is quite simple. The bound-state eigenvalues are fixed in time as is also the magnitude of the reflection coefficient. Another feature of these forced integrable systems is that the Lax pair relation[6]

$$L_t = [M,L], \tag{3}$$

which for the free system is satisfied everywhere is now satisfied "almost everywhere" instead of "everywhere". Equation (3) is violated at those points where the system is being forced. It is this Lax relation which guaranteed the integrability of the free system in the first place. So if for the forced system Eq. (3) is now satisfied only almost everywhere, could we then not expect such systems to be something like "almost integrable"? Indeed, such is the case. As I shall demonstrate, given the forcing terms and only a few additional pieces of information about the system, the system then becomes completely integrable. This additional information is not independent of the forcing terms and is quite dependent on them. So there is a consistency problem. But once this additional information is obtained or known, then the remainder of the system does become completely integrable.

The remainder of the paper will be devoted to using the forced Toda lattice as an example of an almost integrable system and to illustrate these above ideas. By "forced Toda lattice" I mean the semi-infinite Toda lattice[7]

$$\dot{Q}_n = P_n, \tag{4a}$$

$$\dot{P}_n = -\exp(Q_n - Q_{n+1}) + \exp(Q_{n-1} - Q_n), \quad (n \geqslant 1), \tag{4b}$$

and where Q_0 (and $P_0 = \dot{Q}_0$) are externally controlled. In other words, $Q_0(t)$ determines how the zeroth lattice particle will move and then the motion of all other particles to the right of this particle is determined by Eq. (4). This system was suggested to me by Professor Knopoff,[8] who along with T. G. Hill[9] had observed a fascinatingly regular envelope structure developing out of an apparently chaotic system. (See their Fig. 2.) An example of the same is shown in my Fig. 1, but at a different time. What one should note is the regular envelope structure to the left, whereas as one moves to the right the structure becomes more and more random and chaotic. To say the least, this is a very curious and strange behavior, and one would like to be able to understand what is happening here. In this case, the forcing of the zeroth particle is a very simple uniform forward motion $Q_0(t) = -2b_1 t$, where b_1 is some negative constant. Thus the zeroth particle is being rammed into the other particles, creating a shock wave. The strange behavior is the subsequent creation of a regular envelope from out of this chaotic shock wave.

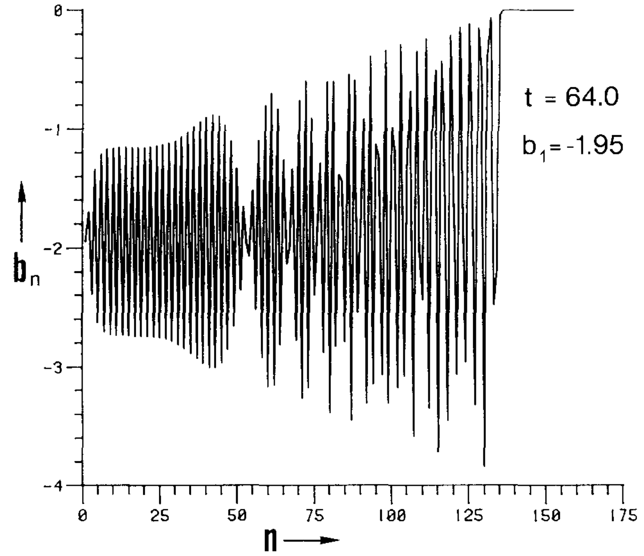

FIG. 1. Plot of b_n vs t in the forced Toda lattice for $b_1 = -1.95$ at $t = 64.0$ where $-2b_{n+1}$ is the velocity of the nth particle. Note the regular envelope structure to the right.

The study of shock waves in one-dimensional lattices is not new. An earlier analysis by Holian and Straub[10] centered on the relaxation toward thermodynamic equilibrium in the wake of shocks. Included in their numerical analysis was the Toda lattice. These numerical results for the Toda lattice have recently been intensively analyzed[11] by using local IST techniques. (By "local" IST techniques it is meant that one takes a small section of the system and analyzes it with the IST, determining what solitons are present inside this section, etc. Of course the section must be sufficiently wide so that an analysis does make sense.) This is in contrast to what I shall do here which would best be described as a "global" analysis. Thus my analysis is a compliment to theirs, and many of our results are of course the same. Mainly we differ in emphasis. Holian, Flaschka, and McLaughlin[11] sought to explain the molecular-dynamics experiments. I am seeking a more general method for determining the time evolution of the scattering data when an integrable system is being forced. Only the model and the specific results are the same. The techniques developed by each of us are different.

Next I shall briefly summarize the IST for the semi-infinite Toda lattice in Sec. II. Then in Sec. III I shall determine the time dependence of the scattering data for the forced Toda lattice. This will not be a solution of the initial-value problem since this solution will require a part of the solution before one can construct the problem. So there will be a consistency problem.

Nevertheless this solution is still useful, and in Sec. IV I shall discuss how one may use it to predict the scattering data for all time. I shall then conclude with some concluding remarks on the consistency problem.

II. THE IST FOR THE FORCED TODA LATTICE

Following Flaschka,[12] we define a_n and b_n by

$$a_{n+1} = \tfrac{1}{2}\exp[-\tfrac{1}{2}(Q_n - Q_{n-1})], \quad (n \geqslant 1), \tag{5a}$$

$$b_n = -\tfrac{1}{2}P_{n-1}; \tag{5b}$$

then from Eq. (4) it follows that

$$\dot{a}_n = a_n(b_n - b_{n-1}), \quad (n \geqslant 2), \tag{6a}$$

$$\dot{b}_n = 2(a_{n+1}^2 - a_n^2), \tag{6b}$$

where $b_1(t)$ and $Q_0(t)$ are to be specified. Equation (6) then determines a_n and b_n for $n \geqslant 2$.

Consider now the eigenvalue problem[12]

$$a_{n+1}V_{n+1} + a_nV_{n-1} + (b_n - \lambda)V_n = 0 \quad (n \geqslant 1), \tag{7}$$

where λ is the eigenvalue and we shall take $a_1 = \frac{1}{2}$ (see Ref. 13). As shown by Case[13] one may define the scattering data in the semi-infinite discrete case as follows. (I shall shift to the AKNS notation, where ψ_n are the right eigenstates and ϕ_n are the left eigenstates.) Take

$$\lambda = \tfrac{1}{2}(z + 1/z) \tag{8}$$

and assume that $a_n - \frac{1}{2}$ and b_n each approach zero sufficiently rapidly that the following results hold. Then the right eigenstate may be defined by

$$\psi_n(z) \rightarrow z^n \quad \text{as } n \rightarrow +\infty, \tag{9}$$

where $\psi_n z^{-n}$ is analytic inside the unit circle of the z-plane. I define

$$\bar{\psi}_n(z) \equiv \psi_n(1/z), \tag{10}$$

which is the second independent right eigenstate of (7).

Now define a left eigenstate by

$$\phi_n \equiv (z - 1/z)^{-1}[\bar{\psi}_0(z)\psi_n(z) - \psi_0(z)\bar{\psi}_n(z)]. \tag{11}$$

By construction,

$$\phi_0 = 0, \tag{12a}$$

$$\phi_1 = 1. \tag{12b}$$

Consider using Eqs. (7) and (12) to construct the solution ϕ_n. Clearly ϕ_n will be at most a polynomial in λ, of order $n-1$. Thus it follows that ϕ_n is analytic in λ except for a finite-order pole at $\lambda = \infty$.

Define[13]

$$S(z) = e^{2i\delta(z)} = \bar{\psi}_0(z)/\psi_0(z), \tag{13}$$

where δ is the phase shift. Then the scattering data consists of the values of $\delta(z)$ for z on the unit circle (the continuous spectrum) and the poles of $S(z)$ inside the unit circle (the bound-state spectrum). These poles are the zeros of $\psi_0(z)$ inside the unit circle. The bound-state part of the spectrum is specified by the value of z at the pole (z_i) and value of the normalization constant M_i^2, which is the negative of the residue of $z^{-1}S(z)$ at the pole. The constant M_i is real, whence $M_i^2 \geqslant 0$.

The inverse scattering equations are obtained by considering the contour integral

$$\oint \frac{dz}{2\pi i}\frac{\phi_n(\lambda)}{\psi_0(z)}(z - 1/z)z^{m-1}, \tag{14}$$

where C is an infinitestimal circular contour CCW around the origin. From this and upon expanding ψ_n as

$$\psi_n(z) = K_n \sum_{j=n} \kappa_{nj} z^j, \tag{15}$$

where $\kappa_{nn} = 1$, one obtains the following.[13,12] First construct

$$F_j = \frac{1}{2\pi i}\oint \frac{dz}{z}\,[1 - S(z)]z^j, \tag{16}$$

then for $m > n \geqslant 1$, one has

$$\kappa_{nm} + F_{n+m} + \sum_{j=n+1}^{\infty} \kappa_{nj}F_{j+m} = 0 \tag{17}$$

from which one may solve for $\kappa_{n,j}$. Next construct K_n from

$$(K_n)^{-2} = 1 + F_{2n} + \sum_{j=n+1}^{\infty} \kappa_{nj}F_{j+n}. \tag{18}$$

Then a_n and b_n may be recovered from

$$a_n = \tfrac{1}{2}K_n/K_{n-1}, \tag{19a}$$

$$b_n = \tfrac{1}{2}(\kappa_{n,n+1} - \kappa_{n-1,n}). \tag{19b}$$

From these equations one may construct the direct and inverse scattering transform for the forced Toda lattice. Given (a_n, b_n) for $n \geqslant 2$, by Eqs. (7)–(13) one may map these quantities into the scattering data. And given the scattering data, from Eqs. (16)–(19) one may construct the inverse scattering transform which allows one to reconstruct the potentials (a_n, b_n) for $n \geqslant 2$. Clearly we may do either of these at any time. Now the question is, if (a_n, b_n) for $n \geqslant 2$ evolves according to Eq. (6), how will the scattering data evolve? This we shall answer next.

III. THE TIME DEPENDENCE OF THE SCATTERING DATA

In the absence of forcing and when one has an infinite lattice, Flaschka[12] found that the time evolution of the eigenstates of Eq. (7) was given by

$$\dot{V}_n = a_{n+1}V_{n+1} - a_nV_{n-1} + CV_n, \tag{20}$$

where C is an arbitrary constant. In the infinite case, the integrability condition for (7) and (20) is the infinite Toda lattice [Eq. (6) valid for all n]. But in the semi-infinite case, although we expect Eq. (20) to be valid for large n, one must carefully account for the equations near $n = 0$ since Eq. (6) is only valid for $n \geqslant 2$. Equation (6) just cannot be true for $n = 1$ since a_1 and b_1 are constrained. Carefully accounting for these equations near $n = 0$ shows that for the forced Toda lattice, the equivalent form of (20) is

$$\dot{V}_n = a_{n+1}V_{n+1} - a_nV_{n-1} + CV_n \quad (n \geqslant 2), \tag{21a}$$

$$\dot{V}_1 = (C + \lambda - b_1)V_1 - V_0, \tag{21b}$$

$$\dot{V}_0 = (4a_2^2 - 2\dot{b}_1)V_1 + V_0(b_1 - \lambda + C). \tag{21c}$$

We comment that Eq. (21b) is simply Eq. (21a) for $n = 1$ combined with Eq. (7) for $n = 1$. Equation (21c) follows upon differentiating Eq. (7) with respect to time. One may easily verify that the integrability conditions for Eqs. (7) and (21) are now Eqs. (6).

However, one may not uniquely determine the time evolution of the scattering data from Eq. (21). Note the term a_2^2 present in Eq. (21c). From Eq. (5) we have

$$a_2^2 = \tfrac{1}{4}\exp(Q_1 - Q_0), \tag{22a}$$

$$b_1 = -2\dot{Q}_0. \tag{22b}$$

Although we do know b_1 because $Q_0(t)$ is to be specified, we do not know what a_2^2 will be because $Q_1(t)$ is an unknown.

For the present, let us assume that we do know what a_2^2 is, and continue. To determine the time dependence of the scattering data, $S(z)$, per Eq. (13) we require the time dependence of $\psi_0(z)$. From Eqs. (9) and (21) for n large, I determine that for the eigenstate $\psi_n(z)$, the constant C is

$$C = -\tfrac{1}{2}(z - 1/z). \tag{23}$$

Define the function $\chi(z,t)$ by

$$\psi_1 = \chi e^{+Ct}; \tag{24a}$$

then by (21b),

$$\psi_0 = e^{Ct}[(\lambda - b_1)\chi - \dot{\chi}] \tag{24b}$$

and (21a) gives

$$\ddot{\chi} + \Omega^2(z,t)\chi = 0, \tag{25}$$

where

$$\Omega^2 = 4a_2^2 - \dot{b}_1 - (b_1 - \lambda)^2. \tag{26}$$

Given $\Omega^2(z,t)$ and the initial values of $\chi(z,0)$ and $\dot{\chi}(z,0)$, one may construct the solution for $\chi(z,t)$, and thereby the solution for $\psi_0(z,t)$. From Eq. (13) one may now construct the scattering function $S(z,t)$. However, the value of $4a_2^2(t)$ is required before any of this may be performed. If $4a_2^2(t)$ was known then the remainder of the solution would follow. In this sense, these forced integrable systems are "almost integrable." Some piece of the solution must be provided before the remainder of the solution will follow.

However, if one knows something nontrivial about the properties of $4a_2^2(t)$, then something nontrivial can be said about the scattering data, and thereby something nontrivial about the remainder of the solution. It is in this manner that I shall seek to glean information about this forced system.

IV. THE MOLECULAR-DYNAMICS CASE

Let us now specialize to the molecular dynamics case where one takes

$$Q_0(t) = \begin{cases} 0 & \text{if } t \leqslant 0, \\ -2b_1 t & \text{if } t \geqslant 0, \end{cases} \tag{27}$$

with b_1 as a constant, $-2b_1$ being the velocity of the zeroth particle. For this case the behavior of $4a_2^2$ is quite simple[11] and has two characteristic forms. These are shown in Fig. 2 and Fig. 3. In Fig. 2, I show the characteristic form of $4a_2^2$ for small velocities; in this example $b_1 = -\tfrac{1}{2}$. The main features to note are the initial rise, followed by a decaying ringing, which soon decays to a constant value of approximately 2.25. The value of $b_1 = -1$ is a critical value,[11] and for magnitudes of b_1 larger than this critical value the characteristic form of $4a_2^2$ changes, as one can see in Fig. 3. Here where $b_1 = -2.0$, we see that the ringing does not decay. Instead $4a_2^2$ seems to asymptotically approach an oscillation with an amplitude about 1.0 and with an average value of about 9.0.

In either case, the dominant feature of $4a_2^2$ is that it shifts from 1.0 at $t = 0$ up to some larger asymptotic value, 2.25 for $b_1 = -0.5$ and 9.0 for $b_1 = -2.0$. So as a first approximation one could replace $4a_2^2$ in Eq. (26) by its asymptotic average value and then proceed to solve for χ from Eq. (25). Of course this will not generate the exact solution for the scattering data. But one could expect that it

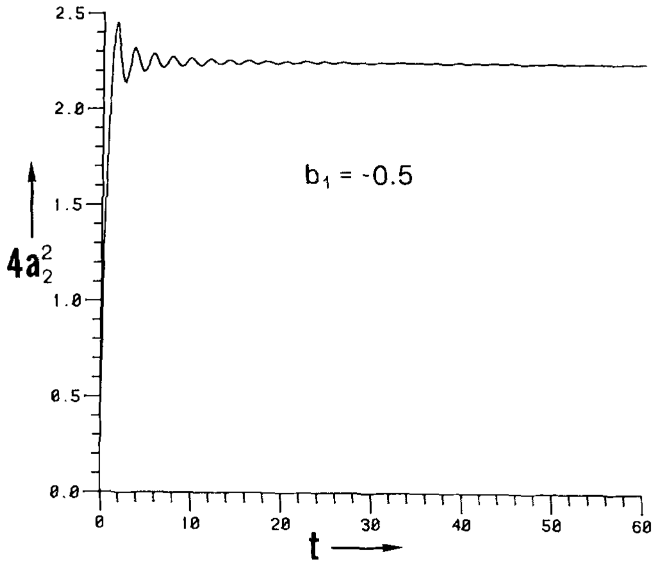

FIG. 2. A plot of $4a_2^2$ vs t when $b_1 = -0.5$ showing the rapid decay of the initial ringing.

would contain the main features of the solution. This is indeed so. We have already determined that this procedure works quite well for predicting the soliton birth rate.[14]

As a final point, I wish to point out that there may be a solution to the consistency problem such that given $b_1(t)$, one may be able to directly determine $4a_2^2$. Let me illustrate this in the molecular-dynamics case, Eq. (27). First, I determine the initial conditions on χ and $\dot{\chi}$. At $t = 0$, we have

$$a_n = \tfrac{1}{2}, \quad b_n = 0 \quad (n \geqslant 2), \tag{28a}$$

$$a_1 = \tfrac{1}{2}, \tag{28b}$$

while b_1 is some nonzero value. Then solving (7) for ψ_n gives

$$\psi_n = z^n \quad (n \geqslant 1), \tag{29a}$$

$$\psi_0 = 1 - 2b_1 z. \tag{29b}$$

So by Eq. (24) we have

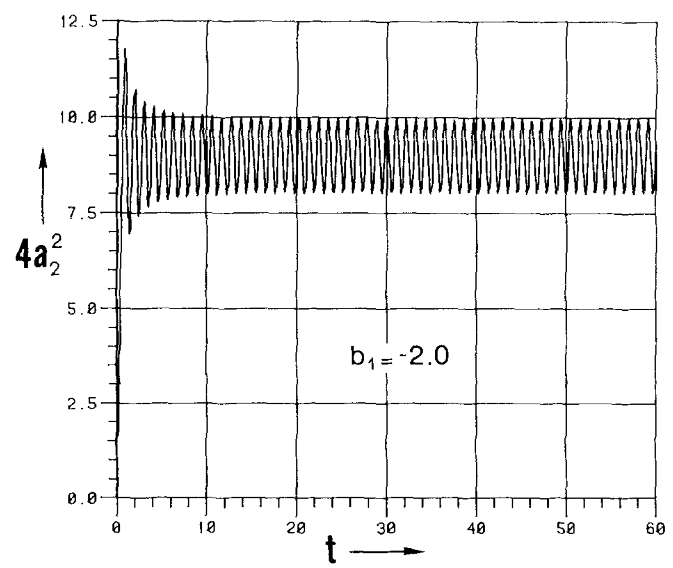

FIG. 3. A plot of $4a_2^2$ vs t when $b_1 = -2.0$ showing the asymptotic oscillations.

$$\chi(t = 0) = z, \tag{30a}$$

$$\dot{\chi}(t = 0) = \tfrac{1}{2}(z^2 - 1) + zb_1. \tag{30b}$$

Since b_1 is a constant, then Eq. (25) may easily be turned into the integral equation

$$\begin{aligned} &2(\lambda - b_1)\chi(t) \\ &= z^2 \exp[(\lambda - b_1)t] + (1 - 2b_1 z)\exp[-(\lambda - b_1)t] \\ &\quad + 2\int_0^t dt'\chi(t')4a_2^2(t')\sinh[(\lambda - b_1)(t' - t)], \end{aligned} \tag{31}$$

where $4a_2^2$ only appears in the kernel.

Now consider the analytical properties of this solution as $|z| \to 0$. In general we would expect essential singularities at $z = 0$ due to the presence of terms like $e^{\pm \lambda t}$. But now consider (24a). We have

$$(1/z)\psi_1 = (1/z)\chi e^{Ct}, \tag{32}$$

where $(1/z)\psi_1$ is known to be analytic inside the unit circle.[13] For arbitrary values of $4a_2^2$ in (31), such will not be so on the right-hand side. One may easily verify this by using a Taylor series expansion about $t = 0$. One would also note that (32) would have the correct analytical properties only if $4a_2^2$ satisfies the equations of motion, Eqs. (6), for the proper value of b_1. (I have only checked this out to second order, but from its form, it seems reasonable that it will be true to all orders.)

This leads us to conjecture that by demanding $z^{-1}\chi e^{Ct}$ to be analytic inside the unit circle, the correct solution for $4a_2^2(t)$ may be determined and obtained without having to solve the equations of motion. Given $b_1(t)$, Eqs. (25) and (26) show that $4a_2^2(t)$ is a potential for χ, while χ satisfies a Schrödinger-like equation on the semi-infinite interval $t \geqslant 0$. Clearly, $4a_2^2$ could be mapped into the scattering data for the problem given by Eq. (25). But whether or not the required analytical properties of χ in Eq. (32) are sufficient to obtain this scattering data remains to be seen.

ACKNOWLEDGMENTS

The author gratefully acknowledges the assistance of Mr. Dan Neuberger in performing the numerical calculations and preparing the figures. He also wishes to thank Professor Knopoff and Professor Busse for their kind hospitality during the author's stay at the Institute of Geophysics and Planetary Physics, UCLA. This research has been supported in part by the Institute of Geophysics and Planetary Physics (UCLA) and the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant or Cooperative Agreement Number AFOSR-82-0154. This material is based upon work supported by the National Science Foundation under Grant No. MCS-8202117. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation.

[1] M. Ablowitz, D. Kaup, A. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).

[2] M. Ablowitz, D. Kaup, A. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1262 (1973).

[3]V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].
[4]V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **64**, 1627 (1973) [Sov. Phys. JETP **37**, 823 (1973)].
[5]M. J. Ablowitz and H. Segur, J. Math. Phys. **16**, 1054 (1975).
[6]P. Lax, Comm. Pure Appl. Math. **21**, 467 (1968).
[7]M. Toda, Prog. Theor. Phys. Suppl. **45**, 174 (1970).
[8]L. Knopoff (private communication).
[9]T. G. Hill and L. J. Knopoff, Geophys. Res. Pap. **85**, 7025 (1980).
[10]B. L. Holian and G. K. Straub, Phys. Rev. B **18**, 1593 (1978).
[11]B. L. Holian, H. Flaschka, and D. W. McLaughlin, Phys. Rev. A **24**, 2595 (1981).
[12]H. Flaschka, Prog. Theor. Phys. **51**, 703 (1974).
[13]K. M. Case, J. Math. Phys. **14**, 916 (1973).
[14]D. J. Kaup and D. H. Neuberger, J. Math. Phys. **25**, 282 (1984).

The soliton birth rate in the forced Toda lattice

D. J. Kaup and D. H. Neuberger
Institute for Nonlinear Studies, Clarkson College of Technology, Potsdam, New York 13676

(Received 26 July 1983; accepted for publication 9 September 1983)

The soliton birth rate in the semi-infinite Toda lattice is studied. The lattice is forced by driving the zeroth particle with a constant velocity into the remainder of the lattice. An approximate solution for the soliton birth rate is derived and it is shown to compare quite favorably with the actual birth rate.

PACS numbers: 03.40.Kf

In a recent paper[1] one of the authors (DJK) discussed and demonstrated how one could solve "almost integrable" systems, one example of which is the forced Toda lattice. This is the semi-infinite Toda lattice[2] where the equation of motion is

$$\dot{Q}_n = \exp(Q_n - Q_{n-1}) - \exp(Q_{n+1} - Q_n) \tag{1}$$

for $n \geqslant 1$. The position of the zeroth particle $Q_0(t)$ is assumed to be driven by some external agent. And the motion of this particle then drives all other particles through Eq. (1). A simple example is a case from molecular dynamics[3] where one starts with a static lattice; then at $t = 0$ one forces the zeroth particle to ram into the remainder of the lattice by imposing upon it a uniform forward velocity. Thus $Q_0 = v_0 t$, where v_0 is a constant.

As this zeroth particle rams into the remainder of the lattice, a shock wave is created, the front part of which consists of a collection of solitons, all with approximately the same velocity. Parts of this shock wave have been analyzed[3] using "local IST" techniques to verify that solitons are present with approximately the same velocities.

With the recently developed method for handling almost integrable systems,[1] it now becomes possible to accurately predict what the soliton structure and spectrum of this shock wave is. The purpose of this paper is to predict this spectrum and to compare the predicted soliton spectrum with the actual observed spectrum. As we shall see, the agreement between the predictions and the numerical results is quite good indeed.

Next we shall summarize those equations and results from Ref. 1 which are applicable to the motion of the soliton spectrum. The solution of these equations requires one to know beforehand what will be the separation between the first two particles as a function of time. We approximate this in a reasonable manner and obtain thereby an approximate solution for the motion of the bound-state (soliton) spectrum. We next numerically compute the lattice motion from Eq. (1), determine what the actual spectrum is at various times, and then compare results.

According to Kaup,[1] the inverse scattering transform (IST) for the forced Toda lattice requires the solution of the eigenvalue problem

$$a_{n+1}\,\psi_{n+1} + a_n\,\psi_{n-1} + (b_n - \lambda)\,\psi_n = 0, \tag{2}$$

where

$$\lambda = \tfrac{1}{2}(z + 1/z), \tag{3}$$

and ψ_n is the eigensolution where

$$\psi_n \rightarrow z^n \quad \text{as} \quad n \rightarrow +\infty. \tag{4}$$

The quantities a_n and b_n in Eq. (2) are related to Q_n by

$$a_{n+1} = \tfrac{1}{2}\exp[-\tfrac{1}{2}(Q_n - Q_{n-1})], \tag{5a}$$

$$b_n = -\tfrac{1}{2}\dot{Q}_{n-1}, \tag{5b}$$

and thus as $n \rightarrow +\infty$,

$$a_n \rightarrow \tfrac{1}{2}, \tag{6a}$$

$$b_n \rightarrow 0. \tag{6b}$$

Note that b_1 is just the negative of one half of the velocity of the driven zeroth particle. Also a_1 cannot be defined by (5a) since the $n = -1$ particle does not exist. Instead we may define it to be $\frac{1}{2}$, as was shown by Case.[4] The bound-state eigenvalues are those values of z where $\psi_0(z)$ is zero.[4,5] These only occur when z is real and is between -1 and $+1$.

As shown by Kaup,[1] if one defines the function χ by

$$\chi = \psi_1 e^{-Ct}, \tag{7}$$

it then follows that

$$\dot{\chi} = (\lambda - b_1)\,\chi - \psi_0 e^{-Ct}, \tag{8}$$

and that χ will satisfy

$$\ddot{\chi} + \Omega^2 \chi = 0, \tag{9}$$

where

$$\Omega^2(z,t) = 4a_2^2(t) - (b_1 - \lambda)^2 - \dot{b}_1 \tag{10}$$

and

$$C = -\tfrac{1}{2}(z - 1/z). \tag{11}$$

Thus if one possessed the function $\chi(z,t)$, from (8) one could construct $\psi_0(z,t)$ thereby obtaining the soliton spectrum (the zeros of ψ_0) as a function of time. However, before we may construct the solution for χ, we must know $4a_2^2$, which by (5a) is

$$4a_2^2 = \exp(Q_0 - Q_1). \tag{12}$$

Although Q_0 is given, Q_1 is not and requires the solution of the problem which we are trying to solve. For the moment we shall simply assume that $4a_2^2$ is known, and continue.

Assuming $4a_2^2(t)$ to be known, then we may solve Eq. (9) as follows. Take a solution of (9) to be of the form

 0022-2488/84/020282-03$02.50

$$\chi = Ae^{\pm i\phi}; \tag{13}$$

then Eq. (9) gives

$$A = \text{const}/(\dot{\phi})^{1/2}, \tag{14}$$

$$\dot{\phi}^2 = \Omega^2 + \mu^2 + \dot{\mu}, \tag{15}$$

where

$$\mu = \dot{A}/A. \tag{16}$$

The initial values for χ and $\dot{\chi}$ follow from the initial values for a_n and b_n as follows. Consider the moment just after $t = 0$ where b_1 has reached its nonzero uniform value. Here

$$a_n = \tfrac{1}{2} \quad (n \geqslant 1), \tag{17a}$$

$$b_n = 0 \quad (n \geqslant 2). \tag{17b}$$

We may now solve (2) for the initial value of ψ_n. We find

$$\psi_n(t=0) = z^n \quad (n \geqslant 1), \tag{18a}$$

$$\psi_0(t=0) = 1 - 2b_1 z, \tag{18b}$$

which by (7) and (8) give the initial values

$$\chi(t=0) = z, \tag{19a}$$

$$\dot{\chi}(t=0) = \lambda z - 1 + b_1 z. \tag{19b}$$

Matching the two possible solutions in (13) to these initial conditions, we determine the correct solution of χ to be

$$\chi = \frac{A}{A_0}\left[z\cos\phi + \frac{\dot{\chi}_0 - \mu_0 z}{\dot{\phi}_0}\sin\phi\right], \tag{20}$$

where the subscripts "0" refer to initial values and we have taken

$$\phi_0 = 0. \tag{21}$$

So far no approximations have been made. From (8), the zeros of ψ_0 will be where

$$\tan\phi = \frac{\dot{\phi}_0(\lambda - b_1 - \mu) + \dot{\phi}(1/z - \lambda - b_1 + \mu_0)}{(\lambda - b_1 - \mu)(1/z - \lambda - b_1 + \mu_0) - \dot{\phi}_0\dot{\phi}}. \tag{22}$$

Define

$$\Delta_0 = \arctan[\dot{\phi}_0/(1/z - \lambda - b_1 + \mu_0)] \tag{23}$$

with which Eq. (22) can be reduced to

$$\phi = \Delta_0 + \arctan[\dot{\phi}/(\lambda - b_1 - \mu)]. \tag{24}$$

Now, let us approximate in the spirit of the WKB method to determine $\dot{\phi}$ and $\dot{\phi}_0$. From (1), (5), (10), (14), and (16) one has that the initial value of μ is

$$\mu_0 = b_1/2\dot{\phi}_0^2. \tag{25}$$

Provided $\dot{\phi}_0$ was not close to zero, the solution of (15) would be $\dot{\phi} = \pm\Omega$. However, if Ω^2 would be close to zero we would have to account for the terms $\mu^2 + \dot{\mu}$. We do this by evaluating them for $\dot{\phi}_0$ small. Otherwise they would have no significant effect and could be ignored. For small $\dot{\phi}_0$, we have

$$\dot{\mu}_0 \simeq b_1^2/\dot{\phi}_0^4 = 4\mu_0^2 \tag{26}$$

so we approximate Eq. (15) initially by

$$\dot{\phi}_0^2 = 1 - (\lambda - b_1)^2 + \tfrac{5}{4}b_1^2/\dot{\phi}_0^4, \tag{27}$$

which is a cubic equation for $\dot{\phi}_0^2$. It has only one positive real root when λ and b_1 are real.

For the later times, we shall simply ignore the effects of μ and $\dot{\mu}$. Thus we take them to be zero in (15) and (24). It only remains to specify the values of $4a_2^2(t)$. To see how best to do this consider $4a_2^2$ vs t as shown in Fig. 1, which is when $b_1 = -0.5$, and Fig. 2 which is when $b_1 = -2.0$. What we observe there is that $4a_2^2$ rapidly shifts from its value of 1.0 at $t = 0$ to a larger average value. Clearly the most dominant feature is this definite shift in the average value. So we shall approximate the value of $4a_2^2$ required in the calculation of $\dot{\phi}$, Eq. (15), by its average value. Thus

$$\dot{\phi}^2 \simeq \langle 4a_2^2 \rangle - (b_1 - \lambda)^2. \tag{28}$$

From Figs. 1 and 2, we have

$$b_1 = -0.5, \quad \langle 4a_2^2 \rangle \simeq 2.25, \tag{29a}$$

$$b_1 = -2.0, \quad \langle 4a_2^2 \rangle \simeq 9.0, \tag{29b}$$

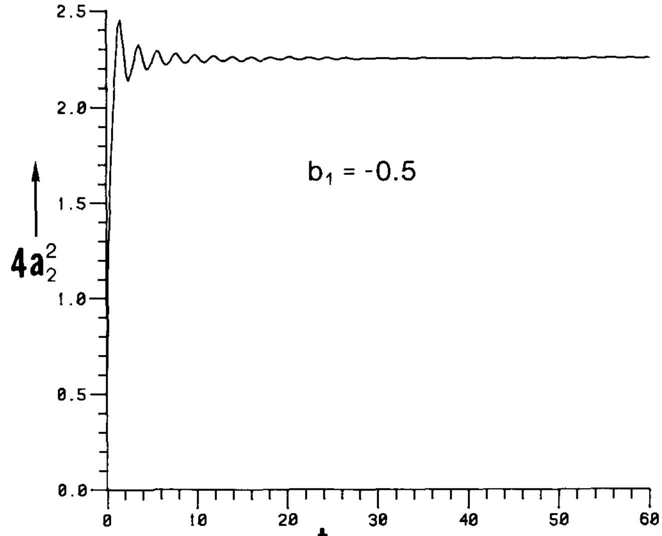

FIG. 1. A plot of $4a_2^2$ vs t when $b_1 = -0.5$ showing the rapid decay of the initial ringing.

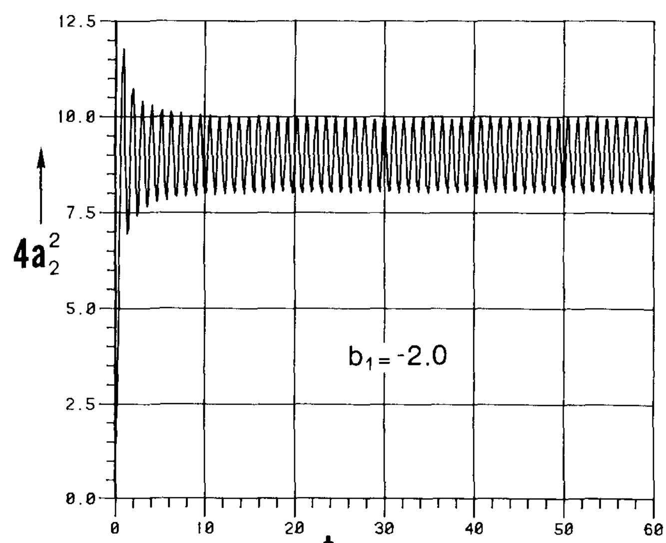

FIG. 2. A plot of $4a_2^2$ vs t when $b_1 = -2.0$ showing the asymptotic oscillations.

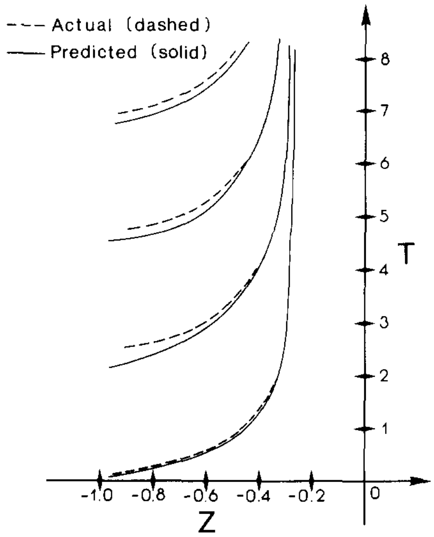

FIG. 3. The soliton birth rate when $b_1 = -0.5$ as predicted by Eq. (31) (solid line) and as actually is (dashed line).

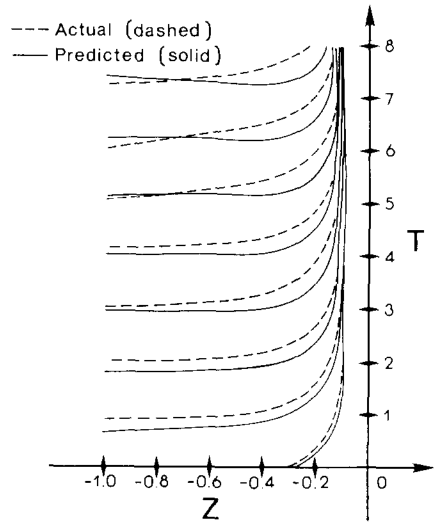

FIG. 4. The soliton birth rate when $b_1 = -2.0$ as predicted by Eq. (31) (solid line) and as actually is (dashed line).

which are the only values that we shall consider here.

Now

$$\phi = \dot{\phi} t, \tag{30}$$

and Eq. (24) gives

$$t = (\dot{\phi})^{-1}[\Delta_0 + \arctan(\dot{\phi}/(\lambda - b_1))]. \tag{31}$$

This equation gives all possible times associated with a given possible value for a bound-state eigenvalue.

A plot of these t values vs z is shown by the solid lines in Fig. 3 for $b_1 = -0.5$ and in Fig. 4 for $b_1 = -2.0$. In Fig. 3, these curves are easily interpreted as being the motion of the eigenvalues of individual solitons. The first soliton is created at $t \sim 0.1$ with an eigenvalue of just above $z = -1$. (A soliton with $z = -1$ would have a zero velocity, zero amplitude, and an infinite width. When z is just greater than -1, then these values become finite and nonzero.) This eigenvalue moves rapidly toward the limiting value of -0.29 at which all bound-state eigenvalues eventually tend to collect, as seen in Fig. 3. The motion in Fig. 4 is quite similar, except that the solitons are created at a faster rate, and the first soliton already exists at $t = 0$. The limiting value is now -0.10, which means faster and narrower solitons as one would expect.

To see how good these predictions are, let us compare this with the actual soliton spectrum. To determine this, we shall numerically integrate Eq. (1) up to some time t. At this time, we shall calculate the a_n's and the b_n's as given by Eq. (5), then solve Eq. (2) numerically for $\psi_0(z)$, plotting $\psi_0(z)$ vs z from $z = -1$ to $z = +1$. One may then easily pick out the zeros of $\psi_0(z)$ which are the bound-state eigenvalues.

The result of this are the dashed lines in Figs. 3 and 4. As seen in Fig. 3, the agreement is quite good, the only difference being slight phase shift in the initial birth times. Otherwise the eigenvalue motion is quite accurately predicted by Eq. (24). Figure 4 does not show as good an agreement, although the general shape and motion still quite accurately reflect the actual curves. This discrepancy may arise in part from ignoring the oscillations in $4a_2^2$ (see Fig. 2) which do not seem to be decaying away. They do rapidly decay away in Fig. 1, and for that value of b_1, the results shown in Fig. 3 gave excellent results.

In conclusion we have demonstrated that one can solve for the soliton spectrum, and its subsequent motion, when an integrable system is driven by forcing terms. The method does require having some particular information about the solution, so it is not a method for solving the initial-value problem. However, the information required for finding the soliton spectrum need not be detailed, and we found average values to be adequate to reproduce at least the gross features of the curves.

ACKNOWLEDGMENTS

This research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant or Cooperative Agreement Number AFOSR-82-0154. This material is based upon work supported by the National Science Foundation under Grant No. MCS-8202117. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[1]D. J. Kaup, J. Math. Phys. **25**, 277 (1984).

[2]M. Toda, Prog. Theor. Phys. Suppl. **45**, 174 (1970).

[3]B. L. Holian, H. Flaschka, and D. W. McLaughlin, Phys. Rev. A **24**, 2595 (1981).

[4]K. M. Case, J. Math. Phys. **14**, 916 (1973).

[5]H. Flaschka, Prog. Theor. Phys. **51**, 703 (1974).

Derivation and application of extended parabolic wave theories. I. The factorized Helmholtz equation

Louis Fishman and John J. McCoy
The Catholic University of America, Washington, D.C. 20064

(Received 1 June 1982; accepted for publication 24 June 1983)

The reduced scalar Helmholtz equation for a transversely inhomogeneous half-space supplemented with an outgoing radiation condition and an appropriate boundary condition on the initial-value plane defines a direct acoustic propagation model. This elliptic formulation admits a factorization and is subsequently equivalent to a first-order Weyl pseudodifferential equation which is recognized as an extended parabolic propagation model. Perturbation treatments of the appropriate Weyl composition equation result in a systematic development of approximate wave theories which extend the narrow-angle, weak-inhomogeneity, and weak-gradient ordinary parabolic (Schrödinger) approximation. The analysis further provides for the formulation and exact solution of a multidimensional nonlinear inverse problem appropriate for ocean acoustic and seismic studies. The wave theories foreshadow computational algorithms, the inclusion of range-dependent effects, and the extension to (1) the vector formulation appropriate for elastic media and (2) the bilinear formulation appropriate for acoustic field coherence.

PACS numbers: 03.40.Kf, 02.30.Jr, 92.10.Vz, 43.25. – x

I. INTRODUCTION

A significant advance in wave propagation modeling in recent years has been the introduction and the widespread application of the "parabolic approximation."[1] Since the original formulation approximately 35 years ago by Fock and Leontovich,[2,3] parabolic approximations have found application in studies of electromagnetic,[4–12] seismic,[13–18] and acoustic[1,19–37] propagation processes. Its direct application as a basis of computational algorithms has been most successfully accomplished in conjunction with the split-step FFT algorithm of Tappert and Hardin.[38]

Within the framework of the reduced scalar Helmholtz governing wave equation, the parabolic approximation can be termed a forward-scattering approximation. It is applicable to a medium characterized by slowly varying (on a wavelength scale) material inhomogeneities with respect to a distinguished global principal propagation, or range, direction and valid for small propagation angles with respect to this established horizontal. Furthermore, the medium must be weakly inhomogeneous with slowly varying material inhomogeneities with respect to the appropriate perpendicular, or cross-range, directions. The parabolic approximation is generally distinct from (1) geometric acoustics (optics), a small wavelength theory which neglects diffraction effects, and (2) separation of variables methods which adopt a picture of horizontal stratification with the subsequent neglect of waveguide mode coupling. The parabolic is a full-wave approximation which retains both diffraction effects associated with a given geometry and coupling between waveguide modes.

It is expected that in some experimental situations the concept of a distinguished principal propagation direction remains valid while the weak-inhomogeneity, weak-gradient, and narrow-angle limitations of the parabolic approximation will be strained and often exceeded. Thus, for example, while the parabolic approximation is ideally suited for describing the propagation of sound in the water column, in a range and depth-dependent ocean, a realistic consideration of the sea-bottom interaction problem raises several troubling points. These involve, but are not necessarily limited to, the sharp discontinuity in the "sound speed" at the interface of the water column and the sea floor and the potentially much larger variations in sound speed in the sea floor itself. An additional complication results from the fact that the sea floor can transmit some energy in a shear mode. Further, with regard to a number of seismic experiments the following two situations are significant: (1) large variations in the sound speed profile occurring over large distances, and (2) significant beam wander, violating the narrow-angle restriction measured relative to a global direction, while still valid measured relative to a local direction.

In discussing the literature treating a failure of the parabolic approximation it is convenient to distinguish between attempts to correct solutions of the parabolic equation, through iteration or asymptotic methods[21,23,25,26] or environmental transformation approaches,[27,28] and attempts to extend a parabolic propagation theory itself.[1,5,6,13,18,20] A third approach is, of course, to return to the full elliptic formulation, the Helmholtz equation, for a direct numerical solution[39] or approximate numerical techniques.[40–42] The approach presented here is to extend the parabolic propagation theory in the spirit of the works of Tappert,[1] Cornes,[5,6,18] Claerbout,[13] and McDaniel.[20]

The extended theories to be developed are parabolic in the sense that the range coordinate can be treated in an incremental manner. This is the crucial characteristic of a parabolic model that makes it so amenable to numerical implementation. The nature of the "extensions" is in the manner in which the cross-range coordinates are treated. Unlike in

 0022-2488/84/020285-12$02.50

the ordinary parabolic wave theory, in which the cross-range dependence enters through a strictly local operator, the cross-range dependence in the extended parabolic theories enters, in general, through a nonlocal operator.

While it might be natural to expect to derive extended theories starting from the ordinary parabolic equation, it seems more likely that such theories must arise from a systematic sequence of equations starting from the Helmholtz and proceeding, with the introduction of a hierarchy of approximations, to the ordinary parabolic. For transversely inhomogeneous environments it is readily established that the governing Helmholtz equation can be exactly factorized into equations that are parabolic in the sense described (see Sec. II and Appendix A). Thus, there is an extended parabolic wave theory that is the exact equivalent to the Helmholtz equation wave theory, as applied to forward propagation in a transversely inhomogeneous half-space. Two of the principal results of this paper, then, are (1) the development of a nonperturbative framework for the explicit construction of the formal square root operator that exactly factorizes the Helmholtz equation and (2) the subsequent systematic derivation of approximate extended parabolic wave theories. The Fourier-type analysis provides for two further developments. The formulation and exact solution of an associated multidimensional nonlinear inverse problem follows, in a complementary fashion, from the direct propagation model. The related construction of a phase space path integral representation for the propagator provides, in conjunction with alternative path integral constructions, both a global and stochastic perspective of the extended parabolic wave theories as well as the basis for computational algorithms. The path integral analysis is presented in Paper II.[43]

II. SUGGESTIVE DERIVATION

To motivate the previous considerations more explicitly, consider the following formal factorization calculation. The Helmholtz equation for a transversely inhomogeneous medium can be written as

$$\{\partial_x^2 + [\bar{k}^2 K^2(\mathbf{x}_\perp) + \nabla_\perp^2]\}\phi(\mathbf{x}) = 0, \tag{2.1}$$

where $\phi(\mathbf{x})$ is the wave field, $k(\mathbf{x}_\perp) = \bar{k}K(\mathbf{x}_\perp)$ is a spatially varying wave number field, x is the principal propagation or range coordinate, and $\nabla_\perp^2$ is the transverse Laplacian associated with the perpendicular, or cross-range, coordinates $\{x_\perp\}$. For acoustics, $\phi(\mathbf{x})$ is a pressure field, $K(\mathbf{x}_\perp) = c_0/c(\mathbf{x}_\perp)$ is a dimensionless sound speed profile or refractive index field, and $\bar{k} = \omega/c_0$ is an appropriate average or reference wavenumber, ω being the signal frequency, $c(\mathbf{x}_\perp)$ the medium sound speed, and c_0 an appropriate average or reference sound speed. The wave field can be expressed as the sum of a forward and backward propagating wave:

$$\phi(\mathbf{x}) = \phi^+(\mathbf{x}) + \phi^-(\mathbf{x}). \tag{2.2}$$

The separate wave fields satisfy uncoupled equations,

$$(i/\bar{k})\partial_x\phi^+(\mathbf{x}) + [K^2(\mathbf{x}_\perp) + (1/\bar{k}^2)\nabla_\perp^2]^{1/2}\phi^+(\mathbf{x}) = 0 \tag{2.3a}$$

and

$$-(i/\bar{k})\partial_x\phi^-(\mathbf{x}) + [K^2(\mathbf{x}_\perp) + (1/\bar{k}^2)\nabla_\perp^2]^{1/2}\phi^-(\mathbf{x}) = 0. \tag{2.3b}$$

That this factorization is exact when there is no range dependence in the index of refraction and the boundary conditions follows from the physical requirement that range variations are necessary to couple forward and backward waves. In the general case of a range-dependent environment, the forward and backward wave propagation is coupled (see Appendix A).

The parabolic approximation is readily recovered. A Taylor series expansion suggests the approximation

$$[K^2(\mathbf{x}_\perp) + (1/\bar{k}^2)\nabla_\perp^2]^{1/2} \approx \{1 + \tfrac{1}{2}[K^2(\mathbf{x}_\perp) - 1] + (1/2\bar{k}^2)\nabla_\perp^2\}, \tag{2.4}$$

which upon substitution in Eq. (2.3a) gives the ordinary parabolic wave theory,

$$\{(i/\bar{k})\partial_x + (1/2\bar{k}^2)\nabla_\perp^2 + \tfrac{1}{2}[K^2(\mathbf{x}_\perp) + 1]\}\phi^+(\mathbf{x}) = 0. \tag{2.5}$$

Truncation after the first-order terms requires both (1) $[K^2(\mathbf{x}_\perp) - 1]$ and (2) $(1/\bar{k}^2)\nabla_\perp^2$ to be in some sense small in addition to implicitly requiring that (3) the action $(1/\bar{k}^2)\nabla_\perp^2$ on $[K^2(\mathbf{x}_\perp) - 1]$ not result in terms that are appropriately large. The first restriction is the weak-inhomogeneity limitation and the second, the narrow-angle limitation, while the third and final restriction corresponds to the weak-gradient limitation.

Equation (2.3a) is the formally exact wave equation for propagation in a transversely inhomogeneous half-space supplemented with appropriate outgoing wave radiation and initial-value conditions. For the theory to be well defined and computational, an explicit representation of the square root operator, interpreted as an integral operator

$$\left[K^2(\mathbf{x}_\perp) + \frac{1}{\bar{k}^2}\nabla_\perp^2\right]^{1/2}\phi^+(x,\mathbf{x}_\perp) = \int d\mathbf{x}'_\perp\, B(\mathbf{x}_\perp,\mathbf{x}'_\perp)\phi^+(x,\mathbf{x}'_\perp), \tag{2.6}$$

must be constructed. This construction must address the operator-ordering problem presented by the noncommuting operators $K^2(\mathbf{x}_\perp)$ and $(1/\bar{k}^2)\nabla_\perp^2$.

The parabolic form of Eq. (2.3a) allows for the kernel $B(\mathbf{x}_\perp,\mathbf{x}'_\perp)$ to be recovered from the propagator G^+. This suggests a natural inverse algorithm which reconstructs the refractive index field from wave field data taken on a plane. Symbolically,

$$G^+ \rightarrow B(\mathbf{x}_\perp,\mathbf{x}'_\perp), \tag{2.7}$$

followed by

$$\begin{aligned} B(\mathbf{x}_\perp,\mathbf{x}'_\perp)B(\mathbf{x}_\perp,\mathbf{x}'_\perp) &\doteq [K^2(\mathbf{x}_\perp) + (1/\bar{k}^2)\nabla_\perp^2]^{1/2}[K^2(\mathbf{x}_\perp) + (1/\bar{k}^2)\nabla_\perp^2]^{1/2} \\ &\doteq K^2(\mathbf{x}_\perp) + (1/\bar{k}^2)\nabla_\perp^2. \end{aligned} \tag{2.8}$$

An explicit computational statement of Eqs. (2.7) and (2.8) requires the Fourier construction of the square root operator.

Equations (2.3a) and (2.3b) have been written in a form which emphasizes their correspondence with the Schrödinger equation of quantum mechanics, in particular, the analogous roles played by $1/\bar{k}$ and $\hbar$ (Planck's constant divided by 2π). More precisely, the reformulation of the second-order Helmholtz equation as a coupled first-order Schrödinger system in terms of a splitting matrix $T(\mathbf{x})$, as presented in Appendix A, is structurally analogous to the two-component representation of the Klein–Gordon equation of relativistic physics.[44] The diagonalization implicit in deriving the decoupled equations for ϕ^+ and ϕ^- [Eqs. (A8a) and (A8b)] parallels that in the Klein–Gordon theory for the complete decoupling into positive and negative frequency solutions for time-independent scalar and vector potentials. Finally, Eq. (2.5) is recognized as a nonrelativistic Schrödinger equation.

III. HOMOGENEOUS HALF-SPACE

For the special limiting case of forward wave propagation in a homogeneous half-space, $K^2(\mathbf{x}_\perp) = K_0^2$, the square root operator is readily constructed. It follows upon Fourier transforming the Helmholtz equation and applying the radiation condition that the transformed forward propagating wave field $\hat{\phi}^+(x,\mathbf{p}_\perp)$ satisfies the equation

$$[(i/\bar{k})\partial_x + (K_0^2 - \mathbf{p}_\perp^2)^{1/2}]\hat{\phi}^+(x,\mathbf{p}_\perp) = 0, \tag{3.1}$$

where $\{p_\perp\}$ is the set of perpendicular coordinates conjugate to $\{x_\perp\}$. Inverse transforming, the wave equation, Eq. (3.1), in n spatial dimensions takes the form

$$\frac{i}{\bar{k}}\partial_x\phi^+(x,\mathbf{x}_\perp) + \int d^{n-1}\mathbf{p}_\perp\,(K_0^2 - \mathbf{p}_\perp^2)^{1/2} \times \exp(i\bar{k}\mathbf{p}_\perp\cdot\mathbf{x}_\perp)\hat{\phi}^+(x,\mathbf{p}_\perp) = 0, \tag{3.2}$$

where

$$\phi^+(x,\mathbf{x}_\perp) = \int d^{n-1}\mathbf{p}_\perp\,\exp(i\bar{k}\mathbf{p}_\perp\cdot\mathbf{x}_\perp)\hat{\phi}^+(x,\mathbf{p}_\perp) \tag{3.3}$$

relates ϕ^+ and its Fourier transform $\hat{\phi}^+$ and where the square root in Eq. (3.2) is chosen to correspond to the exponentially decaying branch for the forward (outgoing) wave.[1] [In Eq. (3.2) and all subsequent equations, the integrations are understood to be over the interval $(-\infty,\infty)$.] The kernel of Eq. (2.6), subsequently, can be expressed formally as

$$B(\mathbf{y}) = \left(\frac{\bar{k}}{2\pi}\right)^{n-1}\int d^{n-1}\mathbf{p}_\perp\,(K_0^2 - \mathbf{p}_\perp^2)^{1/2}\exp(i\bar{k}\mathbf{p}_\perp\cdot\mathbf{y}), \tag{3.4}$$

where $\mathbf{y} = \mathbf{x}_\perp - \mathbf{x}'_\perp$ and takes the explicit form

$$B(y) = (K_0/2|y|)H_1^{(1)}(\bar{k}K_0|y|) \tag{3.5}$$

in two dimensions and

$$B(\mathbf{y}) = (K_0/2\pi|\mathbf{y}|^2)\exp(i\bar{k}K_0|\mathbf{y}|)(-1 + 1/i\bar{k}K_0|\mathbf{y}|) \tag{3.6}$$

in three dimensions. $H_1^{(1)}(\rho)$ is the first-order Hankel function of the first kind. The square root function and its subsequent Fourier transform, the kernel, are understood in the context of the theory of generalized functions.[45]

The extended parabolic wave equation is local for fixed Fourier components in the cross-range directions. The nonlocality in the physical space representation results from the manner of superposing the perpendicular Fourier components; there is no operator-ordering question in the homogeneous limit.

In anticipating the generalization to the transversely inhomogeneous half-space, it is natural to ask whether allowing $K_0^2 \to K^2(\mathbf{x}_\perp)$ in Eq. (3.2) provides the correct result. In general, the answer is no! This generalization does not allow for the action of the differential operator $(1/\bar{k}^2)\nabla_\perp^2$ on the refractive index field; i.e., it does not correspond to the proper order of applying the two noncommuting operators. Not surprisingly, then, it will ultimately be seen that this generalization is correct in the $\bar{k}\to\infty$ (high-frequency) limit, where the refractive index field appears essentially constant on the wavelength scale.

IV. TRANSVERSELY INHOMOGENEOUS HALF-SPACE

For the case of forward wave propagation in a transversely inhomogeneous half-space, the operator-ordering question presents itself. Considering the two-dimensional case for notational convenience, a Taylor series expansion of the wave operator

$$\begin{aligned}(1/\bar{k})\mathbf{A} &= [K^2(z) + (1/\bar{k}^2)\partial_z^2]^{1/2} \\ &= 1 + \tfrac{1}{2}\{[K^2(z) - 1] + (1/\bar{k}^2)\partial_z^2\} \\ &\quad - \tfrac{1}{8}\{[K^2(z) - 1] + (1/\bar{k}^2)\partial_z^2\}^2 + \cdots\end{aligned} \tag{4.1}$$

illustrates the ambiguity for the terms beyond first order. For different orderings of the noncommuting operators in the higher-order terms, Eq. (4.1) represents in general different integral operators. The ordinary parabolic approximation does not, in this context, directly address the ordering question, and, hence, the explicit definition of the wave operator. It may be said to indirectly address the ordering question when considering estimates of the range of its validity.

The square root of an operator is defined through its square[46]; i.e., $(1/\bar{k})\mathbf{A}$ is that operator which satisfies the operator equation

$$[(1/\bar{k})\mathbf{A}][(1/\bar{k})\mathbf{A}] = K^2(z) + (1/\bar{k}^2)\partial_z^2. \tag{4.2}$$

Application of Eq. (4.2) to Eq. (4.1) yields the symmetrical ordering or Weyl operator producing,

$$\begin{aligned}(1/\bar{k})\mathbf{A} &= 1 + \tfrac{1}{2}(\epsilon + \mu) - \tfrac{1}{8}(\epsilon^2 + \epsilon\mu + \mu\epsilon + \mu^2) \\ &\quad + \tfrac{1}{16}(\epsilon^3 + \epsilon^2\mu + \epsilon\mu\epsilon + \mu\epsilon^2 \\ &\quad + \epsilon\mu^2 + \mu\epsilon\mu + \mu^2\epsilon + \mu^3) - \cdots,\end{aligned} \tag{4.3}$$

where $\epsilon = K^2(z) - 1$ is the field strength and $\mu = (1/\bar{k}^2)\partial_z^2$. While the series representation of Eq. (4.3) is unambiguously defined, it suffers several drawbacks. Its validity is at best asymptotic in some, as yet, unspecified sense. Moreover, as the aim is to construct extended parabolic wave theories involving a wide range of conditions on the refractive index field and the propagation angle, the use of an asymptotic formulation linked to a particular physical limit (narrow angle, weak inhomogeneity, weak gradient) is clearly too restrictive.

The basis for constructing a nonperturbative framework is provided by the analogy between the forward wave propagation problem and quantum mechanics as suggested by the form of Eq. (2.3a) and the subsequent parabolic (Schrödinger) approximation. In both cases the proper meaning of an operator which is a function of the two noncommuting operators $\mathbf{Q} = q$ and $\mathbf{P} = (-i/\bar{k})\partial_q$ $(1/\bar{k}\leftrightarrow\hbar)$ must be provided. The mapping of Cohen[47] or, equivalently, of Agarwal and Wolf,[48] provides the relevant operator representation

$$\mathbf{H}(\mathbf{P},\mathbf{Q}) = \int\!\!\int du\, dv\, F(u,v)\hat{h}_{\mathbf{H}}(u,v) \exp[i\bar{k}(v\mathbf{Q} + u\mathbf{P})]. \tag{4.4}$$

In the quantum mechanical problem,

$$h_{\mathbf{H}}(p,q) = \int\!\!\int du\, dv\, \hat{h}_{\mathbf{H}}(u,v)\exp[i\bar{k}(vq + up)] \tag{4.5}$$

is the prescribed classical Hamiltonian, the physical starting point. $F(u,v)$ is a transformation function which determines the **P–Q** operator orderings. The desired operator $\mathbf{H}(\mathbf{P},\mathbf{Q})$ in nonrelativistic quantum mechanics in an electromagnetic field follows from $h_{\mathbf{H}}(p,q)$ upon the separate physical requirement of gauge covariance or, equivalently, Hermiticity, which then determines the equivalence class of functions $F(u,v)$.[49,50] Moreover, the resulting class of transformation functions has a stochastic interpretation.[51] A more detailed account of the general properties of the correspondence relationship has been given by Mizrahi.[49] In the wave propagation problem, however, the operator (or more properly, its square) is given and $h_{\mathbf{H}}(p,q)$ and $F(u,v)$ have no apparent physical significance. Only the product

$$\hat{\Omega}_{\mathbf{H}}(u,v) = F(u,v)\hat{h}_{\mathbf{H}}(u,v) \tag{4.6}$$

is relevant. In this regard, the wave propagation problem is analogous to (1) the Schrödinger equation for particle motion on a Riemannian space[50] and (2) the thermodynamic (Fokker–Planck) equation for particle diffusion.[52]

Utilizing the Baker–Campbell–Hausdorff formula,

$$\exp[i\bar{k}(v\mathbf{Q} + u\mathbf{P})] = \exp(i\bar{k}v\mathbf{Q})\exp(i\bar{k}u\mathbf{P})\exp(\tfrac{1}{2}i\bar{k}uv), \tag{4.7}$$

leads to the normal-ordered form of Eq. (4.4) and, subsequently, allows for the symbolic operator $\mathbf{H}(\mathbf{P},\mathbf{Q})$ to be written as an integral operator in the form

$$\mathbf{H}\left(\frac{-i}{\bar{k}}\partial_{q'}, q'\right)f(q') = \int du\, A_{\mathbf{H}}(q',u)f(u), \tag{4.8}$$

where the kernel is given by

$$A_{\mathbf{H}}(q',u) = \frac{\bar{k}}{2\pi}\int dp\, \Omega_{\mathbf{H}}\left(p, \frac{q'+u}{2}\right)\exp[i\bar{k}p(q'-u)] \tag{4.9}$$

and $\Omega_{\mathbf{H}}(p,q)$ is the inverse Fourier transform of $\hat{\Omega}_{\mathbf{H}}(u,v)$. Equation (4.9), like its homogeneous medium limit counterpart, Eq. (3.4), is understood in the distribution sense.[45] The formal wave equation (2.3a) can then be explicitly written as

$$\frac{i}{\bar{k}}\partial_x\phi^+(x,z) + \int dz'\left\{\frac{\bar{k}}{2\pi}\int dp\, \Omega_{\mathbf{H}}\left(p, \frac{z+z'}{2}\right)\right.$$
$$\left.\times\exp[i\bar{k}p(z-z')]\right\}\phi^+(x,z') = 0, \tag{4.10}$$

where the "symbol" $\Omega_{\mathbf{H}}(p,q)$ associated with the square root operator $(1/\bar{k})\mathbf{A} = \mathbf{H} = [K^2(q) + (1/\bar{k}^2)\partial_q^2]^{1/2}$ satisfies

$$\begin{aligned}\Omega_{\mathbf{H}^2}(p,q) &= K^2(q) - p^2\\ &= \left(\frac{\bar{k}}{\pi}\right)^2\int\!\!\int\!\!\int\!\!\int dt\, dx\, dy\, dz\, \Omega_{\mathbf{H}}(t+p, x+q)\\ &\quad\times\Omega_{\mathbf{H}}(y+p, z+q)\exp[2i\bar{k}(xy - tz)],\end{aligned} \tag{4.11}$$

$\Omega_{\mathbf{H}^2}(p,q)$ being the "symbol" associated with the square of $(1/\bar{k})\mathbf{A}$, $(1/\bar{k}^2)\mathbf{A}^2 = \mathbf{H}^2 = [K^2(q) + (1/\bar{k}^2)\partial_q^2]$. Equation (4.10) is a first-order Weyl pseudodifferential equation. Equation (4.11) is recognized as the composition equation in the Weyl pseudodifferential operator calculus[53] and embodies the definition of an operator square root in terms of its square, that is, Eq. (4.2) in conjunction with the Cohen formalism, Eqs. (4.8) and (4.9). Equation (4.11) can be written in an equivalent operator notation as,[53]

$$K^2(q) - p^2 = \lim_{\substack{\eta\to p\\ y\to q}}\cos\left[\frac{i}{2\bar{k}}(\partial_\eta\partial_q - \partial_p\partial_y)\right]\Omega_{\mathbf{H}}(p,q)\Omega_{\mathbf{H}}(\eta,y). \tag{4.12}$$

In general, the appropriate branch of the multivalued functions which solve the Weyl composition equation must be chosen to correspond to forward (outgoing) wave propagation as was done in the homogeneous medium limit.

The construction procedure for the direct propagation problem can be summarized pictorially by the following correspondence diagram:

$$\begin{array}{ccc} \mathbf{H}^2 & \Leftrightarrow & \Omega_{\mathbf{H}^2} \\ \uparrow & & \Updownarrow \\ \mathbf{H} & \Leftrightarrow & \Omega_{\mathbf{H}} \end{array}$$

where the arrows symbolize the correspondence between the appropriate quantities. A single arrow ($\to$) indicates that an operational or, algorithmic, definition of the transformation is given for the direction indicated; a double arrow ($\Leftrightarrow$) indicates that the transformation is defined in both directions. For prescribed $K^2(q)$, Eq. (4.11) must be inverted to determine $\Omega_{\mathbf{H}}(p,q)$, which then determines the wave equation through Eq. (4.10). The direct propagation algorithm proceeds around the correspondence diagram in a clockwise fashion.

Equation (4.11) is exact and provides a nonperturbative basis for the construction of the square root operator and subsequently the exact extended parabolic wave theory. In the homogeneous medium limit, Eq. (4.11) readily gives $\Omega_{\mathbf{H}}(p,q) = (K_0^2 - p^2)^{1/2}$, which in conjunction with Eq. (4.10) reproduces the results in Eqs. (3.2) and (3.4). Perturbative treatments of Eq. (4.11) or (4.12) lead to approximate extended parabolic wave theories.

In the high-frequency limit, $\bar{k}\to\infty$, taking

$$\Omega_{\mathbf{H}}(p,q) = \Omega_{\mathbf{H}}^{(0)}(p,q) + (1/\bar{k}^2)\Omega_{\mathbf{H}}^{(2)}(p,q) + \cdots \tag{4.13}$$

in conjunction with the calculations in Appendix B results in the extended parabolic equation

$$\frac{i}{\bar{k}}\partial_x\phi^+(x,z) + \int dp\, (K^2(z) - p^2)^{1/2}\exp(i\bar{k}pz)\hat{\phi}^+(x,p) = 0 \tag{4.14}$$

corresponding to the "classical" limit of $\Omega_{\mathbf{H}}(p,q)$. Equation (4.14), in its arbitrary-dimensional generalization, is exact for the homogeneous medium limit for both even and odd spatial dimension and thus extends the nonuniform WKB approximation in even dimension. The wave equation, as a wide-angle theory, can be said to incorporate the effects of diffraction that are due to an inhomogeneous source field at $x = 0$. Thus, Eq. (4.14) can be distinguished from geometric acoustics (as distinct from the WKB propagator and integration over the source field) in that the latter is not, in general, exact for a homogeneous medium since it does not incorporate the effects of diffraction that are due to an inhomogeneous source field. The high-frequency approximation of Eq. (4.14) is, in general, distinct from the geometrical approximations. Since the path integral representation of the solution to Eq. (4.14) contains cntributions from "all of the paths,"[43] the high-frequency theory incorporates, in an approximate manner, diffraction effects due to medium inhomogeneity. While medium diffraction effects could be further incorporated into the high-frequency wave equation through the extended treatment outlined in Appendix B, this would not correspond to a uniform asymptotic treatment. Rather, Eq. (4.11) must be solved in a uniform manner in the $\bar{k}\to\infty$ limit to provide the appropriate kernel function for incorporating diffraction phenomena.

In the limits corresponding to (1) narrow angle, weak inhomogeneity, and weak gradient, (2) narrow angle, arbitrary inhomogeneity (field strength), and weak gradient, and (3) arbitrary angle, weak inhomogeneity, and weak gradient,

$$\begin{aligned}\Omega_{\mathrm{H}}(p,q)\Omega_{\mathrm{H}}(\eta,y) = {}& \Omega_{\mathrm{H}}^{(0)}(p,q)\Omega_{\mathrm{H}}^{(0)}(\eta,y)\\ &+\Delta\,[\Omega_{\mathrm{H}}^{(0)}(p,q)\Omega_{\mathrm{H}}^{(1)}(\eta,y)\\ &+\Omega_{\mathrm{H}}^{(0)}(\eta,y)\Omega_{\mathrm{H}}^{(1)}(p,q)]\\ &+\Delta^2[\Omega_{\mathrm{H}}^{(0)}(p,q)\Omega_{\mathrm{H}}^{(2)}(\eta,y)\\ &+\Omega_{\mathrm{H}}^{(0)}(\eta,y)\Omega_{\mathrm{H}}^{(2)}(p,q)+\Omega_{\mathrm{H}}^{(1)}(p,q)\Omega_{\mathrm{H}}^{(1)}(\eta,y)]\\ &+\cdots. \end{aligned} \tag{4.15}$$

Here Δ is a formal expansion parameter introduced into Eq.(4.11) or (4.12) by setting

(1) $\Omega_{\mathrm{H}^2}(p,q) = 1+\Delta\,\{[K^2(q)-1]-p^2\}$,

(2) $\Omega_{\mathrm{H}^2}(p,q) = K^2(q)-\Delta p^2$,

(3) $\Omega_{\mathrm{H}^2}(p,q) = (1-p^2)+\Delta\,[K^2(q)-1]$,

respectively. The subsequent perturbation theory for the transversely inhomogeneous medium provides the appropriate generalizations to the homogeneous medium limit expansions of $\Omega_{\mathrm{H}}(p,q)$ given, respectively, in the form

(1) $\Omega_{\mathrm{H}}(p,q) = \{1+\Delta\,[(K_0^2-1)-p^2]\}^{1/2}$,

(2) $\Omega_{\mathrm{H}}(p,q) = K_0(1-\Delta p^2/K_0^2)^{1/2}$,

(3) $\Omega_{\mathrm{H}}(p,q) = (1-p^2)^{1/2}[1+\Delta\,(K_0^2-1)/(1-p^2)]^{1/2}$.

The calculations are presented in Appendix C; the resulting wave equations are summarized here. In the limit of narrow angle, weak inhomogneity, and weak gradient, the ordinary parabolic wave equation,

$$(i/\bar{k})\partial_x\phi^+(x,z)+\{(1/2\bar{k}^2)\partial_z^2+\tfrac{1}{2}[K^2(z)+1]\}\phi^+(x,z)=0, \tag{4.16}$$

is recovered. This perturbative equation is formally analogous to the nonrelativistic limit in the Klein–Gordon theory resulting from the inverse mass expansion of the Hamiltonian achieved through the application of the Foldy–Wouthuysen approximate diagonalization.[44] Equation (4.16) is further suggestive of weak-coupling (simple diffusion) approximations to the Master equation in statistical mechanics.[54,55] The ordinary parabolic wave theory is a full-wave approximation applicable in the high-frequency regime, although as a narrow-angle theory it is not appropriate for wide-angle source fields.

In the limit of narrow angle, arbitrary inhomogeneity, and weak gradient, the extended parabolic equation is local in the form

$$\begin{aligned}&\frac{i}{\bar{k}}\partial_x\phi^+(x,z)+\frac{1}{2\bar{k}^2}\partial_z\left(\frac{1}{K(z)}\partial_z\phi^+(x,z)\right)\\ &\quad+\left[K(z)+\frac{1}{4\bar{k}^2}\left(\frac{[K'(z)]^2}{[K(z)]^3}\right.\right.\\ &\quad\left.\left.-\frac{K''(z)}{[K(z)]^2}\right)\right]\phi^+(x,z)=0,\end{aligned} \tag{4.17}$$

reproducing the result previously derived by Tappert.[1] Equation (4.17) is suggestive of weak-coupling (generalized diffusion) approximations to the Master equation[54,55] and particle motion on curved spaces (constrained systems).[56] The wave theory is again a high-frequency theory inappropriate for extended source fields. Equation (4.17) is of the general form considered by Langouche, Roekaerts, and Tirapegui.[52]

In the limit of arbitrary angle, weak inhomogeneity, and weak gradient, the extended parabolic equation takes the form

$$\begin{aligned}&\frac{i}{\bar{k}}\partial_x\phi^+(x,z)+\int dz'\,\frac{1}{2|z-z'|}H_1^{(1)}(\bar{k}\,|z-z'|)\phi^+(x,z')\\ &\quad+\int dz'\left(\int dq\,[K^2(q)-1]\hat{R}(z'-q,q-z)\right)\phi^+(x,z')=0,\end{aligned} \tag{4.18}$$

where

$$\hat{R}(\alpha,\beta)=\left(\frac{\bar{k}}{2\pi}\right)^2\int\!\!\int d\xi\,d\eta\times\frac{\exp(-i\bar{k}\xi\alpha)\exp(-i\bar{k}\eta\beta)}{(1-\xi^2)^{1/2}+(1-\eta^2)^{1/2}} \tag{4.19}$$

is a generalization of a diffraction integral discussed by Watson.[57] The Hankel function kernel is the homogeneous medium result, and the square root functions are to be taken to correspond to the exponentially decaying branch for the forward (outgoing) wave consistent with the treatment of the homogeneous medium limit. Equation (4.18) is a nonlocal extended parabolic wave equation and makes explicit Tappert's symbolic equation.[1] Further, Eq. (4.18) is a full-wave approximation, a wide-angle theory appropriate for extended source fields, not inherently a high-frequency theory, and suggestive of strong-coupling approximations to the Master equation in linear-gas relaxation theory.[54,55] A first-order (in the field strength) solution to Eq. (4.18) gives the Born approximation.[58] The transformed wave field satisfies

$$\frac{i}{\bar{k}}\partial_x\hat{\phi}^+(x,p)+(1-p^2)^{1/2}\hat{\phi}^+(x,p)+\int dp'\,\frac{\hat{\epsilon}(p-p')\hat{\phi}^+(x,p')}{(1-p^2)^{1/2}+(1-p'^2)^{1/2}}=0, \tag{4.20}$$

where $\hat{\epsilon}$ is the Fourier transform of the field strength. Equation (4.20) illustrates the nonlocality for a fixed Fourier component introduced by the presence of medium inhomogeneity.

The Weyl composition equation is amenable to uniform asymptotic analysis. Of particular interest is a slowly varying medium with an isolated region of rapid, or even discontinuous, variation. Moreover, in higher spatial dimensions

$(n > 2)$, the transverse coordinates can be treated unsymmetrically leading to extended parabolic wave equations of the form considered by Palmer.[21-23] Expressing the formal wave equation (2.3a) in the equivalent form

$$(i/\bar{k})[K^2(z) + (1/\bar{k}^2)\partial_z^2]^{1/2}\partial_x \phi^+(x,z) + [K^2(z) + (1/\bar{k}^2)\partial_z^2]\phi^+(x,z) = 0 \tag{4.21}$$

and subsequently applying the approximate representations of the square root operator leads to a sequence of wide-angle propagation theories. In particular, the ordinary parabolic approximation gives

$$(i/\bar{k})\{\tfrac{1}{2}[K^2(z) + 1] + (1/2\bar{k}^2)\partial_z^2\}\partial_x \phi^+(x,z) + [K^2(z) + (1/\bar{k}^2)\partial_z^2]\phi^+(x,z) = 0. \tag{4.22}$$

Equation (4.22) is no longer parabolic in the usual sense, but bears a striking similarity to strong-coupling limit equations in linear-gas relaxation theory.[54,55] This wide-angle wave equation, formally derivable from a rational operator function approximation to the square root operator, is equivalent to an infinite-order approximation (in p^2) to $\Omega_{\mathbf{H}}(p,q)$. The Weyl composition equation thus provides the analytic framework for the rational approximation to the wave equation developed by Greene.[59]

The Weyl composition equation can, in fact, be inverted for several nontrivial $K^2(q)$ profiles. In particular, solutions follow for linear, quadratic, delta function, and discontinuity profiles which provide for an exact analysis of strong refractive and diffractive effects.[60]

The arbitrary-dimensional generalizations of the results in this section follow immediately.

V. INVERSE FORMULATION

The factorization analysis presented in Sec. IV makes explicit the symbolic inverse formulation outlined in Sec. II. Mathematically, the refractive index field (or its square) is reconstructed from the full-space Helmholtz Green's function G.[60] The reflection principle (or method of images) relates the half-space propagator G^+ and the full-space Green's function G through

$$G^+(x,\mathbf{x}_\perp|0,\mathbf{x}'_\perp) = -2\partial_x G(x,\mathbf{x}_\perp|0,\mathbf{x}'_\perp). \tag{5.1}$$

The parabolic form of the wave equation (4.10) then relates the kernel $B(\mathbf{x}_\perp, \mathbf{x}'_\perp)$ to G^+ through

$$B(\mathbf{x}_\perp,\mathbf{x}'_\perp) = \frac{-i}{\bar{k}} \lim_{x\to 0}[\partial_x G^+(x,\mathbf{x}_\perp|0,\mathbf{x}'_\perp)]. \tag{5.2}$$

Combining Eqs. (5.1) and (5.2) then reconstructs the kernel function from Green's function data taken on the initial-value plane in the form

$$B(\mathbf{x}_\perp,\mathbf{x}'_\perp) = \frac{2i}{\bar{k}} \lim_{x\to 0}[\partial_x^2 G(x,\mathbf{x}_\perp|0,\mathbf{x}'_\perp)]. \tag{5.3}$$

The symbol $\Omega_{\mathbf{H}}(p,q)$ (in its arbitrary-dimensional form) is then constructed through an inverse Fourier transform of the kernel function as expressed in Eq. (4.9) and subsequently yields the refractive index field upon a direct application of the Weyl composition equation (4.11), for $p = 0$. The inverse algorithm proceeds around the correspondence diagram (pictorial summary) in a counterclockwise fashion. The direct propagation algorithm requires the inversion of Eq. (4.11) while the inverse propagation algorithm only requires a direct computation of Eq. (4.11).

For the physical experiment, a point source is introduced into the medium defining the initial-value $(x = 0)$ plane. The second derivative with respect to the range of the pressure field is then determined as a function of the point source position. This set of measurements can, in fact, be accomplished on a downfield plane.

This inverse formulation can be distinguished from those associated with plane wave sources and far-field data. The location of both the field source (finite) and the data measurements is within the scattering region. In this regard, the analysis is similar to the stratified environmental model of Stickler and Deift[61] and thus applicable to ocean environments and certain seismic (bore-hole) experiments. Most importantly, the method is a direct inversion of an arbitrary-dimensional propagation equation which requires less symmetry than those models (i.e., Stickler–Deift) reducible to the standard one-dimensional formulation of Deift and Trubowitz[62] or Gel'fand and Levitan.[63] Thus, for example, in a general n-dimensional Cartesian formulation, the refractive index field can be a function of as many as $(n - 1)$ coordinates in the factorization model, while a function of only one coordinate in an "effective one-dimensional" model.

The factorization algorithm exactly inverts the inherently nonlinear relationship between the measured data and the refractive index field as reflected in the Lippmann–Schwinger equation for the propagator.[58] Approximate inversion algorithms follow readily from the perturbative treatments of the Weyl composition equation. $K^2(q)$ is related to $\Omega_{\mathbf{H}}(0,q)$ in a quadratic fashion and through a linear integral relationship, respectively, in the high-frequency $(\bar{k}\to\infty)$ and weak-inhomogeneity (Born) limits. In conjunction with a weak backscatter perturbation theory suggested by the form of Eqs. (A7a) and (A7b), the factorization inversion algorithm could be extended to weakly range-dependent environments.

VI. DISCUSSION

The association of Eqs. (4.8)–(4.12) with the theory of pseudodifferential operators[64-67] provides for a mathematical framework for the evaluation of the approximate extended parabolic wave theories. The Cohen formalism is readily connected to the theory of pseudodifferential operators. The choice of the standard $[F(u,v) = \exp(-\tfrac{1}{2}i\bar{k}uv)]$, antistandard $[F(u,v) = \exp(\tfrac{1}{2}i\bar{k}uv)]$, and Weyl $[F(u,v) = 1]$ orderings in Eqs. (4.6), (4.8), and (4.9) results, respectively, in the standard,

$$\mathbf{H}^s\Big(\frac{-i}{\bar{k}}\partial_{q'},q'\Big)f(q') = \frac{\bar{k}}{2\pi}\iint du\, dp\, h_{\mathbf{H}}(p,q') \times \exp[i\bar{k}p(q'-u)]f(u), \tag{6.1}$$

the antistandard,

$$\mathbf{H}^a\Big(\frac{-i}{\bar{k}}\partial_{q'},q'\Big)f(q') = \frac{\bar{k}}{2\pi}\iint du\, dp\, h_{\mathbf{H}}(p,u) \times \exp[i\bar{k}p(q'-u)]f(u), \tag{6.2}$$

and the Weyl,

$$\mathbf{H}^{w}\left(\frac{-i}{\bar{k}}\partial_{q'},q'\right)f(q')=\frac{\bar{k}}{2\pi}\int\int du\,dp\,h_{\mathbf{H}}\left(p,\frac{q'+u}{2}\right)\times\exp[i\bar{k}p(q'-u)]f(u),\quad(6.3)$$

pseudodifferential operators. To be more mathematically precise, the pseudodifferential operators in Eqs. (6.1)–(6.3) and the corresponding operator calculi are understood for appropriate symbol $(h_{\mathbf{H}}(p,q))$ classes defined through estimates on the symbol derivatives. Roughly speaking, the pseudolocal character of the operators is reflected in the large $|p|$ behavior of the symbol and its derivatives; a rapid decay at infinity for high p derivatives is required.

For symbols which are polynomials in p, $\mathbf{H}^{s}((-i/\bar{k})\partial_{q'},q')$ corresponds to $h_{\mathbf{H}}(p,q')$ with p replaced by $(-i/\bar{k})\partial_{q'}$ put to the right of the coefficients, $\mathbf{H}^{a}((-i/\bar{k})\partial_{q'},q')$ corresponds to $h_{\mathbf{H}}(p,q')$ with p replaced by $(-i/\bar{k})\partial_{q'}$ put to the left of the coefficients, and $\mathbf{H}^{w}((-i/\bar{k})\partial_{q'},q')$ corresponds to the symmetric compromise which associates the term $\chi(q')p^{n}$ with the operator

$$\frac{1}{2^{n}}\sum_{j=0}^{n}\binom{n}{j}\left(\frac{-i}{\bar{k}}\partial_{q'}\right)^{j}\chi(q')\left(\frac{-i}{\bar{k}}\partial_{q'}\right)^{n-j}.$$

Equations (6.1)–(6.3) then provide extensions of partial differential operators with nonconstant coefficients through the enlargement of the class of admissible symbols beyond polynomials in p. In the general case then, for example, Eq. (6.1) loosely corresponds to allowing the differential operator to act first followed by the operation of multiplication by functions of q'. The antistandard and Weyl pseudodifferential operators can be viewed as rearrangements of the more common standard pseudodifferential operator. In the wave propagation problem the Weyl representation, through Eqs. (4.6), (4.8), and (4.9), is canonical $[\Omega_{\mathbf{H}}(p,q)=h_{\mathbf{H}}(p,q)]$. Standard or antistandard representations of the determined forward wave operator $(\Omega_{\mathbf{H}}(p,q))$ correspond to different functions $h_{\mathbf{H}}(p,q)$ and follow from Eq. (4.6).

From a strictly mathematical viewpoint, this analysis has been formal in nature. Equation (4.11) must be examined in detail with respect to classes of functions $K^{2}(q)$ to establish the appropriate estimates on $\Omega_{\mathbf{H}}(p,q)$ necessary for the proper formulation of Weyl pseudodifferential operators. Further, the subsequent first-order Weyl pseudodifferential wave equation is subject to questions concerning existence and uniqueness. In this regard, the relevant physics of forward wave propagation provides the constructive guide.

The n-dimensional Helmholtz equation for a transversely inhomogeneous medium is naturally related to parabolic propagation models through (1) the n-dimensional extended parabolic (pseudodifferential) equation and (2) an imbedding in an $(n+1)$-dimensional parabolic (Schrödinger) equation. The first relationship provides the basis for the operator analysis while the interplay between these two formulations suggests the development of parabolic and elliptic-based path integral representations. The path integrals provide a global perspective of the transition from elliptic to parabolic wave theory and further allow for the natural introduction of the concept of an underlying stochastic process and the notion of strong- and weak-coupling regimes, in addition to an interpretation in terms of free motion on curved spaces. Specifically, the Hamiltonian phase space path integral representation of the propagator,

$$G^{+}(x,z|0,z_s)=\lim_{N\to\infty}\int\prod_{j=1}^{N-1}dz_j\prod_{j=1}^{N}\frac{dp_j^{z}}{2\pi/\bar{k}}\times\exp\left\{i\bar{k}\sum_{j=1}^{N}\left[p_j^{z}(z_j-z_{j-1})+\frac{x}{N}\Omega_{\mathbf{H}}\left(p_j^{z},\frac{z_j+z_{j-1}}{2}\right)\right]\right\},\qquad(6.4)$$

in conjunction with the representations associated with Feynman and Fradkin, Feynman and Garrod, and Feynman and DeWitt-Morette provide a unifying framework for dynamical approximations, resolution of the square root operator, and the concept of an underlying stochastic process.[43] Computationally, the Hamiltonian phase space path integrals corresponding to approximate resolutions of the square root operator $[\Omega_{\mathbf{H}}(p,q)]$ should provide the basis for a marching algorithm in a manner analogous to that provided by the split-step FFT algorithm[38] for the ordinary parabolic approximation.

The principal extension is the inclusion of backscatter effects. The exact formalism developed for the transversely inhomogeneous medium can provide the basis for perturbation treatments in two distinct ways. The formal field splitting analysis in Appendix A [Eqs. (A7a) and (A7b) as mentioned in Sec. V] suggests the inclusion of weak backscatter effects in an obvious manner.[1] The imbedding of the n-dimensional Helmholtz equation in an $(n+1)$-dimensional parabolic problem, in conjunction with recent work by DeSanto[24,25] on the imbedding of the elliptic radiation problem in an appropriate n-dimensional parabolic model, suggests the inclusion of backscatter effects through imbedding methods[68] focusing on the spatial dimension as a variable. This imbedding can be viewed as a dimensional perturbation theory. The inclusion of backscatter effects in the direct propagation problem would have its natural parallel in the inverse formulation.

There are several areas of direct application. The structural similarity between the wave propagation problem and the two-component formulation of the Klein–Gordon equation indicates the applicability of the factorization analysis to hyperbolic wave equations in the time domain. This is reinforced by Davison's[69] general approach to field splitting and invariant imbedding for linear wave equations in conjunction with the suggestive analysis in Appendix A.

The operator and corresponding path integral forms of the factorization analysis provide the framework to extend the narrow-angle, weak-inhomogeneity, and weak-gradient acoustic field coherence formulation.[70] The incorporation of the elliptic effects of the Helmholtz theory would subsequently lead to the development of extended coherence theories of which the ordinary parabolic would be but one particular limiting case. The pseudodifferential coherence formulation then suggests the development of an analogous inverse formulation. A time-domain coherence formulation (broad-band signals) is also amenable to a pseudodifferential operator analysis which would extend the scope of a naturally suggested "geometric" theory.

Wave propagation in elastic media can also be addressed. The analysis in this case would involve the extension of the scalar factorization methods to the appropriate vector formulation. The development of the appropriate nonlocal wave theory would serve to both clarify and extend the recent work of Corones, DeFacio, and Krueger[6,18] based on field splitting techniques. The validity of the ordinary scalar parabolic wave theory requires that the chosen reference wave number $\bar{k}$ be approximately equal to the actual wave number k. In the extended formulation the value chosen for the reference wave number is not material. A difficulty in constructing a parabolic stress wave theory results because of an ambiguity in an appropriate choice of the reference wave number, since there is more than one propagation mode with each mode suggesting a different reference.[15–18] The calculations in this paper suggest that a systematically derived vector parabolic wave theory, even to lowest order, will be nonlocal in the manner in which the cross-range coordinates are incorporated. Apparently, local parabolic approximations only result upon mode separation. Further, Eq. (4.20) can be directly derived from the Lippmann–Schwinger equation for the propagator at the level of the Born approximation in conjunction with Eq. (5.3). Inasmuch as this procedure can be interpreted to be a derivation by range incrementing, this suggests the use of a range incrementing derivation of a vector parabolic wave theory.[17] It is not difficult to show, however, that this would not be correct for a wide-angle treatment due to the difficulty of mode coupling at the boundary of the half-space. It may prove to be valid for a narrow-angle treatment, but justification of this requires further study. Finally, inverse algorithms have already been developed for problems that are reducible to one-dimensional formulations.[71,72]

ACKNOWLEDGMENTS

The work reported here was supported under a grant from the National Science Foundation (EAR-8025367). In addition, L. F. spent one summer with the Large Aperture Acoustics Branch of the Naval Research Laboratory.

APPENDIX A: FIELD SPLITTINGS AND THE WAVE EQUATION

The reduced scalar Helmholtz equation

$$[\nabla^2 + \bar{k}^2 K^2(\mathbf{x})]\phi(\mathbf{x}) = 0 \tag{A1}$$

can be written as a first order system

$$\partial_x \begin{pmatrix} \phi(\mathbf{x}) \\ \partial_x \phi(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -(\nabla_\perp^2 + \bar{k}^2 K^2(\mathbf{x})) & 0 \end{pmatrix} \begin{pmatrix} \phi(\mathbf{x}) \\ \partial_x \phi(\mathbf{x}) \end{pmatrix}, \tag{A2}$$

where x is the distinguished or range coordinate. A formally arbitrary splitting matrix $T(\mathbf{x})$ defines a decomposition of the total wave field into forward (+) and backward (−) components,

$$\begin{pmatrix} \phi^+(\mathbf{x}) \\ \phi^-(\mathbf{x}) \end{pmatrix} = T(\mathbf{x}) \begin{pmatrix} \phi(\mathbf{x}) \\ \partial_x \phi(\mathbf{x}) \end{pmatrix}, \tag{A3}$$

with the subsequent transformation of the wave equation (A2) into the form

$$\partial_x \begin{pmatrix} \phi^+(\mathbf{x}) \\ \phi^-(\mathbf{x}) \end{pmatrix} = \Big\{ \partial_x T(\mathbf{x})[T^{-1}(\mathbf{x})] + T(\mathbf{x}) \begin{pmatrix} 0 & 1 \\ -(\nabla_\perp^2 + \bar{k}^2 K^2(\mathbf{x})) & 0 \end{pmatrix} T^{-1}(\mathbf{x}) \Big\} \times \begin{pmatrix} \phi^+(\mathbf{x}) \\ \phi^-(\mathbf{x}) \end{pmatrix}. \tag{A4}$$

For an arbitrary splitting, the forward and backward wave fields do not have an obvious physical interpretation in terms of propagation solely in the positive and negative x directions, respectively. Taking the splitting matrix $T(\mathbf{x})$ to be

$$T_1(\mathbf{x}) = \begin{pmatrix} 1 & -i/\mathbf{A} \\ 1 & i/\mathbf{A} \end{pmatrix}, \tag{A5}$$

where

$$(1/\bar{k})\mathbf{A} = [K^2(\mathbf{x}) + (1/\bar{k}^2)\nabla_\perp^2]^{1/2}, \tag{A6}$$

gives the following set of coupled equations for the forward and backward wave fields:

$$\partial_x \phi^+(\mathbf{x}) = (-\tfrac{1}{2} i \partial_x \mathbf{A}^{-1} + 1)(i\mathbf{A})\phi^+(\mathbf{x}) + \tfrac{1}{2} i \partial_x \mathbf{A}^{-1}(i\mathbf{A})\phi^-(\mathbf{x}) \tag{A7a}$$

and

$$\partial_x \phi^-(\mathbf{x}) = \tfrac{1}{2} i \partial_x \mathbf{A}^{-1}(i\mathbf{A})\phi^+(\mathbf{x}) + (-\tfrac{1}{2} i \partial_x \mathbf{A}^{-1} - 1)(i\mathbf{A})\phi^-(\mathbf{x}). \tag{A7b}$$

When $K^2(\mathbf{x})$ is range-independent, $K^2(\mathbf{x}) = K^2(\mathbf{x}_\perp)$, Eqs. (A7a) and (A7b) decouple and yield the exact equations for the forward and backward wave fields as given by Eqs. (2.3a) and (2.3b),

$$(i/\bar{k})\partial_x \phi^+(\mathbf{x}) = (-1/\bar{k})\mathbf{A}\phi^+(\mathbf{x}) \tag{A8a}$$

and

$$(i/\bar{k})\partial_x \phi^-(\mathbf{x}) = (1/\bar{k})\mathbf{A}\phi^-(\mathbf{x}). \tag{A8b}$$

In this physically obvious case, the forward wave field corresponds to propagation in the direction of positive x while the backward wave field corresponds to propagation in the direction of negative x. If $K^2(\mathbf{x})$ is a slowly varying function of range, then Eqs. (A8a) and (A8b) serve as extended parabolic approximations to the Helmholtz equation. Despite the nonlocal nature of the operator in the splitting matrix $T_1(\mathbf{x})$, this particular field decomposition is consistent with the principle of localization of Bellman and Kalaba.[20]

By applying other splitting matrices in the range-independent case and subsequently decoupling the equations by effectively setting the off-diagonal terms equal to zero, local approximations to the exact nonlocal extended parabolic equation are obtained. For example, the splitting matrix

$$T_2(\mathbf{x}) = \frac{1}{2} \begin{pmatrix} 1 & -i/\bar{k}K(\mathbf{x}_\perp) \\ 1 & i/\bar{k}K(\mathbf{x}_\perp) \end{pmatrix} \tag{A9}$$

considered by Corones[5] gives the coupled set of equations

$$\partial_x \phi^+(\mathbf{x}) = \Big(i\bar{k}K(\mathbf{x}_\perp) + \frac{i}{2\bar{k}K(\mathbf{x}_\perp)} \nabla_\perp^2 \Big)\phi^+(\mathbf{x}) + \Big(\frac{i}{2\bar{k}K(\mathbf{x}_\perp)} \nabla_\perp^2 \Big)\phi^-(\mathbf{x}) \tag{A10a}$$

and

$$\partial_x \phi^-(\mathbf{x}) = \left(\frac{-i}{2\bar{k}K(\mathbf{x}_\perp)} \nabla_\perp^2 \right) \phi^+(\mathbf{x}) + \left(-i\bar{k}K(\mathbf{x}_\perp) - \frac{i}{2\bar{k}K(\mathbf{x}_\perp)} \nabla_\perp^2 \right) \phi^-(\mathbf{x}). \tag{A10b}$$

While Eqs. (A10a) and (A10b) are exact, it is seen that, in this splitting, the forward and backward fields do not decouple in this range-independent case. The spurious or "kinematical" reflections are an artifact of the arbitrary (and thus physically approximate) identification of the forward and backward wave fields. Equations (A8a) and (A8b) can thus be said to fully take into account the kinematics of propagation.[6] Neglecting the "reflection operators" decouples the equations, giving

$$(i/\bar{k})\partial_x \phi^+(\mathbf{x}) = -\{K(\mathbf{x}_\perp) + [1/2\bar{k}^2 K(\mathbf{x}_\perp)]\nabla_\perp^2\}\phi^+(\mathbf{x}), \tag{A11}$$

which corresponds to the approximation

$$\{K^2(\mathbf{x}_\perp) + (1/\bar{k}^2)\nabla_\perp^2\}^{1/2} \simeq \{K(\mathbf{x}_\perp) + [1/2\bar{k}^2 K(\mathbf{x}_\perp)]\nabla_\perp^2\}. \tag{A12}$$

While an infinite number of such "extended parabolic type" equations can be derived in this manner, they do not in general correspond to an *a priori* systematic perturbation procedure in terms of an appropriate small parameter. Also, the proper operator-ordering inherent in the definition of the square root operator is often not explicitly addressed in these approaches. Noting the perturbation result summarized in Eq. (C17), it is seen *a posteriori* that Eqs. (A11) and (A12) correspond to the limit of narrow angle and arbitrary field strength for an approximately constant refractive index field.

APPENDIX B: HIGH-FREQUENCY PERTURBATION THEORY

Substituting the expansion in Eq. (4.13) into Eq. (4.12) results in the $\mathscr{O}(1)$ equation

$$K^2(q) - p^2 = \lim_{\substack{\eta\to p\\ y\to q}} [\Omega_{\mathrm{H}}^{(0)}(p,q)\Omega_{\mathrm{H}}^{(0)}(\eta,y)] \tag{B1}$$

with the subsequent solution

$$\Omega_{\mathrm{H}}^{(0)}(p,q) = [K^2(q) - p^2]^{1/2}. \tag{B2}$$

The $\mathscr{O}(1/\bar{k}^2)$ equation follows as

$$\Omega_{\mathrm{H}}^{(0)}(p,q)\Omega_{\mathrm{H}}^{(2)}(p,q) = \frac{1}{16} \lim_{\substack{\eta\to p\\ y\to q}} (\partial_\eta^2\partial_q^2 + \partial_p^2\partial_y^2 - 2\partial_\eta\partial_y\partial_p\partial_q) \times (\Omega_{\mathrm{H}}^{(0)}(p,q)\Omega_{\mathrm{H}}^{(0)}(\eta,y)) \tag{B3}$$

and has the solution

$$\Omega_{\mathrm{H}}^{(2)}(p,q) = \frac{-1}{8} \frac{K^3(q)K''(q)}{[K^2(q) - p^2]^{5/2}}. \tag{B4}$$

The superscript primes in Eq. (B4) denote differentiation with respect to the argument. Thus, to $\mathscr{O}(1/\bar{k}^2)$,

$$\Omega_{\mathrm{H}}(p,q) = [K^2(q) - p^2]^{1/2} - \frac{1}{8\bar{k}^2} \frac{K^3(q)K''(q)}{[K^2(q) - p^2]^{5/2}}. \tag{B5}$$

The approximation inherent in Eq. (B5) clearly breaks down in regions where the refractive index field varies rapidly on the wavelength scale and is not in general uniformly valid as follows from the form of the $\mathscr{O}(1/\bar{k}^2)$ term. This asymptotic development does not include terms of exponential order.

The operator can be expressed in terms of the Weyl symbol in the following symbolic manner[53]:

$$\mathbf{H}(\mathbf{P},\mathbf{Q}) = \{\exp[(1/2i\bar{k})\partial_p\partial_q]\Omega_{\mathrm{H}}(p,q)\}\Big|_{\substack{p\to\mathbf{P}\\ q\to\mathbf{Q}\\ \mathbf{Q}\text{before}\mathbf{P}}}. \tag{B6}$$

With the result of Eq. (B5), Eq. (B6) takes the form

$$\mathbf{H}(\mathbf{P},\mathbf{Q}) = (1/\bar{k})\mathbf{A}_{\bar{k}\to\infty} ([1 + (1/2i\bar{k})\partial_p\partial_q + \mathscr{O}(1/\bar{k}^2)] \times \{[K^2(q) - p^2]^{1/2} + \mathscr{O}(1/\bar{k}^2)\})\Big|_{\substack{p\to\mathbf{P}\\ q\to\mathbf{Q}\\ \mathbf{Q}\text{before}\mathbf{P}}}, \tag{B7}$$

which to the lowest order yields

$$\mathbf{H}(\mathbf{P},\mathbf{Q}) = (1/\bar{k})\mathbf{A}_{\bar{k}\to\infty} [K^2(q) - p^2]^{1/2}\Big|_{\substack{p\to\mathbf{P}\\ q\to\mathbf{Q}\\ \mathbf{Q}\text{before}\mathbf{P}}}. \tag{B8}$$

Equation (B8) is equivalent to the standard pseudodifferential operator corresponding to the symbol $[K^2(q) - p^2]^{1/2}$, and thus the forward propagating wave equation is given by Eq. (4.14) in the high-frequency limit.

The kernel associated with Eq. (4.14) is that given by Eq. (3.5) with the identification $K_0 \to K(z)$. This correspondence results since the commutator $[\mathbf{Q},\mathbf{P}] \to 0$ in the $\bar{k}\to\infty$ limit, or, in a more physical sense, since the refractive index field appears constant on the wavelength scale. The high-frequency wave equation, in retaining p^2 and $K^2(z) - 1$ terms to all orders, explicitly addresses the operator-ordering question, although in the simplifying "classical" limit. This approximate theory is thus applicable for arbitrary angle and field strength in a sufficiently high-frequency regime. In the same manner, for example, the Weyl and antistandard pseudodifferential operators approach the standard, or "**Q** before **P**," operator in the $\bar{k}\to\infty$ limit.

APPENDIX C: EXTENDED PARABOLIC PERTURBATION THEORY

For the limiting case of narrow angle, weak inhomogeneity, and weak gradient, Eqs. (4.12) and (4.15) to $\mathscr{O}(1)$ result in the equation

$$1 = \lim_{\substack{\eta\to p\\ y\to q}} \cos[(i/2\bar{k})(\partial_\eta\partial_q - \partial_p\partial_y)]\Omega_{\mathrm{H}}^{(0)}(p,q)\Omega_{\mathrm{H}}^{(0)}(\eta,y), \tag{C1}$$

which has the solution

$$\Omega_{\mathrm{H}}^{(0)}(p,q) = 1, \tag{C2}$$

consistent with the Weyl correspondences $g_1(\mathbf{P}) \leftrightarrow g_1(p)$ and $g_2(\mathbf{Q}) \leftrightarrow g_2(q)$. The $\mathscr{O}(\Delta)$ equation is given by

$$[K^2(q) - 1] - p^2 = \lim_{\substack{\eta\to p\\ y\to q}} \cos[(i/2\bar{k})(\partial_\eta\partial_q - \partial_p\partial_y)] \times (\Omega_{\mathrm{H}}^{(1)}(p,q) + \Omega_{\mathrm{H}}^{(1)}(\eta,y)) \tag{C3}$$

$$= \lim_{\substack{\eta\to p\\ y\to q}} [\Omega_{\mathrm{H}}^{(1)}(p,q) + \Omega_{\mathrm{H}}^{(1)}(\eta,y)], \tag{C4}$$

which has the solution

$$\Omega_{\mathrm{H}}^{(1)}(p,q)=\tfrac{1}{2}\{[K^2(q)-1]-p^2\}. \tag{C5}$$

The $\mathscr{O}(\Delta^2)$ equation is given by

$$2\Omega_{\mathrm{H}}^{(2)}(p,q)=-\lim_{\substack{\eta\to p\\ y\to q}}\cos[(i/2\bar{k})(\partial_\eta\partial_q-\partial_p\partial_y)] \times\Omega_{\mathrm{H}}^{(1)}(p,q)\Omega_{\mathrm{H}}^{(1)}(\eta,y) \tag{C6}$$

$$=-\lim_{\substack{\eta\to p\\ y\to q}}[1-(1/8\bar{k}^2)(\partial_\eta\partial_q-\partial_p\partial_y)^2] \times\Omega_{\mathrm{H}}^{(1)}(p,q)\Omega_{\mathrm{H}}^{(1)}(\eta,y), \tag{C7}$$

which has the solution

$$\Omega_{\mathrm{H}}^{(2)}(p,q)=-\tfrac{1}{8}\{[K^2(q)-1)]-p^2\}^2-(1/16\bar{k}^2)[K^2(q)]''. \tag{C8}$$

Thus, to $\mathscr{O}(\Delta^2)$,

$$\Omega_{\mathrm{H}}(p,q)=1+\tfrac{1}{2}\Delta\{[K^2(q)-1]-p^2\} -\tfrac{1}{8}\Delta^2(\{[K^2(q)-1]-p^2\}^2+(1/2\bar{k}^2)[K^2(q)]''). \tag{C9}$$

To $\mathscr{O}(\Delta)$, Eq. (C9) gives the ordinary parabolic approximation, Eq. (4.16). In addition to the limits of narrow angle and weak inhomogeneity, Eq. (C9) clearly demonstrates that this approximation requires that the variations in the refractive index field, measured on the wavelength scale, must not be too large. The secular nature of the expansion is apparent, in particular, the nonuniformity associated with the angle. In an appropriate high-$\bar{k}$ regime, the small p approximation to $\Omega_{\mathrm{H}}(p,q)$ provides the dominant contribution to the wave operator in the narrow-angle region. In retaining p^2 and $K^2(q)-1$ terms only to first order, this approximation does not explicitly address the ordering question in the definition of the square root wave operator. The expansion is inherently a multiscale expansion; if the $K^2(q)$ variations occur on the order of a characteristic length l, then the Δ expansion implicitly assumes that $1/\bar{k}^2l^2$ is $\mathscr{O}(1)$ on the Δ scale. Further, the expansion corresponding to the homogeneous medium limit is readily apparent.

For the limiting case of narrow angle, arbitrary inhomogeneity, and weak gradient, Eq. (4.12) to $\mathscr{O}(1)$ takes the form

$$K^2(q)=\lim_{\substack{\eta\to p\\ y\to q}}\cos[(i/2\bar{k})(\partial_\eta\partial_q-\partial_p\partial_y)]\Omega_{\mathrm{H}}^{(0)}(p,q)\Omega_{\mathrm{H}}^{(0)}(\eta,y) \tag{C10}$$

and has the solution

$$\Omega_{\mathrm{H}}^{(0)}(p,q)=K(q), \tag{C11}$$

consistent with the correspondence $g_2(\mathbf{Q})\leftrightarrow g_2(q)$. The $\mathscr{O}(\Delta)$ equation is given by

$$-p^2=\lim_{\substack{\eta\to p\\ y\to q}}\cos[(i/2\bar{k})(\partial_\eta\partial_q-\partial_p\partial_y)](\Omega_{\mathrm{H}}^{(0)}(q)\Omega_{\mathrm{H}}^{(1)}(\eta,y) +\Omega_{\mathrm{H}}^{(0)}(y)\Omega_{\mathrm{H}}^{(1)}(p,q)). \tag{C12}$$

Equation (C12) can be inverted and solved for $\Omega_{\mathrm{H}}^{(1)}(p,q)$. Expressing Eq. (C12) in a form analogous to Eq. (4.11) and applying standard Fourier integral methods leads to the result

$$\Omega_{\mathrm{H}}^{(1)}(p,q)=\frac{\bar{k}}{2\pi}\int\int dt\,d\xi \times\frac{(-t^2)\exp[i\bar{k}\xi(p-t)]}{\Omega_{\mathrm{H}}^{(0)}(q-\xi/2)+\Omega_{\mathrm{H}}^{(0)}(q+\xi/2)}. \tag{C13}$$

Noting the result of Eq. (C11), the integral in Eq. (C13) is readily evaluated to yield

$$\Omega_{\mathrm{H}}^{(1)}(p,q)=-\left(\frac{p^2}{2K(q)}+\frac{1}{8\bar{k}^2}\frac{K''(q)}{K^2(q)}\right), \tag{C14}$$

thus giving to $\mathscr{O}(\Delta)$ the result

$$\Omega_{\mathrm{H}}(p,q)=K(q)-\Delta\left(\frac{p^2}{2K(q)}+\frac{1}{8\bar{k}^2}\frac{K''(q)}{K^2(q)}\right). \tag{C15}$$

Examination of Eq. (C15) shows that the variations in the refractive index field on the wavelength scale must not be too large for the approximation to remain valid. Moreover, the approximation also breaks down when $K(q)$ is sufficiently small. The nonuniformity with respect to the angle is apparent as well as the implicit requirement that $1/\bar{k}^2l^2$ is $\mathscr{O}(1)$ on the Δ scale. The approximation explicitly addresses the operator-ordering question through the summation of $K^2(q)-1$ terms of all orders. Writing $K(q)=\{1+[K^2(q)-1]\}^{1/2}$ and taking $[K^2(q)-1]$ to be $\mathscr{O}(\Delta)$ reduces Eq. (C15) to the ordinary parabolic approximation result. Combining Eqs. (C15) and (4.9) then yields the kernel

$$A_{\mathrm{H}}(q',u) =K((u+q')/2)\delta(u-q')+\Delta\left(\frac{-1}{8\bar{k}^2}\frac{K''((u+q')/2)}{K^2((u+q')/2)} \times\delta(u-q')+\frac{\delta''(u-q')}{2\bar{k}^2K((u+q')/2)}\right), \tag{C16}$$

which then gives (with $\Delta=1$) Eq. (4.17). Equation (4.17), like Eq. (4.16), is valid in an appropriate high-frequency regime. For the limiting case of a homogeneous medium, Eq. (4.17) reduces to

$$(i/\bar{k})\partial_x\phi^+(x,z)+(1/2\bar{k}^2K_0)\partial_z^2\phi^+(x,z)+K_0\phi^+(x,z)=0. \tag{C17}$$

For the limiting case of arbitrary angle, weak inhomogeneity, and weak gradient, Eq. (4.12) to $\mathscr{O}(1)$ takes the form

$$(1-p^2)=\lim_{\substack{\eta\to p\\ y\to q}}\cos[(i/2\bar{k})(\partial_\eta\partial_q-\partial_p\partial_y)] \times\Omega_{\mathrm{H}}^{(0)}(p,q)\Omega_{\mathrm{H}}^{(0)}(\eta,y) \tag{C18}$$

and has the solution

$$\Omega_{\mathrm{H}}^{(0)}(p,q)=(1-p^2)^{1/2}, \tag{C19}$$

consistent with the correspondence $g_1(\mathbf{P})\leftrightarrow g_1(p)$. The $\mathscr{O}(\Delta)$ equation is given by

$$[K^2(q)-1]=\lim_{\substack{\eta\to p\\ y\to q}}\cos[(i/2\bar{k})(\partial_\eta\partial_q-\partial_p\partial_y)] \times[\Omega_{\mathrm{H}}^{(0)}(p)\Omega_{\mathrm{H}}^{(1)}(\eta,y)+\Omega_{\mathrm{H}}^{(0)}(\eta)\Omega_{\mathrm{H}}^{(1)}(p,q)]. \tag{C20}$$

Equation (C20) can be inverted in the same manner as Eq.

(C12) with the result

$$\Omega^{(1)}_{\mathrm{H}}(p,q) = \int dt \frac{\hat{\epsilon}(t)\exp(i\bar{k}qt)}{\Omega^{(0)}_{\mathrm{H}}(p-t/2) + \Omega^{(0)}_{\mathrm{H}}(p+t/2)}. \quad \text{(C21)}$$

Combining Eqs. (C19) and (C21) then gives to $\mathscr{O}(\Delta)$ the result

$$\Omega_{\mathrm{H}}(p,q) = (1-p^2)^{1/2} + \Delta \int dt \times \frac{\hat{\epsilon}(t)\exp(i\bar{k}qt)}{[1-(p-t/2)^2]^{1/2} + [1-(p+t/2)^2]^{1/2}}. \quad \text{(C22)}$$

The nonuniformity with respect to the refractive index field variations associated with the approximation of Eq. (C22) is implicit in the integral correction term. For the limiting case of a homogeneous medium, $\hat{\epsilon}(t) = (K_0^2 - 1)\delta(t)$, and the first two terms in the appropriate expansion readily follow. The reduction to the ordinary parabolic approximation follows upon setting $p = \Delta^{1/2}p$ and $t = \Delta^{1/2}\tau$ in Eq. (C22) to yield

$$\Omega_{\mathrm{H}}(p,q) = (1-\Delta p^2)^{1/2} + \Delta^{3/2}\int d\tau \times \frac{\hat{\epsilon}(\Delta^{1/2}\tau)\exp(i\bar{k}q\Delta^{1/2}\tau)}{[1-\Delta(p-\tau/2)^2]^{1/2} + [1-\Delta(p+\tau/2)^2]^{1/2}}. \quad \text{(C23)}$$

For field strength variations such that $\bar{k}l\Delta^{1/2} \gg 1$, the major contribution to the integral in Eq. (C23) comes from the neighborhood of $\tau = 0$. Thus,

$$\Omega_{\mathrm{H}}(p,q) \approx 1 - \frac{\Delta}{2}p^2 + \frac{\Delta^{3/2}}{2(1-\Delta p^2)^{1/2}} \times \int d\tau\, \hat{\epsilon}(\Delta^{1/2}\tau)\exp(i\bar{k}q\Delta^{1/2}\tau), \quad \text{(C24)}$$

which to first order in Δ is

$$\Omega_{\mathrm{H}}(p,q) = 1 + \tfrac{1}{2}\Delta\{[K^2(q) - 1] - p^2\}, \quad \text{(C25)}$$

the ordinary parabolic result. The approximation of Eq. (C22) explicitly addresses the operator-ordering question through the summation of p^2 terms to all orders.

Combining Eqs. (C22) and (4.9) then gives the kernel

$$A_{\mathrm{H}}(q',u) = \frac{\bar{k}}{2\pi}\int dp\,(1-p^2)^{1/2}\exp[i\bar{k}p(q'-u)] + \Delta\int dq\,\epsilon(q)\hat{R}(u-q,q-q'), \quad \text{(C26)}$$

where $\hat{R}(\alpha,\beta)$ is given by Eq. (4.19), and finally (with $\Delta = 1$) Eq. (4.18). For the limiting case of a homogeneous medium, Eq. (4.18) reduces to

$$\frac{i}{\bar{k}}\partial_x\phi^+(x,z) + \int dz'\left(\frac{1}{2|z-z'|}H_1^{(1)}(\bar{k}|z-z'|) + \frac{\bar{k}}{4}(K_0^2-1)H_0^{(1)}(\bar{k}|z-z'|)\right)\phi^+(x,z') = 0, \quad \text{(C27)}$$

where $H_0^{(1)}(\rho)$ is the zeroth-order Hankel function of the first kind. Equation (4.20) results upon Fourier transformation of Eq. (4.18).

[1]F. D. Tappert, "The parabolic approximation method," in *Wave Propagation in Underwater Acoustics*, Lecture Notes in Physics No. 70, edited by J. B. Keller and J. S. Papadakis (Springer-Verlag, New York, 1977), p. 224. This article contains a brief historical survey of the parabolic approximation method referencing numerous applications made to wide-ranging areas of physics.

[2]V. A. Fock, "The field of a plane wave near the surface of a conducting body," J. Phys. U.S.S.R. **10**, 399 (1946).

[3]M. A. Leontovich and V. A. Fock, "Solution of the problem of propagation of electromagnetic waves along the earth's surface by the method of parabolic equation," J. Phys. U.S.S.R. **10**, 13 (1946).

[4]H. Bremmer, "The W.K.B. approximation as the first term of a geometrical–optical series," Commun. Pure Appl. Math. **4**, 105 (1951).

[5]J. Corones, "Bremmer series that correct parabolic approximations," J. Math. Anal. Appl. **50**, 361 (1975).

[6]J. Corones and R. J. Krueger, "Higher order parabolic approximations to time-independent wave equations," J. Math. Phys. **24**, 2301 (1983).

[7]V. I. Klyatskin and V. I. Tatarskii, Zh. Eksp. Teor. Fiz. **58**, 624 (1970) ["The parabolic equation approximation for propagation of waves in a medium with random inhomogeneities," Sov. Phys. JETP **31**, 335 (1970)].

[8]P.-L. Chow, "Application of function space integrals to problems in wave propagation in random media," J. Math. Phys. **13**, 1224 (1972).

[9]P.-L. Chow, "Perturbation methods in stochastic wave propagation," SIAM Rev. **17**(1), 57 (1975).

[10]P.-L. Chow, "A functional phase-integral method and applications to the laser beam propagation in random media," J. Stat. Phys. **12**(2), 93 (1975).

[11]C. M. Rose and I. M. Besieris, "*N*th-order multifrequency coherence functions: A functional path integral approach," J. Math. Phys. **20**(7), 1530 (1979).

[12]C. M. Rose and I. M. Besieris, "*N*th-order multifrequency coherence functions: A functional path integral approach. II.," J. Math. Phys. **21**(8), 2114 (1980).

[13]J. F. Claerbout, "Course grid calculations of waves in inhomogeneous media with applications to delineation of complicated seismic structure," Geophysics **35**, 407 (1970).

[14]J. F. Claerbout, "Toward a unified theory of reflector mapping," Geophysics **36**, 467 (1971).

[15]T. Landers and J. F. Claerbout, "Numerical calculations of elastic waves in laterally inhomogeneous media," J. Geophys. Res. **77**, 1476 (1972).

[16]J. A. Hudson, "A parabolic approximation for elastic waves," Wave Motion **2**, 207 (1980).

[17]J. J. McCoy, "A parabolic theory of stress wave propagation through inhomogeneous linearly elastic solids," J. Appl. Mech. **44**, 462 (1977).

[18]J. Corones, B. DeFacio, and R. J. Krueger, "Parabolic approximations to the time-independent elastic wave equation," J. Math. Phys. **23**(4), 577 (1982).

[19]S. T. McDaniel, "Propagation of normal modes in the parabolic approximation," J. Acoust. Soc. Am. **57**, 307 (1975).

[20]S. T. McDaniel, "Parabolic approximation for underwater sound propagation," J. Acoust. Soc. Am. **58**, 1178 (1975).

[21]D. R. Palmer, "Eikonal approximation and the parabolic equation," J. Acoust. Soc. Am. **60**, 343 (1976).

[22]D. R. Palmer, "An Introduction to the Application of Feynman Path Integrals to Sound Propagation in the Ocean," Naval Research Laboratory Report No. 8148, 1978.

[23]D. R. Palmer, "A path-integral approach to the parabolic approximation. I," J. Acoust. Soc. Am. **66**(3), 862 (1979).

[24]J. A. DeSanto, "Connection between the solutions of the Helmholtz and parabolic equations for sound propagation," in *Proceedings of the Oceanic Acoustic Modelling Conference*, No. 17, edited by W. Bachmann and R. B. Williams (STI, Saclanteen, 1975), p. 43–1.

[25]J. A. DeSanto, "Relation between the solutions of the Helmholtz and parabolic equations for sound propagation," J. Acoust. Soc. Am. **62**(2), 295 (1977).

[26]J. A. DeSanto, J. S. Perkins, and R. N. Baer, "A correction to the parabolic approximation," J. Acoust. Soc. Am. **64**(6), 1664 (1978).

[27]H. K. Brock, R. N. Buchal, and C. W. Spofford, "Modifying the sound speed profile to improve the accuracy of the parabolic-equation technique," J. Acoust. Soc. Am. **62**(3), 543 (1977).

[28]L. B. Dozier and C. W. Spofford, "Extensions of the parabolic equation model for high-angle bottom-interacting paths," Report SAI-78-712-WA, Science Applications, Inc., December 1977.

[29]S. M. Flatté, R. F. Dashen, W. H. Munk, K. M. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge U. P., Cambridge, 1978).

[30]R. Dashen, "Path integrals for waves in random media," J. Math. Phys. **20**(5), 894 (1979).

[31]M. J. Beran and J. J. McCoy, "Propagation through an anisotropic random medium," J. Math. Phys. **15**, 1901 (1974).

[32]M. J. Beran and J. J. McCoy, "Propagation through an anisotropic random medium—An integro-differential formulation," J. Math. Phys. **17**, 1186 (1976).
[33]J. J. McCoy and M. J. Beran, "Propagation of beamed signals through inhomogeneous media: A diffraction theory," J. Acoust. Soc. Am. **59**, 1142 (1976).
[34]J. J. McCoy and M. J. Beran, "Directional spectral spreading in randomly inhomogeneous media," J. Acoust. Soc. Am. **66**, 1468 (1979).
[35]I. M. Besieris and F. D. Tappert, "Kinetic equation for the quantized motion of a particle in a randomly perturbed potential field," J. Math. Phys. **14**, 1829 (1973).
[36]I. M. Besieris and F. D. Tappert, "Stochastic wave-kinetic theory in the Liouville approximation," J. Math. Phys. **17**, 734 (1976).
[37]H. L. Wilson and F. D. Tappert, "Acoustic propagation in random oceans using the transport equation," Report SAI-78-639-LJ, Science Applications, Inc., April 1978.
[38]F. D. Tappert and R. H. Hardin, "Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations," SIAM Rev. **15**, 423(A) (1973).
[39]H. K. Brock, unpublished work, as reported in Ref. 28.
[40]R. M. Fitzgerald, "Helmholtz equation as an initial value problem with application to acoustic propagation," J. Acoust. Soc. Am. **57**, 839 (1975).
[41]L. E. Estes and G. Fain, "Numerical technique for computing the wide-angle acoustic field in an ocean with range-dependent velocity profiles," J. Acoust. Soc. Am. **62**(1), 38 (1977).
[42]F. R. DiNapoli and R. L. Deavenport, "Numerical models of underwater acoustic propagation," in *Ocean Acoustics*, edited by J. A. DeSanto (Springer-Verlag, New York, 1979), p. 79.
[43]L. Fishman and J. J. McCoy, "Derivation and application of extended parabolic wave theories. II. Path integral representations," J. Math. Phys., (following paper).
[44]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).
[45]I. M. Gel'fand and G. E. Shilov, *Generalized Functions. Vol. 1. Properties and Operations* (Academic, New York, 1964).
[46]M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis* (Academic, New York, 1972).
[47]L. Cohen, "Generalized phase-space distribution functions," J. Math. Phys. **7**, 781 (1966).
[48]B. S. Agarwal and E. Wolf, "Calculus for functions of noncommuting operators and general phase-space methods in quantum mechanics. I. Mapping theorems and ordering of functions of noncommuting operators," Phys. Rev. D **2**(10), 2161 (1970).
[49]M. M. Mizrahi, "Correspondence rules and path integrals," in *Feynman Path Integrals*, Lecture Notes in Physics No. 106, edited by S. A. Albeverio *et al.* (Springer-Verlag, New York, 1979), p. 234.
[50]J. S. Dowker, "Path integrals and ordering rules," J. Math. Phys. **17**(10), 1873 (1976).
[51]W. Garczynski, "Dependence of the Feynman path integral on discretization: the case of a spinless particle in an external electromagnetic field," in *Functional Integration: Theory and Applications*, edited by J.-P. Antoine and E. Tirapegui (Plenum, New York, 1980), p. 175.
[52]F. Langouche, D. Roekaerts, and E. Tirapegui, "Discretization problems of functional integrals in phase space," Phys. Rev. D **20**(2), 419 (1979).
[53]L. Hörmander, "The Weyl calculus of pseudo-differential operators," Comm. Pure Appl. Math. **32**, 359 (1979).
[54]M. R. Hoare, "The linear gas," Adv. Chem. Phys. **20**, 135 (1971).
[55]L. Fishman, "Mathematical Methods in Quantum and Statistical Mechanics," Ph.D. thesis, Harvard University, 1977.
[56]C. DeWitt-Morette, A. Maheshwari, and B. Nelson, "Path integration in nonrelativistic quantum mechanics," Phys. Rep. **50**(5), (March 1979).
[57]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., Cambridge, 1966).
[58]J. R. Taylor, *Scattering Theory: The Quantum Theory on Nonrelativistic Collisions* (Wiley, New York, 1972).
[59]R. R. Greene, "The rational approximation to the acoustic wave equation with bottom interaction," J. Acoust. Soc. Am. (to be published).
[60]L. Fishman and J. J. McCoy, "Direct and inverse wave propagation theories and the factorized Helmholtz equation" (1982) (unpublished).
[61]D. C. Stickler and P. A. Deift, "Inverse problem for a stratified ocean and bottom," J. Acoust. Soc. Am. **70**(6), 1723 (1981).
[62]P. Deift and E. Trubowitz, "Inverse scattering on the line," Comm. Pure Appl. Math. **32**, 121 (1979).
[63]I. M. Gel'fand and B. M. Levitan, "On the determination of a differential equation from its spectral function," Izv. Akad. Nauk SSSR **15**, 309 (1951) [Am. Math. Soc. Transl. **1**, 253 (1956)].
[64]M. E. Taylor, *Pseudodifferential Operators* (Princeton U. P., Princeton, NJ, 1981).
[65]F. Treves, *Introduction to Pseudodifferential and Fourier Integral Operators. Vol. 1. Pseudodifferential Operators* (Plenum, New York, 1980).
[66]F. Treves, *Introduction to Pseudodifferential and Fourier Integral Operators. Vol. 2. Fourier Integral Operators* (Plenum, New York, 1980).
[67]References to significant works in the theory of pseudodifferential operators can be found in two recent book reviews: R. Beals, Bull. Am. Math. Soc. **3**(3), 1069 (1980); M. E. Taylor, *ibid.*, **5**(1), 73 (1981).
[68]R. Bellman and G. M. Wing, *An Introduction to Invariant Imbedding* (Wiley, New York, 1975).
[69]M. E. Davison, "A general approach to splitting and invariant imbedding for linear wave equations," Iowa State University, Ames, Iowa, 1982 (preprint).
[70]S. Frankenthal, M. J. Beran, and A. Whitman, "Caustic corrections using coherence theory," J. Acoust. Soc. Am. **71**, 348 (1982).
[71]S. Coen, "On the elastic profiles of a layered medium from reflection data. I. Plane-wave sources," J. Acoust. Soc. Am. **70**(1), 172 (1981).
[72]S. Coen, "On the elastic profiles of a layered medium from reflection data. II. Impulsive point source," J. Acoust. Soc. Am. **70**(5), 1473 (1981).

Derivation and application of extended parabolic wave theories. II. Path integral representations

Louis Fishman and John J. McCoy
The Catholic University of America, Washington, D. C. 20064

(Received 14 December 1982; accepted for publication 24 June 1983)

The n-dimensional reduced scalar Helmholtz equation for a transversely inhomogeneous medium is naturally related to parabolic propagation models through (1) the n-dimensional extended parabolic (Weyl pseudodifferential) equation and (2) an imbedding in an $(n+1)$-dimensional parabolic (Schrödinger) equation. The first relationship provides the basis for the parabolic-based Hamiltonian phase space path integral representation of the half-space propagator. The second relationship provides the basis for the elliptic-based path integral representations associated with Feynman and Fradkin, Feynman and Garrod, and Feynman and DeWitt-Morette. Exact and approximate path integral constructions are derived for the homogeneous and transversely inhomogeneous cases corresponding to both narrow- and wide-angle extended parabolic wave theories. The path integrals allow for a global perspective of the transition from elliptic to parabolic wave theory in addition to providing a unifying framework for dynamical approximations, resolution of the square root operator, and the concept of an underlying stochastic process.

PACS numbers: 03.40.Kf, 02.30.Jr, 92.10.Vz, 43.25. − x

I. INTRODUCTION

The reduced scalar Helmholtz equation for a transversely inhomogeneous half-space supplemented with an outgoing radiation condition and an appropriate boundary condition on the initial-value plane defines a direct acoustic propagation model. This elliptic formulation admits a factorization and is subsequently equivalent to a formal first-order forward-propagating wave equation,

$$(i/\bar{k})\partial_x \phi^+(x,\mathbf{x}_\perp) + [K^2(\mathbf{x}_\perp) + (1/\bar{k}^2)\nabla_\perp^2]^{1/2}\phi^+(x,\mathbf{x}_\perp) = 0, \tag{1.1}$$

which is recognized as an extended parabolic propagation model with respect to a distinguished global principal propagation, or range, direction x. In Eq. (1.1) $K(\mathbf{x}_\perp)$ is a dimensionless sound speed profile or refractive index field, $\bar{k}$ is an appropriate average or reference wavenumber, and $\{x_\perp\}$ denotes the perpendicular, or cross-range, coordinates. In a previous paper[1] it was established that the formal wave equation, Eq. (1.1), can be explicitly written (in an obvious two-dimensional notation) as

$$\frac{i}{\bar{k}}\partial_x \phi^+(x,z) + \int dz' \left\{ \frac{\bar{k}}{2\pi}\int dp\, \Omega_{\mathbf{H}}\left(p,\frac{z+z'}{2}\right) \times \exp[i\bar{k}p(z-z')]\right\}\phi^+(x,z') = 0. \tag{1.2}$$

Here the "symbol" $\Omega_{\mathbf{H}}(p,q)$ associated with the square root operator $\mathbf{H} = [K^2(q) + (1/\bar{k}^2)\partial_q^2]^{1/2}$ satisfies

$$\Omega_{\mathbf{H}^2}(p,q) = K^2(q) - p^2 = \left(\frac{\bar{k}}{\pi}\right)^2 \iiiint dt\, dx\, dy\, dz\, \Omega_{\mathbf{H}}(t+p,x+q) \times \Omega_{\mathbf{H}}(y+p,z+q)\exp[2i\bar{k}(xy-tz)], \tag{1.3}$$

$\Omega_{\mathbf{H}^2}(p,q)$ being the "symbol" associated with the square of $\mathbf{H}$, $\mathbf{H}^2 = [K^2(q) + (1/\bar{k}^2)\partial_q^2]$. Equation (1.2) is a first-order Weyl pseudodifferential equation; Eq. (1.3) is recognized as the composition equation in the Weyl pseudodifferential operator calculus and embodies the definition of an operator square root in terms of its square. Perturbation treatments of Eq. (1.3) result in a systematic development of approximate wave equation representations of the propagation theory.

The wave equation form of the extended parabolic wave theory, however, does not provide the only representation. Path integrals can provide a qualitative as well as quantitative equivalent representation of the propagation theory. While it is perhaps somewhat ironic that the development of extended parabolic *theories* is then primarily based upon an apparent *solution* representation, it should be emphasized that path integration does not merely provide a solution to a partial differential equation. The construction of the path integrals provides a global perspective of the propagation/scattering experiment and thus a natural means for cumulative error estimates. Furthermore, these representations also provide a stochastic perspective of the phenomenon and subsequently a stochastic interpretation of the transition from a completely deterministic elliptic to a completely deterministic parabolic wave theory. Asymptotic analysis of differential equation formulations does not naturally lead to cumulative error estimates nor to the stochastic interpretation referred to above.

Configuration and phase space path integral representations of the half-space Helmholtz propagator follow from both the parabolic form of Eq. (1.2) and the elliptic form of the Helmholtz equation. The factorization analysis provides the basis for a Hamiltonian phase space representation which can be termed direct (see Sec. II for a precise definition). This construction explicitly addresses the operator-ordering question inherent in the formal square root wave operator and resolved in Eqs. (1.2) and (1.3). Exact direct path integral representations for the forward propagator in a range-independent half-space are constructed for a homogeneous environment. For a transversely inhomogeneous environment, direct path integrals are constructed approximating the exact propagator to within a well-defined perturbative resolution of the wave operator and corresponding to both narrow and wide-angle propagation theories.

The elliptic-based analysis results in an indirect (see Sec. II) path integral representation (Feynman–Fradkin), which is formally exact and well defined. Thus, in principle, this representation contains the resolution of the operator-ordering question. However, this resolution is not transparent in the indirect representation, the principal utility of which is the derivation of approximate solutions with inherently *a posteriori* justifications. Further, two direct path integrals follow from the indirect representation by a series of formal manipulations. As obtained, then, these direct path integral representations are to be accepted as either approximate (Feynman–Garrod) or symbolic (Feynman–DeWitt-Morette); their construction does not explicitly settle the operator-ordering question. The primary significance of the direct path integrals heuristically constructed, then, is in their physical and mathematical suggestiveness.

In conjuction, then, the path integral representations provide a unifying framework for dynamical approximations, resolution of the square root operator, and the concept of an underlying stochastic process.

II. APPLICATION OF PATH INTEGRALS TO THE HELMHOLTZ EQUATION

The n-dimensional reduced scalar Helmholtz equation for a transversely inhomogeneous medium,

$$[\nabla^2 + \bar{k}^2 K^2(\mathbf{x}_\perp)]\,\phi(\mathbf{x}) = 0, \tag{2.1}$$

is naturally related to parabolic propagation models through (1) the n-dimensional extended parabolic (Weyl pseudodifferential) equation and (2) an imbedding in an $(n+1)$-dimensional parabolic (Schrödinger) equation. Relationships (1) and (2) provide the basis for parabolic- and elliptic-based path integral constructions, respectively.

A. Parabolic constructions

Equation (1.1) is in the form of a Schrödinger equation with an arbitrary Hamiltonian operator. Using arguments based upon the Hamiltonian formulation of classical mechanics[2,3] and in a manner analogous to the Feynman construction,[4–6] the half-space Helmholtz propagator can be symbolically represented as a path integral taken over phase space,[7]

$$G^+(x,\mathbf{x}_\perp|0,\mathbf{x}_\perp') = \int_C D(\mathbf{p}_\perp,\mathbf{z}_\perp)\exp\left\{i\bar{k}\int_0^x d\tau\left[\mathbf{p}_\perp\cdot\frac{d\mathbf{z}_\perp}{d\tau} - H_{\rm cl}(\mathbf{p}_\perp,\mathbf{z}_\perp)\right]\right\}. \tag{2.2}$$

In Eq. (2.2), in analogy with the quantum mechanical case, the "classical Hamiltonian" $H_{\rm cl}(\mathbf{p}_\perp,\mathbf{z}_\perp) = -[K^2(\mathbf{z}_\perp) - \mathbf{p}_\perp^2]^{1/2}$ and C is the set of phase space paths $(\mathbf{p}_\perp,\mathbf{z}_\perp)$ such that $\mathbf{z}_\perp(0) = \mathbf{x}_\perp'$ and $\mathbf{z}_\perp(x) = \mathbf{x}_\perp$ with $\mathbf{p}_\perp(\tau)$, the "conjugate momentum," unrestricted.

The Cohen/Agarwal–Wolf construction provides the basis for an algorithmic, or operational, representation of the general Schrödinger propagator, removing the inherent ambiguity in the Feynman formulation by properly accounting for the operator-ordering.[8–10] The operator representation[1] (in a two-dimensional notation)

$$\mathbf{H}(\mathbf{P},\mathbf{Q}) = \int\!\!\int du\, dv\, F(u,v)\hat{h}_{\mathbf{H}}(u,v)\exp[i\bar{k}(v\mathbf{Q} + u\mathbf{P})], \tag{2.3}$$

where

$$\hat{h}_{\mathbf{H}}(u,v) = \left(\frac{\bar{k}}{2\pi}\right)^2 \int\!\!\int dp\, dq\, h_{\mathbf{H}}(p,q)\exp[-i\bar{k}(vq + up)], \tag{2.4}$$

in conjunction with the Markov property of the propagator forms the basis of the path integral construction. This has been considered in detail in a series of papers by Mayes and Dowker.[11–13] For a Schrödinger propagator corresponding to a Hamiltonian operator given by Eq. (2.3), the result can be expressed as

$$G^+(x,z|0,z_s) = \lim_{N\to\infty}\int\prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N}\frac{dp_j^z}{2\pi/\bar{k}}\exp\left\{i\bar{k}\sum_{j=1}^{N}\left[p_j^z(z_j - z_{j-1}) + \frac{x}{N}H(p_j^z,z_j,z_{j-1})\right]\right\}, \tag{2.5}$$

where

$$H(p,q'',q') = \frac{\bar{k}}{2\pi}\int\!\!\int ds\, dt\, F(q'' - q',s)h_{\mathbf{H}}\left(p,\frac{q''+q'}{2} - t\right)\exp(i\bar{k}st). \tag{2.6}$$

More specifically, the path in phase space in Eq. (2.5) is defined by the prescription $z_j = z(j/N)$, $p_j^z = p^z[(j-\frac{1}{2})/N]$, placing p_j^z halfway between z_j and z_{j-1}, with the further constraints $z_0 = z_s$ and $z_N = z$.[10] Further, all integrations in Eqs. (2.3)–(2.6) are taken over the interval $(-\infty,\infty)$.

For the wave propagation problem,

$$\hat{\Omega}_{\mathrm{H}}(u,v) = F(u,v)\hat{h}_{\mathrm{H}}(u,v), \tag{2.7}$$

with $\Omega_{\mathrm{H}}(p,q)$, the inverse Fourier transform of $\hat{\Omega}_{\mathrm{H}}(u,v)$, determined through Eq. (1.3) and the forward (outgoing) wave propagation condition. That the representation for G^+ is not unique is seen with the realization that $H(p,q'',q')$ is not uniquely defined by $\Omega_{\mathrm{H}}(p,q)$, but is in general a function of the chosen transformation function $F(u,v)$. This is as it must be. Equation (2.5) can be viewed as the discretization of a symbolic functional integral. It is well known that different discretizations, corresponding to different $F(u,v)$ functions, lead to different "Hamiltonian" operators and, thus, propagators.[10] This is compensated for by the discretized "effective Hamiltonian" in Eq. (2.6), which is just such as to preserve the total Hamiltonian and result in unique equations of motion. The functional integral is then independent of the discretization as it should be.[12,13] A particular representation corresponds to the choice of the Weyl ordering, $F(u,v) = 1$, which subsequently identifies $\Omega_{\mathrm{H}}(p,q) = h_{\mathrm{H}}(p,q)$ and $H(p,q'',q') = \Omega_{\mathrm{H}}(p,(q''+q')/2)$. Thus, Eq. (2.5) can be written as

$$G^+(x,z|0,z_s) = \lim_{N\to\infty} \int \prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N} \frac{dp_j^z}{2\pi/\bar{k}} \exp\left\{ i\bar{k} \sum_{j=1}^{N} \left[p_j^z(z_j - z_{j-1}) + \frac{x}{N}\Omega_{\mathrm{H}}\left(p_j^z, \frac{z_j + z_{j-1}}{2}\right)\right]\right\}. \tag{2.8}$$

Well-defined perturbative resolutions of $\Omega_{\mathrm{H}}(p,q)$ lead to corresponding approximate path integral representations of the half-space Helmholtz propagator. Extension of Eq. (2.8) to arbitrary dimension follows readily. The path integral representation in Eq. (2.8) is termed "direct" in that it is expressed directly as a functional integral.

B. Elliptic constructions

The elliptic-based constructions proceed from the Green's function for the n-dimensional Helmholtz equation in Cartesian coordinates for an infinite medium with arbitrary inhomogeneity satisfying

$$[\nabla^2 + \bar{k}^2 K^2(\mathbf{x})]G(\mathbf{x},\mathbf{x}_s) = -\delta^{(n)}(\mathbf{x}-\mathbf{x}_s), \tag{2.9}$$

supplemented with an outgoing wave radiation condition. The acoustic radiation problem formulated in Eq. (2.9) can be readily imbedded into an initial-value problem expressed in the form of a parabolic equation in one higher dimension. This just expresses the well-known quantum mechanical Fourier transform relationship between the parabolic (Schrödinger) equation propagator and the Helmholtz (fixed energy) equation Green's function. Thus,

$$G(\mathbf{x},\mathbf{x}_s) = \frac{i}{2\bar{k}} \int_0^\infty d\tau \exp(\tfrac{1}{2}i\bar{k}\tau)\Phi(\tau,\mathbf{x}|0,\mathbf{x}_s), \tag{2.10}$$

where $\Phi(\tau,\mathbf{x}|0,\mathbf{x}_s)$ satisfies the parabolic equation

$$(i/\bar{k})\partial_\tau \Phi(\tau,\mathbf{x}|0,\mathbf{x}_s) + [(1/2\bar{k}^2)\nabla^2 + \tfrac{1}{2}(K^2(\mathbf{x}) - 1)]\Phi(\tau,\mathbf{x}|0,\mathbf{x}_s) = 0 \tag{2.11}$$

with the initial condition

$$\Phi(0,\mathbf{x}|0,\mathbf{x}_s) = \delta^{(n)}(\mathbf{x}-\mathbf{x}_s). \tag{2.12}$$

A configuration space path integral representation follows on inserting the Feynman representation for $\Phi(\tau,\mathbf{x}|0,\mathbf{x}_s)$[4–6] resulting in

$$G(\mathbf{x},\mathbf{x}_s) = \frac{i}{2\bar{k}} \int_0^\infty d\tau \exp(\tfrac{1}{2}i\bar{k}\tau) \int_P D(\mathbf{x}(\sigma)) \exp\left\{ i\bar{k} \int_0^\tau d\sigma \left[\tfrac{1}{2}\left(\frac{d\mathbf{x}(\sigma)}{d\sigma}\right)^2 - V(\mathbf{x}(\sigma))\right]\right\}, \tag{2.13}$$

where P is the set of continuous paths from $(\mathbf{x}_s,0)$ to $(\mathbf{x},\tau)$. A corresponding phase space representation follows from Eqs. (2.2) and (2.11) in the form

$$G(\mathbf{x},\mathbf{x}_s) = \frac{i}{2\bar{k}} \int_0^\infty d\tau \exp(\tfrac{1}{2}i\bar{k}\tau) \int_C D(\mathbf{p}(\sigma),\mathbf{x}(\sigma)) \exp\left(i\bar{k} \int_0^\tau d\sigma \left\{ \mathbf{p}(\sigma)\cdot\frac{d\mathbf{x}(\sigma)}{d\sigma} - \left[\frac{\mathbf{p}^2(\sigma)}{2} + V(\mathbf{x}(\sigma))\right]\right\}\right). \tag{2.14}$$

In Eqs. (2.13) and (2.14) $\bar{k}$ has an infinitesimally small, positive, imaginary part, $\bar{k} \to \bar{k} + i\epsilon$, enforcing the outgoing wave radiation condition. The identification, $V(\mathbf{x}) = -\tfrac{1}{2}[K^2(\mathbf{x}) - 1]$, is further noted. The path integral representations in Eqs. (2.13) and (2.14) are termed Feynman–Fradkin representations.[14,15] These representations are termed "indirect" in that they are expressed through an integration over a path integral as opposed to "directly" as a path integral. Further, the Feynman–Fradkin representations are formally exact.

The construction of a direct phase space path integral representation of $G(\mathbf{x},\mathbf{x}_s)$ proceeds from Eq. (2.14). Introducing the algorithmic representation (the lattice approximation) for the path integral in Eq. (2.14) corresponding to Eq. (2.8) with $\Omega_{\mathrm{H}}(p,q) = -[\tfrac{1}{2}p^2 + V(q)]$ and formally carrying out the τ integration results in[3,16]

$$G(\mathbf{x},\mathbf{x}_s) \doteq \frac{-1}{2\bar{k}^2} \lim_{N\to\infty} \int \prod_{j=1}^{N-1} d\mathbf{x}_j \times \prod_{j=1}^{N} \frac{d\mathbf{p}_j}{(2\pi/\bar{k})^n} \frac{\exp(i\bar{k}S_N)}{\tfrac{1}{2} - \mathscr{E}}, \tag{2.15}$$

where

$$S_N = \sum_{j=1}^{N} [\, \mathbf{p}_j\cdot(\mathbf{x}_j - \mathbf{x}_{j-1})] \tag{2.16}$$

and

$$\mathscr{E} = \frac{1}{N}\left\{\sum_{j=1}^{N}\left[\frac{\mathbf{p}_j^2}{2} + V\left(\frac{\mathbf{x}_j + \mathbf{x}_{j-1}}{2}\right)\right]\right\}. \tag{2.17}$$

S_N corresponds to an appropriate discretized action, and $\mathscr{E}$ plays a role analogous to an average energy. The phase space path in Eq. (2.15) is the n-dimensional generalization of that given for Eq. (2.5). Thus, the order of accomplishing the integrations is prescribed and cannot be arbitrarily interchanged. The paths are not classical paths.[6,17] Schulman has provided a detailed discussion on the limitation attached to applying a literal physical path picture to formal canonical path integral representations.[6] Equation (2.15) is the Feynman–Garrod representation.

The interchange of operations and subsequent integration in the derivation of Eq. (2.15) is clearly heuristic; more importantly, the resulting expression can only approximate the exact result. The formal integration of the lattice approximation in some sense bypasses the detailed consideration of the inherent operator-ordering question. Gutzwiller's formal derivation of the WKB propagator starting from the Feynman–Garrod representation[16] suggests that Eq. (2.15) only addresses the ordering question in some high-frequency limit. To signify this as yet uncharacterized approximate relationship, the symbol $\doteq$ is used in the appropriate equations.

Following Gutzwiller,[16] the phase space integration in Eq. (2.15) can be decomposed into an integration over the variables $\{\mathbf{x}_j, \mathbf{p}_j\}$ on a hypersurface of constant "average energy" [$\mathscr{E}$ as given by Eq. (2.17)] followed by an integration over all values of the "average energy." As $\bar{k} \to \infty$, the stationary path $\mathbf{x}_s = \mathbf{x}_0, \mathbf{p}_1, \mathbf{x}_1, \ldots, \mathbf{p}_N, \mathbf{x}_N = \mathbf{x}$ for S_N [Eq. (2.16)] must be determined under the subsidiary constraint of the fixed "average energy" condition expressed by the denominator in Eq. (2.15). In the $N \to \infty$, or functional, limit the isoperimetric problem is thus to determine paths $\mathbf{x}(\alpha)$, $\mathbf{p}(\alpha)$ in phase space for which $S = \int_0^1 d\alpha\, \mathbf{p}(\alpha) d\mathbf{x}(\alpha)/d\alpha$ is stationary, given the end points at $\mathbf{x}_s$ and $\mathbf{x}$, as well as the "average energy." The Euler equations for this problem are the usual Hamiltonian equations of motion. However, the usual Hamiltonian variational principle demands the far more restrictive condition that the trajectories always remain on the energy surface. Garrod has established that the usual constant energy Hamiltonian variational principle is unnecessarily restrictive and can, in fact, be extended to include trajectories with a fixed average energy as suggested by the path integral representation.[3] In conjunction with the Jacobi form of the least action principle,[18] which reformulates the principle in terms of the element of path length, the extended variational principle suggests a formal direct path integral representation in configuration space.

Thus, following Feynman,[4] Davies,[2] and Garrod,[3] the path integral formally results from the variational principle upon exponentiating the Jacobi form of the action analog and summing over all paths from $\mathbf{x}_s$ to $\mathbf{x}$ with a fixed "average energy." Specifically, this takes the form[19]

$$G(\mathbf{x},\mathbf{x}_s) = \frac{-1}{2\bar{k}^2}\int_E D(\mathbf{x}')\exp[i\bar{k}\overline{W}(\mathbf{x}')], \tag{2.18}$$

where

$$\overline{W} = \int_{\mathbf{x}_s}^{\mathbf{x}} \|d\mathbf{x}'\|\,[1 - 2V(\mathbf{x}')]^{1/2} \tag{2.19}$$

is the analog of the action associated with a "free particle" on a space with metric

$$dl^2 = [1 - 2V(\mathbf{x}')]\,\|d\mathbf{x}'\|^2 \tag{2.20}$$

and where E represents the space of paths from $\mathbf{x}_s$ to $\mathbf{x}$ such that

$$\frac{1}{2} = \frac{1}{\tau}\int_0^\tau dt\left[\frac{1}{2}\left\|\frac{d\mathbf{x}'(t)}{dt}\right\|^2 + V(\mathbf{x}'(t))\right] \tag{2.21}$$

with the constraints

$$\begin{aligned}\mathbf{x}'(0) &= \mathbf{x}_s,\\ \mathbf{x}'(\tau) &= \mathbf{x}.\end{aligned} \tag{2.22}$$

The conditions expressed in Eqs. (2.21) and (2.22) are analogous to a condition of fixed average energy along the path in quantum mechanics. The expression in Eq. (2.18) is termed the Feynman–DeWitt-Morette representation; it is a symbolic representation. A formal analysis of Eq. (2.18) in the $\bar{k} \to \infty$ limit indeed yields the correct WKB result. Thus, for the parabolic equation, the path space is constructed from paths going from $\mathbf{x}_s$ to $\mathbf{x}$ in a fixed time τ, while for the Helmholtz equation the path space is constructed from paths going from $\mathbf{x}_s$ to $\mathbf{x}$ with a fixed "average energy." This is suggested by the construction of the Feynman–Garrod representation; in particular, from the form of the "energy denominator" in Eq. (2.15).

The Feynman–DeWitt-Morette representation formally effects the τ integration over both the measure and the exponentiated action functional in the Feynman–Fradkin representation. Furthermore, the representation is explicit in its underlying stochastic foundations. The stochastic process with characteristics given by Eqs. (2.21) and (2.22) with τ taken as a stochastic variable plays a role analogous to the one that Brownian motion plays with respect to the stochastic foundations of the parabolic equation.[19,20] This is true in a heuristic sense, however, since the process in question has not been studied in detail,[19] and, consequently, it remains an open question as to whether it can provide a rigorous mathematical basis for the Helmholtz path integral in the manner that Brownian motion provides for the parabolic path integral.

For environments which are range-independent in that $V(\mathbf{x})$ is independent of one or more of the Cartesian components of $\mathbf{x}$, reduced path integral representations can be constructed. The appropriate homogeneous medium propagators factor from the Feynman path integral in the Feynman–Fradkin representations resulting in lower spatial-dimensional path integrals. The Feynman–DeWitt-Morette representation also symbolically reduces in a straightforward fashion. For the Feynman–Garrod representation a partial integration over the phase space effects the reduction. This is a heuristic operation since in a strict mathematical sense the $\mathbf{p}$-space and $\mathbf{x}$-space integrations cannot be performed independently of one another—even in the $V(\mathbf{x}) = 0$ case.[7,8,19] It is, in fact, equivalent to formally integrating the Feynman–Fradkin representation at the level of the lattice approximation, taking explicit account of the reduced form of the Feynman path integral. Thus, the resulting propagators are approximate in the sense previously discussed.

For the two-dimensional case, $n = 2$, for an x- or range-independent potential, $V(x,z) = V(z)$, the partial integrations over the x and p^x variables in Eq. (2.15) proceed in a straightforward manner (the details are given in the Appendix), leading to a reduced representation for the Green's function in the form

$$G(x,z|x_s,z_s) \doteq \frac{i}{2^{3/2}\bar{k}} \lim_{N\to\infty} \int \prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N} \frac{dp_j^z}{2\pi/\bar{k}} \times \frac{\exp\{i\bar{k}\,[S_N + 2^{1/2}(\tfrac{1}{2} - \mathscr{E})^{1/2}|x - x_s|\,]\}}{(\tfrac{1}{2} - \mathscr{E})^{1/2}}. \tag{2.23}$$

Here S_N and $\mathscr{E}$ are the appropriate one-dimensional (in z and p^z) action and average energy analogs of Eqs. (2.16) and (2.17) with the corresponding path boundary conditions given by $z_0 = z_s$ and $z_N = z$.

For the three-dimensional case, $n = 3$, two cases are of immediate interest. For an x-independent potential, $V(x, y,z) = V(y,z)$, corresponding to a distinguished direction taken along the x axis, the resulting reduced representation is of the same form as Eq. (2.23). Specifically, the Liouville measure, action, and average energy analogs take on their corresponding two-dimensional forms in the variables z, y, p^z, and p^y. For a potential that is also independent of one cross-range coordinate, $V(x,y,z) = V(z)$, a further reduction is possible following the same procedure as outlined in the Appendix. The result of this further reduction is written as

$$G(x, y,z|x_s, y_s,z_s) \doteq \frac{i}{4} \lim_{N\to\infty} \int \prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N} \frac{dp_j^z}{2\pi/\bar{k}} \exp(i\bar{k}S_N) \times H_0^{(1)}(2^{1/2}\bar{k}(\tfrac{1}{2} - \mathscr{E})^{1/2}R_2), \tag{2.24}$$

where again S_N and $\mathscr{E}$ take on their one-dimensional forms, $z_0 = z_s$ and $z_N = z$, and $R_2 = [(x - x_s)^2 + (y - y_s)^2]^{1/2}$ is the appropriate range coordinate. Taking the source location along the z axis and writing $r = (x^2 + y^2)^{1/2}$ in Eq. (2.24), the representation then corresponds to the case of cylindrical symmetry with a range (r)-independent potential,

$$G(r,z|0,z_s) \doteq \frac{i}{4} \lim_{N\to\infty} \int \prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N} \frac{dp_j^z}{2\pi/\bar{k}} \exp(i\bar{k}S_N) \times H_0^{(1)}(2^{1/2}\bar{k}(\tfrac{1}{2} - \mathscr{E})^{1/2}r). \tag{2.25}$$

Equation (2.25) can also be derived starting from DeSanto's integral representation,[21] which for the range-independent environment reduces to the Feynman–Fradkin representation upon the introduction of a Feynman path integral representation for the parabolic equation solution.

The reflection principle, or method of images, relates the half-space Helmholtz propagator G^+ and the full-space Helmholtz Green's function G for the symmetric extension through

$$G^+(x,\mathbf{x}_\perp|0,\mathbf{x}_\perp') = -2\partial_x\, G(x,\mathbf{x}_\perp|0,\mathbf{x}_\perp'). \tag{2.26}$$

Equation (2.26) in conjunction with the representations of Feynman and Fradkin [Eqs. (2.13) and (2.14)], Feynman and Garrod [Eq. (2.15)], and Feynman and DeWitt-Morette [Eq. (2.18)] for the symmetrized environment results in path integral representations for the arbitrary half-space problem. For a transversely inhomogeneous half-space, the reduced representations allow for the appropriate constructions. For example, in the two-dimensional case, the Feynman–Garrod representation follows immediately from Eqs. (2.23) and (2.26) as

$$G^+(x,z|0,z_s) \doteq \lim_{N\to\infty} \int \prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N} \frac{dp_j^z}{2\pi/\bar{k}} \times \exp\{i\bar{k}\,[S_N + 2^{1/2}(\tfrac{1}{2} - \mathscr{E})^{1/2}x\,]\}. \tag{2.27}$$

The representation expressed in Eq. (2.27) can also be derived starting from DeSanto's integral relationship for the waveguide problem.[21] A representation analogous to Eq. (2.27) corresponds to the propagator for the half-space boundary-value problem in its three-dimensional formulation.

III. HOMOGENEOUS HALF-SPACE

For the homogeneous half-space, $K^2(\mathbf{x}_\perp) = K_0^2$, the Hamiltonian phase space representation of the propagator follows from Eq. (2.8) on noting the exact result[1] $\Omega_{\mathrm{H}}(p,q) = (K_0^2 - p^2)^{1/2}$. The propagator takes the form

$$G^+(x,z|0,z_s) = \lim_{N\to\infty} \int \prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N} \frac{dp_j^z}{2\pi/\bar{k}} \exp\Big(i\bar{k} \sum_{j=1}^{N} \Big\{p_j^z(z_j - z_{j-1}) + \frac{x}{N}\,[(1 - 2V_0) - (p_j^z)^2]^{1/2}\Big\}\Big), \tag{3.1}$$

where $V_0 = -\frac{1}{2}(K_0^2 - 1)$ as before. The Feynman–Garrod representation follows from Eq. (2.27) in the form

$$G^+(x,z|0,z_s) = \lim_{N\to\infty} \int \prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N} \frac{dp_j^z}{2\pi/\bar{k}} \exp\Big\{i\bar{k} \sum_{j=1}^{N} [\,p_j^z(z_j - z_{j-1})]\Big\} \exp\Big(i\bar{k}x\Big\{1 - \frac{1}{N}\sum_{j=1}^{N} [(p_j^z)^2 + 2V_0]\Big\}^{1/2}\Big). \tag{3.2}$$

The equivalence of the two path integral representations can be directly demonstrated by integrating over the $\{z_j\}$ coordinates in Eqs. (3.1) and (3.2) in the manner of the Appendix. The resulting expressions are written as

$$G^{+}(x,z|0,z_s) = \lim_{N\to\infty} \frac{\bar{k}}{2\pi} \int dp_N^z \int \prod_{j=1}^{N-1} dp_j^z \exp[i\bar{k}(p_N^z z - p_1^z z_s)]$$
$$\times \exp\left\{i\bar{k}\frac{x}{N}\sum_{j=1}^{N}[(1-2V_0)-(p_j^z)^2]^{1/2}\right\}\delta(p_1^z - p_2^z)\delta(p_2^z - p_3^z)\cdots$$
$$\times\delta(p_{N-1}^z - p_N^z) \tag{3.3}$$

and

$$G^{+}(x,z|0,z_s) = \lim_{N\to\infty} \frac{\bar{k}}{2\pi} \int dp_N^z \int \prod_{j=1}^{N-1} dp_j^z \exp[i\bar{k}(p_N^z z - p_1^z z_s)]$$
$$\times \exp\left(i\bar{k}x\left\{1 - \frac{1}{N}\sum_{j=1}^{N}[(p_j^z)^2 + 2V_0]\right\}^{1/2}\right)\delta(p_1^z - p_2^z)\delta(p_2^z - p_3^z)\cdots$$
$$\times\delta(p_{N-1}^z - p_N^z), \tag{3.4}$$

which are clearly equivalent. It is thus seen that, despite its heuristic derivation, Eq. (3.2) is operationally correct for the homogeneous medium case. This is not surprising due to the absence of the operator-ordering question.

Formally defining a weighted averaging over the appropriate function space by

$$\langle J(p(\alpha))\rangle \equiv \int_C D(p(\alpha),z(\alpha))\exp\left\{i\bar{k}\int_0^1 d\alpha\left[p(\alpha)\frac{dz(\alpha)}{d\alpha}\right]\right\} J(p(\alpha)), \tag{3.5}$$

supplemented with the normalization condition

$$\langle 1\rangle = \delta(z(1) - z(0)) = \delta(z - z_s), \tag{3.6}$$

the path integral equivalence can then be expressed as

$$\langle\exp\{i\bar{k}x\,\overline{[-H(p(\alpha))]}\}\rangle = \langle\exp(i\bar{k}x\{\overline{[-H(p(\alpha))]^2}\}^{1/2})\rangle. \tag{3.7}$$

The "effective Hamiltonian" in Eq. (3.7) is given by $H(p(\alpha)) = -\{K_0^2 - [p(\alpha)]^2\}^{1/2}$ and $\overline{f(\alpha)}$ denotes the average of $f(\alpha)$ on the interval [0,1]. Thus, the equivalence is in equating the functional averages of the average "effective Hamiltonian" and the root mean square "effective Hamiltonian."

Completing the integrations in Eqs. (3.3) and (3.4) leads to the result

$$G^{+}(x,z|0,z_s) = (i\bar{k}K_0 x/2R_2)\, H_1^{(1)}(\bar{k}K_0R_2), \tag{3.8}$$

where $R_2 = [x^2 + (z - z_s)^2]^{1/2}$ and $H_1^{(1)}(\eta)$ is the appropriate Hankel function. The three-dimensional case follows in an analogous fashion, leading to the final result

$$G^{+}(x,y,z|0,y_s,z_s) = (i\bar{k}K_0x/2\pi R_3^2)\exp(i\bar{k}K_0R_3)(-1 + 1/i\bar{k}K_0R_3), \tag{3.9}$$

where $R_3 = [x^2 + (y - y_s)^2 + (z - z_s)^2]^{1/2}$.

Configuration space path integral representations follow readily from both the parabolic- and elliptic-based phase space constructions. Performing the $\{p_j^z\}$ integrations in Eq. (3.1) with $V_0 = 0$ taken and noting the result in Eq. (3.8) leads to

$$G^{+}(x,z|0,z_s) = \lim_{N\to\infty}\int\prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N}\left\{\left(\frac{i\bar{k}}{2\{1 + [(z_j - z_{j-1})/\delta]^2\}^{1/2}}\right)[H_1^{(1)}(\bar{k}\delta\{1 + [(z_j - z_{j-1})/\delta]^2\}^{1/2})]\right\}, \tag{3.10}$$

where $\delta = x/N$. The corresponding three-dimensional result follows in an analogous fashion. Equation (3.10) is just an explicit statement of the composition law associated with the parabolic equation,

$$G^{+}(x,z|0,z_s) = \lim_{N\to\infty}\int\prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N} G^{+}(\delta,z_j|0,z_{j-1}), \tag{3.11}$$

and is termed a path sum.[22]

For the elliptic-based construction it is useful to start from Eq. (2.15) written in the form

$$G_n(\mathbf{x},\mathbf{x}_s) = G_n(\mathbf{x} - \mathbf{x}_s) = -\lim_{N\to\infty}\int\prod_{j=1}^{N-1} d\mathbf{x}_j \prod_{j=1}^{N}\frac{d\mathbf{p}_j}{(2\pi)^n}\,\frac{\exp\left\{i\sum_{j=1}^{N}[\mathbf{p}_j\cdot(\mathbf{x}_j - \mathbf{x}_{j-1})]\right\}}{\bar{k}^2 - (1/N)\sum_{j=1}^{N}(\mathbf{p}_j)^2}, \tag{3.12}$$

where the subscript n has been affixed to emphasize the spatial dimensionality and $V_0 = 0$. The $\mathbf{p}$-space integral in Eq. (3.12) is recognized as the inverse Fourier representation of the Helmholtz Green's function in nN dimensions; thus

$$G_n(\mathbf{x} - \mathbf{x}_s) = \lim_{N\to\infty}\int\prod_{j=1}^{N-1} d\mathbf{x}_j\,[N^{nN/2}G_{nN}(N^{1/2}\mathbf{\Delta})], \tag{3.13}$$

where $\mathbf{\Delta}$ is an nN-dimensional vector $(x^{(1)} - x_{N-1}^{(1)}, x^{(2)} - x_{N-1}^{(2)}, \ldots, x^{(n)} - x_{N-1}^{(n)}, x_{N-1}^{(1)} - x_{N-2}^{(1)}, x_{N-1}^{(2)} - x_{N-2}^{(2)}, \ldots, x_{N-1}^{(n)} - x_{N-2}^{(n)}, \ldots, x_1^{(1)} - x_s^{(1)}, x_1^{(2)} - x_s^{(2)}, \ldots, x_1^{(n)} - x_s^{(n)})$ with $x^{(i)}$ denoting the ith component of the n-dimensional vector $\mathbf{x}$. The well-known result for the Helmholtz Green's function of arbitrary dimension,[23]

$$G_{nN}(N^{1/2}\Delta) = \tfrac{1}{4}i(\bar{k}/2\pi R_{nN})^{(nN-2)/2} H^{(1)}_{(nN-2)/2}(\bar{k}R_{nN}), \tag{3.14}$$

where

$$R_{nN} = N^{1/2}\left[\sum_{j=1}^{N} (\mathbf{x}_j - \mathbf{x}_{j-1})^2\right]^{1/2} \tag{3.15}$$

and $H^{(1)}_{(nN-2)/2}(\eta)$ is the Hankel function of the first kind of order $(nN-2)/2$, then provides the final representation:

$$G_n(\mathbf{x} - \mathbf{x}_s) = \lim_{N\to\infty} \int \prod_{j=1}^{N-1} d\mathbf{x}_j \left[\frac{i}{4} N^{nN/2}\left(\frac{\bar{k}}{2\pi R_{nN}}\right)^{(nN-2)/2} H^{(1)}_{(nN-2)/2}(\bar{k}R_{nN})\right]. \tag{3.16}$$

Specifically, for the two-dimensional case,

$$G_2(x,z|x_s,z_s) = \lim_{N\to\infty} \int \prod_{j=1}^{N-1} dx_j\, dz_j \left[\frac{i}{4} N^{N}\left(\frac{\bar{k}}{2\pi R_{2N}}\right)^{N-1} H^{(1)}_{N-1}(\bar{k}R_{2N})\right], \tag{3.17}$$

where

$$R_{2N} = N^{1/2}\left\{\sum_{j=1}^{N} [(x_j - x_{j-1})^2 + (z_j - z_{j-1})^2]\right\}^{1/2}. \tag{3.18}$$

The propagator G_n^+ follows on integrating over the range coordinate in Eq. (3.12) and noting the result in Eq. (2.26). In analogy with Eq. (3.13) the representation takes the form

$$G_n^+(x,x^{(1)},x^{(2)},\ldots,x^{(n-1)}|0,x_s^{(1)},x_s^{(2)},\ldots,x_s^{(n-1)}) = \lim_{N\to\infty} \int \prod_{j=1}^{N-1} d\mathbf{x}_j (N^{(n-1)N/2} G^+_{(n-1)N+1}), \tag{3.19}$$

and in a fashion similar to Eq. (3.16),

$$\begin{aligned} G_n^+(x,x^{(1)},x^{(2)},\ldots,x^{(n-1)}|0,x_s^{(1)},x_s^{(2)},\ldots,x_s^{(n-1)}) &= \lim_{N\to\infty} \int \prod_{j=1}^{N-1} d\mathbf{x}_j \Bigg[i\pi x N^{(n-1)N/2} \\ &\times \left(\frac{\bar{k}}{2\pi R_{(n-1)N+1}}\right)^{[(n-1)N+1]/2} H^{(1)}_{[(n-1)N+1]/2}(\bar{k}R_{(n-1)N+1})\Bigg]. \end{aligned} \tag{3.20}$$

In Eqs. (3.19) and (3.20) x is the range coordinate, $\mathbf{x}_j$ is an $(n-1)$-dimensional vector, and

$$R_{(n-1)N+1} = \left(N \sum_{j=1}^{N} (\mathbf{x}_j - \mathbf{x}_{j-1})^2 + x^2\right)^{1/2}. \tag{3.21}$$

Specifically, for the two-dimensional case,

$$G_2^+(x,z|0,z_s) = \lim_{N\to\infty} \int \prod_{j=1}^{N-1} dz_j \left[i\pi x N^{N/2}\left(\frac{\bar{k}}{2\pi R_{N+1}}\right)^{(N+1)/2} H^{(1)}_{(N+1)/2}(\bar{k}R_{N+1})\right], \tag{3.22}$$

where

$$R_{N+1} = \left[N \sum_{j=1}^{N} (z_j - z_{j-1})^2 + x^2\right]^{1/2}. \tag{3.23}$$

IV. TRANSVERSELY INHOMOGENEOUS HALF-SPACE

For the transversely inhomogeneous half-space, the parabolic-based construction provides the principal results. Since $\Omega_{\mathrm{H}}(p,q)$ has only been determined approximately, in general, for the square root operator, the Hamiltonian phase space path integral can only provide an approximate representation of the propagator. The perturbation limits previously considered in Paper I[1] are again appropriate.

In the $\bar{k}\to\infty$, or high-frequency, limit, the forward propagating wave equation corresponds to the choice $F(u,v) = \exp(-\tfrac{1}{2}i\bar{k}uv)$ and $h_{\mathrm{H}}(p,q) = [K^2(q) - p^2]^{1/2}$ in accordance with the identification of the operator as a standard pseudodifferential operator.[1] It then follows from Eqs. (2.5) and (2.6) that the path integral for the propagator corresponding to the $\bar{k}\to\infty$ wave equation, taken as a high-frequency approximation to the full propagator G^+, is given by

$$G^+(x,z|0,z_s) \approx \lim_{N\to\infty} \int \prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N} \frac{dp_j^z}{2\pi/\bar{k}} \exp\left(i\bar{k} \sum_{j=1}^{N} \left\{p_j^z(z_j - z_{j-1}) + \frac{x}{N}[K^2(z_{j-1}) - (p_j^z)^2]^{1/2}\right\}\right). \tag{4.1}$$

Equation (4.1) could equally well be expressed in terms of a Weyl discretization $[F(u,v) = 1]$, in which case $h_{\mathrm{H}}(p,q)$ would no longer be simply given by $[K^2(q) - p^2]^{1/2}$, but rather in terms of an infinite series in $(1/\bar{k})$. It follows from Appendix B of Paper I, in fact, that the Weyl discretization with $h_{\mathrm{H}}(p,q) = [K^2(q) - p^2]^{1/2}$ provides an approximate propagator, which is valid through the next order, in the form

$$G^+(x,z|0,z_s) \approx \lim_{N\to\infty} \int \prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N} \frac{dp_j^z}{2\pi/\bar{k}} \exp\left(i\bar{k} \sum_{j=1}^{N} \left\{p_j^z(z_j - z_{j-1}) + \frac{x}{N}\left[K^2\left(\frac{z_j + z_{j-1}}{2}\right) - (p_j^z)^2\right]^{1/2}\right\}\right). \tag{4.2}$$

Performing the $\{p_j^z\}$ integrations in Eq. (4.1) then leads to the approximate path sum representation

$$G^{+}(x,z|0,z_s)\approx \lim_{N\to\infty}\int\prod_{j=1}^{N-1}dz_j\prod_{j=1}^{N}\left\{\left(\frac{i\bar{k}K(z_{j-1})}{2\{1+[(z_j-z_{j-1})/\delta]^2\}^{1/2}}\right)[H_1^{(1)}(\bar{k}\delta K(z_{j-1})\{1+[(z_j-z_{j-1})/\delta]^2\}^{1/2})]\right\}, \quad (4.3)$$

which for the limit $K=1$ reduces to the exact result given by Eq. (3.10). Since the high-frequency propagator contains contributions form "all of the paths" in the path integral, it is a full-wave approximation, incorporating, in an approximate manner, what are commonly referred to as medium inhomogeneity diffraction effects.

For the homogeneous half-space the functional averages of the average and root mean square "effective Hamiltonian" are equivalent [Eq. (3.7)]. This can be viewed as a lack of "dispersion" resulting from the absence of the operator-ordering question associated with the noncommuting operators **Q** and **P**. Since in the $\bar{k}\to\infty$ limit the commutator $[\mathbf{Q},\mathbf{P}]\to 0$, the Feynman–Garrod representation should provide a useful approximate representation for the propagator for the transversely inhomogeneous half-space at high frequencies. Thus,

$$G^{+}(x,z|0,z_s)\approx \lim_{N\to\infty}\int\prod_{j=1}^{N-1}dz_j\prod_{j=1}^{N}\frac{dp_j^z}{2\pi/\bar{k}}\exp\left\{i\bar{k}\sum_{j=1}^{N}[p_j^z(z_j-z_{j-1})]\right\}\exp\left(i\bar{k}x\left\{\frac{1}{N}\sum_{j=1}^{N}[K^2(z_{j-1})-(p_j^z)^2]\right\}^{1/2}\right), \quad (4.4)$$

and following from Eqs. (3.22) and (3.23),

$$G^{+}(x,z|0,z_s)\approx \lim_{N\to\infty}\int\prod_{j=1}^{N-1}dz_j\left(i\pi xN^{N/2}\left\{\frac{\bar{k}[(1/N)\Sigma_{j=1}^{N}K^2(z_{j-1})]^{1/2}}{2\pi R_{N+1}}\right\}^{(N+1)/2}H_{(N+1)/2}^{(1)}\left(\bar{k}\left[\frac{1}{N}\sum_{j=1}^{N}K^2(z_{j-1})\right]^{1/2}R_{N+1}\right)\right). \quad (4.5)$$

The Feynman–Garrod representation (4.4) is not directly based on a large $\bar{k}$ approximation to $\Omega_{\mathrm{H}}(p,q)$. Mizrahi has provided a detailed treatment of the WKB approximation to the propagator for arbitrary Hamiltonian operators.[24]

In the limit of narrow angle, weak inhomogeneity, and weak gradient, at the level of the ordinary parabolic approximation,[1]

$$\Omega_{\mathrm{H}}(p,q)=1+\tfrac{1}{2}\{[K^2(q)-1]-p^2\}. \quad (4.6)$$

The corresponding approximation to the propagator following from Eq. (2.8) then takes the form

$$G^{+}(x,z|0,z_s)\approx\exp(i\bar{k}x)\lim_{N\to\infty}\int\prod_{j=1}^{N-1}dz_j\prod_{j=1}^{N}\frac{dp_j^z}{2\pi/\bar{k}}\exp\left(i\bar{k}\sum_{j=1}^{N}\left\{p_j^z(z_j-z_{j-1})-\frac{x}{N}\left[\frac{(p_j^z)^2}{2}+V\left(\frac{z_j+z_{j-1}}{2}\right)\right]\right\}\right), \quad (4.7)$$

which, following from the Feynman–Fradkin construction, can be written as

$$G^{+}(x,z|0,z_s)\approx\exp(i\bar{k}x)\Phi(x,z|0,z_s), \quad (4.8)$$

where $\Phi(x,z|0,z_s)$ is the parabolic propagator of Eqs. (2.11) and (2.12). Performing the $\{p_j^z\}$ integrations in Eq. (4.7) leads to the Lagrangian path integral representation for the parabolic propagator corresponding to the Feynman construction.[4–6]

In the limit of narrow angle and weak gradient for arbitrary field strength,[1]

$$\Omega_{\mathrm{H}}(p,q)=K(q)-\left[\frac{p^2}{2K(q)}+\frac{1}{8\bar{k}^2}\frac{K''(q)}{K^2(q)}\right], \quad (4.9)$$

where the superscript primes denote differentiation with respect to the argument. The approximate propagator then follows from Eq. (2.8) as

$$G^{+}(x,z|0,z_s)\approx \lim_{N\to\infty}\int\prod_{j=1}^{N-1}dz_j\prod_{j=1}^{N}\frac{dp_j^z}{2\pi/\bar{k}}\exp\left\{i\bar{k}\sum_{j=1}^{N}[p_j^z(z_j-z_{j-1})]\right\}\exp\left\{i\bar{k}\frac{x}{N}\sum_{j=1}^{N}\left[\frac{-(p_j^z)^2}{2K(\bar{z}_j)}+\Gamma(\bar{z}_j)\right]\right\}, \quad (4.10)$$

where

$$\Gamma(q)=K(q)-\frac{1}{8\bar{k}^2}\frac{K''(q)}{K^2(q)} \quad (4.11)$$

and $\bar{z}_j=\frac{1}{2}(z_j+z_{j-1})$ is the average coordinate. A configuration space representation follows from Eq. (4.10) on integrating over the $\{p_j^z\}$ variables. This results in the approximate representation

$$G^{+}(x,z|0,z_s)\approx \lim_{N\to\infty}\int\prod_{j=1}^{N-1}dz_j\prod_{j=1}^{N}\left\{\left[\frac{N\bar{k}K(\bar{z}_j)}{2\pi ix}\right]^{1/2}\right\}\exp\left\{i\bar{k}\sum_{j=1}^{N}\left[\frac{(z_j-z_{j-1})^2}{2(x/N)}K(\bar{z}_j)+\frac{x}{N}\Gamma(\bar{z}_j)\right]\right\}, \quad (4.12)$$

which has the form of a Lagrangian path integral following from the quadratic dependence of the "effective Hamiltonian" on the p^z variable. Equation (4.12) can be shown to be equivalent to a result based upon a direct configuration space derivation in terms of an appropriate "short-time" propagator.[12]

For arbitrary angle in the limit of weak inhomogeneity and gradient,[1]

$$\Omega_{\mathrm{H}}(p,q)=(1-p^2)^{1/2}+\int dt\frac{\hat{\epsilon}(t)\exp(i\bar{k}qt)}{[1-(p-t/2)^2]^{1/2}+[1-(p+t/2)^2]^{1/2}}, \quad (4.13)$$

where $\hat{\epsilon}(t)$ is the Fourier transform of the field strength $\epsilon(q)=K^2(q)-1$. The approximate propagator follows from Eq. (2.8) in the form

$$G^{+}(x,z|0,z_s)\approx \lim_{N\to\infty}\int \prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N} \frac{dp_j^z}{2\pi/\bar{k}} \exp\Big\{i\bar{k}\sum_{j=1}^{N}\big[p_j^z(z_j-z_{j-1})\big]\Big\}$$
$$\times\exp\bigg(i\bar{k}\frac{x}{N}\sum_{j=1}^{N}\bigg\{[1-(p_j^z)^2]^{1/2}+\int dt\,\frac{\hat{\epsilon}(t)\exp(i\bar{k}\bar{z}_j\,t)}{[1-(p_j^z-t/2)^2]^{1/2}+[1-(p_j^z+t/2)^2]^{1/2}}\bigg\}\bigg). \tag{4.14}$$

A first-order perturbation evaluation of Eq. (4.14) yields the appropriate Born approximation. The "short-time" propagator corresponding to Eq. (4.14) takes the form

$$G^{+}(\delta,z|0,z')\approx\frac{\bar{k}}{2\pi}\int dp\,\exp[i\bar{k}p(z-z')]$$
$$\times\exp\bigg(i\bar{k}\delta\bigg\{(1-p^2)^{1/2}+\int dt\,\frac{\hat{\epsilon}(t)\exp[i\bar{k}t\,\tfrac{1}{2}(z+z')]}{[1-(p-t/2)^2]^{1/2}+[1-(p+t/2)^2]^{1/2}}\bigg\}\bigg). \tag{4.15}$$

The arbitrary-dimensional generalizations of the path integrals constructed follow immediately. For any given approximation to $\Omega_{\mathrm{H}}(p,q)$ a corresponding approximate propagator follows through Eq. (2.8). This would include, for example, uniform asymptotic treatments, unsymmetrical treatments of the transverse variables in higher spatial dimensions ($n>2$), and the wide-angle rational approximation to the wave equation. For the cases where $\Omega_{\mathrm{H}}(p,q)$ can be determined exactly,[1] formally exact Hamiltonian phase space path integral representations of the half-space Helmholtz propagator can be constructed.

V. DISCUSSION

In the development of extended parabolic wave theories, the path integrals provide the basis to relate dynamical approximations, the resolution of the square root wave operator, and the concept of an underlying stochastic process. The indirect Feynman–Fradkin path integrals provide formally exact representations for both the full and half-space Helmholtz propagators. For the half-space propagation problem, the one-sided Fourier transform of the parabolic (Schrödinger) propagator implicitly contains the proper resolution of the operator-ordering question as expressed through the construction summarized in Eqs. (1.2) and (1.3). This is seen through the equivalence of the Hamiltonian phase space and Feynman–Fradkin representations for a transversely inhomogeneous medium,

$$\lim_{N\to\infty}\int\prod_{j=1}^{N-1}dz_j\prod_{j=1}^{N}dp_j^z/(2\pi/\bar{k})\exp\bigg\{i\bar{k}\sum_{j=1}^{N}\bigg[p_j^z(z_j-z_{j-1})+\frac{x}{N}\Omega_{\mathrm{H}}\Big(p_j^z,\frac{z_j+z_{j-1}}{2}\Big)\bigg]\bigg\}$$
$$=\exp\Big(-\frac{i\pi}{4}\Big)\Big(\frac{\bar{k}}{2\pi}\Big)^{1/2}x\int_0^\infty d\tau\,\tau^{-3/2}\exp\Big[\frac{i\bar{k}}{2}\Big(\tau+\frac{x^2}{\tau}\Big)\Big]\bigg[\lim_{N\to\infty}\int\prod_{j=1}^{N-1}dz_j\prod_{j=1}^{N}\frac{dp_j^z}{2\pi/\bar{k}}$$
$$\times\exp\bigg(i\bar{k}\sum_{j=1}^{N}\bigg\{p_j^z(z_j-z_{j-1})-\frac{\tau}{N}\bigg[\frac{(p_j^z)^2}{2}+V\Big(\frac{z_j+z_{j-1}}{2}\Big)\bigg]\bigg\}\bigg)\bigg], \tag{5.1}$$

in the two-dimensional notation. Equations (5.1) and (1.3) can be said to formally accomplish the τ integration in the Feynman–Fradkin representation.

Equation (5.1) relates the linear Feynman–Fradkin Fourier analysis and the nonlinear pseudodifferential analysis. Approximate wave theories can be viewed in this light. Approximate evaluations of the τ integral in Eq. (5.1) correspond to appropriate perturbation solutions of the Weyl composition equation. In the high-frequency limit, the WKB half-space propagator follows from the Feynman–Fradkin representation through a functional stationary phase approximation on the parabolic path integral followed by an ordinary stationary phase approximation on the resulting τ integral.[5,6,16,19,25] The corresponding Hamiltonian phase space path integral analysis involves functional stationary phase methods in conjunction with Eq. (4.1) and its extensions.[1] The ordinary parabolic approximation is a singular perturbation; it can be viewed as a unidirectional asymptotic expansion of the half-space Feynman–Fradkin path integral.[15,23,25] This results in an ordinary stationary phase approximation on the τ integral in Eq. (5.1) with the parabolic propagator assumed slowly varying. This is equivalent to Eqs. (4.6) and (4.7). The limit of narrow angle and weak gradient for arbitrary field strength, corresponding to Eqs. (4.9) and (4.10), is equivalent to an ordinary stationary phase treatment of the τ integral accounting for some of the τ variations associated with the "potential" contribution of the parabolic propagator. For arbitrary angle in the limit of weak inhomogeneity and gradient, Eqs. (4.13) and (4.14) correspond to a full τ integration over an approximation to the parabolic propagator. Exactly soluble models provide specific examples of Eq. (5.1). A more detailed treatment of this correspondence will be presented elsewhere.

The approximate Feynman–Garrod path integral provides the basis for the Feynman–DeWitt-Morette path integral. The form of this symbolic representation suggests the formal dynamical character of the propagation theory. The dynamical basis of the Helmholtz equation can be viewed in terms of "free particle" motion on an appropriate curved space or, equivalently, in terms of a stochastic process embodying the fixed "average energy" condition.[19] Thus, in the half-space wave propagation theory, there is no fundamental dynamical significance attached to a limiting ($\bar{k} \to \infty$) "classical Hamiltonian" function $h_{\mathrm{H}}(p,q)$ as there is in the quantum mechanical theory ($\hbar \to 0$). This is consistent with the role of the symbol $\Omega_{\mathrm{H}}(p,q)$ as reflected in Eq. (2.7). This dynamical viewpoint is reinforced by the forms of the Hamiltonian phase space representations in the perturbation limits considered. The narrow-angle theories constructed in Eqs. (4.7) and (4.10) are suggestive of weak-coupling approximations to the Master equation in statistical mechanics, corresponding, respectively, to simple and generalized diffusion models.[1] The wide-angle theory of Eq. (4.14) is suggestive of strong-coupling approximations to the Master equation, specifically, those incorporating large discontinuous change.[1] Moreover, the form of Eq. (4.12) is analogous to that corresponding to the quantum mechanical motion of a particle moving in a curved space (the curved space having been imbedded in an appropriate Cartesian space).[6,12,13,26]

The representation presented in Eq. (3.16) is operationally correct; indeed, for any finite N it expresses a Bessel function identity. Moreover, it is a well-defined functional of the path and provides a computational prescription for the symbolic Feynman–DeWitt-Morette representation expressed in Eqs. (2.18)–(2.22) for the homogeneous medium limit. The connection between Eq. (3.16) and the symbolic sum over fixed "average energy" paths must result from an appropriate limiting procedure in the $N \to \infty$ or, functional, limit. The representation expressed in Eq. (3.10) does not in an obvious way approach a well-defined functional of the path in configuration space in the infinite limit. Moreover, attempts to interpret representations such as Eq. (3.10) as the discretized form of a Feynman path integral in terms of an appropriate Lagrangian cannot be carried out in a consistent manner in general.[22,26] Although Klauder,[27] in studying an elementary model of quantum gravity, was able to proceed from a path sum in the form expressed by Eq. (3.11) and directly identify the underlying stochastic process associated with the model, much of this success apparently lies with the process ultimately being a Weiner process in one-higher dimension and thus readily recognizable. A similar analysis applied to Eq. (3.10) is not quite so readily transparent. Equation (3.22), however, for finite N is a well-defined functional of the path and is suggestive of a formal path integral in the $N \to \infty$ limit. Equation (4.5) provides an extension to transversely inhomogeneous media in a high-frequency limit.

There are several final points. Backscatter effects associated with a range-dependent refractive index field are accounted for explicitly in the Feynman–Fradkin representations and symbolically in the Feynman–DeWitt-Morette representation. Further, for the two-dimensional half-space problem, conformal mapping techniques reduce range-dependent media to transversely inhomogeneous media for a special class of refractive index fields.[28] The path integral analysis explicitly introduces into the inverse problem, in a natural manner, the concept of an underlying stochastic process and the notion of strong- and weak-coupling regimes, in addition to an interpretation in terms of free motion on curved spaces. The extensions of the factorization analysis to hyperbolic wave equations in the time domain and to the vector formulation appropriate for wave propagation in elastic media have been discussed in Paper I. For the acoustic field coherence function, phase space path integrals can be constructed in two ways. The bilinear form of the coherence function immediately allows for a representation as a product of two Eq. (2.8) path integrals. The Wigner function, an appropriate Fourier transform of the coherence function, satisfies a composition equation of motion in terms of $\Omega_{\mathrm{H}}(p,q)$.[29] The Cohen/Agarwal–Wolf construction then provides the basis for a phase space representation of the Wigner function. The path integral representations have immediate application in the theory of wave propagation in random media.[30]

ACKNOWLEDGMENTS

The work reported here was supported under a grant from the National Science Foundation (EAR-8025367). In addition, L. F. spent one summer with the Large Aperture Acoustics Branch of the Naval Research Laboratory.

APPENDIX: PARTIAL INTEGRATIONS AND REDUCED REPRESENTATIONS

For the two-dimensional case for an x- or range-independent potential, Eq. (2.15) can be written as

$$G(x,z|x_s,z_s) \doteq \frac{-1}{2\bar{k}^2} \lim_{N\to\infty} \int \prod_{j=1}^{N-1} dz_j \prod_{j=1}^{N} \frac{dp_j^z}{2\pi/\bar{k}} \exp(i\bar{k}S_N^z) \int \prod_{j=1}^{N-1} dx_j \prod_{j=1}^{N} \frac{dp_j^x}{2\pi/\bar{k}}$$

$$\times \frac{\exp(i\bar{k}S_N^x)}{\frac{1}{2} - (1/N)\{\Sigma_{j=1}^{N} [\frac{1}{2}(p_j^z)^2 + \frac{1}{2}(p_j^x)^2 + V(\frac{1}{2}(z_j + z_{j-1}))]\}} \tag{A1}$$

in an obvious notation. Writing the action analog S_N^x as

$$S_N^x = \sum_{j=1}^{N} p_j^x(x_j - x_{j-1}) = p_N^x\, x - p_1^x\, x_s + (p_1^x - p_2^x)x_1 + \cdots + (p_{N-1}^x - p_N^x)x_{N-1} \tag{A2}$$

and noting the integral representation of the delta function,

$$\delta(\rho) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\xi\, \exp(i\xi\rho), \tag{A3}$$

allows the $\{x_j\}$ integrations to be carried out giving

$$G(x,z|x_s,z_s) \doteq \frac{-1}{2\bar{k}^2}\lim_{N\to\infty}\frac{\bar{k}}{2\pi}\int\prod_{j=1}^{N-1}dz_j\prod_{j=1}^{N}\frac{dp_j^z}{2\pi/\bar{k}}\exp(i\bar{k}S_N^z)\int\prod_{j=1}^{N}dp_j^x$$
$$\times\frac{\exp[i\bar{k}(p_N^x\, x - p_1^x\, x_s)]\delta(p_1^x - p_2^x)\cdots\delta(p_{N-1}^x - p_N^x)}{\frac{1}{2} - (1/N)\{\Sigma_{j=1}^N\,[\frac{1}{2}(p_j^z)^2 + \frac{1}{2}(p_j^x)^2 + V(\frac{1}{2}(z_j + z_{j-1}))]\}}. \tag{A4}$$

Carrying out the $\{p_j^x\}$ integrations results in

$$G(x,z|x_s,z_s) \doteq \frac{-1}{2\bar{k}^2}\lim_{N\to\infty}\frac{\bar{k}}{2\pi}\int\prod_{j=1}^{N-1}dz_j\prod_{j=1}^{N}\frac{dp_j^z}{2\pi/\bar{k}}\exp(i\bar{k}S_N^z)$$
$$\times\int_{-\infty}^{\infty}dp\,\frac{\exp[i\bar{k}p(x - x_s)]}{\left(\frac{1}{2} - (1/N)\left\{\sum_{j=1}^{N}[\frac{1}{2}(p_j^z)^2 + V(\frac{1}{2}(z_j + z_{j-1}))]\right\}\right) - \frac{1}{2}p^2}. \tag{A5}$$

The remaining p integral is elementary,

$$\frac{\bar{k}}{2\pi}\int_{-\infty}^{\infty}dp\,\frac{\exp[i\bar{k}p(x - x_s)]}{(\frac{1}{2} - \mathscr{E}) - \frac{1}{2}p^2} = \frac{\exp[i\bar{k}\,2^{1/2}(\frac{1}{2} - \mathscr{E})^{1/2}|x - x_s|]}{(\frac{1}{2} - \mathscr{E})^{1/2}}, \tag{A6}$$

and results in the representation of Eq. (2.23). For the homogeneous medium case, $V(x,z) = 0$, the same procedures used in going from Eq. (A1) to Eq. (A5) when applied to Eq. (2.23) lead to the result

$$G(x,z|x_s,z_s) = \frac{i}{4\pi}\int_{-\infty}^{\infty}dp\frac{\exp\{i[p(z - z_s) + (\bar{k}^2 - p^2)^{1/2}|x - x_s|]\}}{(\bar{k}^2 - p^2)^{1/2}}. \tag{A7}$$

Equation (A7) is just a standard integral representation for the Hankel function $H_0^{(1)}(\eta)$[31] and thus the well-known homogeneous medium limit is recovered,

$$G(x,z|x_s,z_s) = \tfrac{1}{4}iH_0^{(1)}(\bar{k}\,[(x - x_s)^2 + (z - z_s)^2]^{1/2}). \tag{A8}$$

The three-dimensional case follows in the same manner. When $V(x,y,z) = V(y,z)$, the steps going from Eq. (A1) to Eq. (A6) lead to the two-dimensional version of Eq. (2.23). When $V(x,y,z) = V(z)$, application of the steps leading from Eq. (2.23) to Eq. (A8) results in the representation of Eq. (2.24). For the homogeneous medium, $V(x,y,z) = 0$, the same procedures applied to Eq. (2.24), noting the result[32]

$$\int_{-\infty}^{\infty}dp\,\exp[ip(z - z_s)]H_0^{(1)}((\bar{k}^2 - p^2)^{1/2}[(x - x_s)^2 + (y - y_s)^2]^{1/2}) = (-2i/R_3)\exp(i\bar{k}R_3), \tag{A9}$$

lead to the well-known expression

$$G(x,y,z|x_s,y_s,z_s) = (1/4\pi R_3)\exp(i\bar{k}R_3), \tag{A10}$$

where $R_3 = [(x - x_s)^2 + (y - y_s)^2 + (z - z_s)^2]^{1/2}$.

[1] L. Fishman and J. J. McCoy, "Derivation and application of extended parabolic wave theories. I. The factorized Helmholtz equation," J. Math. Phys. **25**, 285 (1984).
[2] H. Davies, "Hamiltonian approach to the method of summation over Feynman histories," Proc. Cambridge Phil. Soc. **59**, 147 (1963).
[3] C. Garrod, "Hamiltonian path-integral methods," Rev. Mod. Phys. **38**, 483 (1966).
[4] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[5] J. B. Keller and D. W. McLaughlin, "The Feynman integral," Am. Math. Monthly **82**, 457 (1975).
[6] L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
[7] M. M. Mizrahi, "Phase space path integrals, without limiting procedure," J. Math. Phys. **19**(1), 298 (1978).
[8] M. M. Mizrahi, "Correspondence rules and path integrals," in *Feynman Path Integrals*, Lecture Notes in Physics No. 106, edited by S. A. Albeverio *et al.* (Springer-Verlag, New York, 1979), p. 234.
[9] F. A. Berezin, "Non-Wiener functional integrals," Theor. Math. Phys. **6**, 141 (1971).
[10] F. J. Testa, "Quantum operator ordering and the Feynman formulation," J. Math. Phys. **12**(8), 1471 (1971).
[11] I. W. Mayes and J. S. Dowker, "Canonical functional integrals in general coordinates," Proc. Roy. Soc. London. A **327**, 131 (1972).
[12] I. W. Mayes and J. S. Dowker, "Hamiltonian orderings and functional integrals," J. Math. Phys. **14**(4), 434 (1973).
[13] J. S. Dowker, "Path integrals and ordering rules," J. Math. Phys. **17**(10), 1873 (1976).
[14] V. I. Klyatskin and V. I. Tatarskii, Zh. Eksp. Teor. Fiz. **58**, 624 (1970) ["The parabolic equation approximation for propagation of waves in a

medium with random inhomogeneities," Sov. Phys. JETP **31**, 335 (1970)].
[15]D. R. Palmer, "A path-integral approach to the parabolic approximation. I," J. Acoust. Soc. Am. **66**(3), 862 (1979).
[16]M. C. Gutzwiller, "Phase-integral approximation in momentum space and the bound state of an atom," J. Math. Phys. **8**, 1979 (1967).
[17]M. C. Gutzwiller, "Path integrals and the relation between classical and quantum mechanics," in *Path Integrals and Their Applications in Quantum, Statistical, and Solid State Physics*, edited by G. J. Papadopoulas and J. T. Devreese (Plenum, New York, 1978), p. 163.
[18]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1965).
[19]C. DeWitt-Morette, A. Maheshwari, and B. Nelson, "Path integration in non-relativistic quantum mechanics," Phys. Rep. **50** (5) (March 1979).
[20]A. Friedman, *Stochastic Differential Equations and Applications* (Academic, New York, 1975).
[21]J. A. DeSanto, "Relation between the solutions of the Helmholtz and parabolic equations for sound propagation," J. Acoust. Soc. Am. **62**(2), 295 (1977).
[22]B. Mühlschlegel, "Path integral associated with the Fokker-Planck equation," in *Path Integrals and Their Applications in Quantum, Statistical, and Solid State Physics*, edited by G. J. Papadopoulas and J. T. Devreese (Plenum, New York, 1978), p. 39.
[23]D. R. Palmer, "*An Introduction to the Application of Feynman Path Integrals to Sound Propagation in the Ocean*," Naval Research Laboratory Report No. 8148, 1978.
[24]M. M. Mizrahi, "On the WKB approximation to the propagator for arbitrary Hamiltonians," J. Math. Phys. **22**(1), 102 (1981).
[25]P. -L. Chow, "A functional phase-integral method and applications to the laser beam propagation in random media," J. Stat. Phys. **12**(2), 93 (1975).
[26]M. M. Mizrahi, "The Weyl correspondence and path integrals," J. Math. Phys. **16**(11), 2201 (1975).
[27]J. R. Klauder, "Path integrals for affine variables," in *Functional Integration: Theory and Applications*, edited by J.-P. Antoine and E. Tirapegui (Plenum, New York, 1980), p. 101.
[28]J. A. DeSanto, "Connection between the solutions of the Helmholtz and parabolic equations for sound propagation," in *Proceedings of the Oceanic Acoustic Modelling Conference* (p. 43), LaSpezia, Italy, 1975, edited by W. Bachmann and R. B. Williams, Saclanteen Conference Proceedings No. 17 (STI, Saclanteen, 1975).
[29]S. R. DeGroot and L. G. Suttorp, *Foundations of Electrodynamics* (North-Holland, Amsterdam, 1972), Chap. 6.
[30]R. Dashen, "Path integrals for waves in random media," J. Math. Phys. **20**(5), 894 (1979).
[31]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. 1.
[32]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965).

The Doppler effect: Now you see it, now you don't

Dan Censor
Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel 84105

D. M. Le Vine
Goddard Laboratory for Atmospheric Sciences, Goddard Space Flight Center, Greenbelt, Maryland 20771

(Received 24 February 1983; accepted for publication 11 August 1983)

Two main classes of problems are identified in the theory of electromagnetic scattering in velocity-dependent systems. The first involves transformation of space and time coordinates and field components from the laboratory system of reference to the comoving system of the scatterer, solution of the scattering problem, and inverse transformations. In general, this method displays the Doppler frequency shifts. The second class involves the substitution of Minkowski's constitutive relations into Maxwell's equations for harmonic time variation, heuristically stipulating the absence of Doppler frequency shifts. The interrelation between the two methods is investigated here. It is argued that the second method is a limiting case for very low, as well as very high frequencies, and provided the mean square fluctuation of the dielectric constant is small, and the geometrical boundaries defining the scatterers are fixed. Canonical problems of plane, cylindrical, and spherical stratification are discussed and analytical results for the scattered fields are derived. If the parameters of the problem do not meet the above conditions, the first method should be used, giving rise, in general, to a whole spectrum of frequencies due to the Doppler effect.

PACS numbers: 03.50.De

INTRODUCTION

Since Doppler's original work,[1] no other method has surpassed his as a tool for remote sensing of velocity by means of scattering of waves. Substantial progress in the theory of the Doppler effect has been made by Einstein.[2] A detailed review is given elsewhere.[3] The present discussion is concerned with first-order velocity effects; hence the exact special relativistic formulas become considerably simpler. In particular, we address ourselves to the question of remote sensing of velocity by means of electromagnetic waves, but the same methods are applicable to other wave phenomena. In some cases the motion of a single object is probed, as in the case of a radar observing a moving target. In a variety of cases the remote sensing concerns the collective behavior of an assemblage of objects, i.e., a moving medium. For example, consider the case of measuring the velocity of falling raindrops, or the observation of the Doppler effect produced by particles and irregularities carried along in a moving fluid. Subsequently, it is shown that in certain cases the overall medium effects cancel the Doppler effects produced by individual scatterers, but the velocity effects are still present in the macroscopic constitutive parameters of the medium. It is precisely this aspect of the problem which motivated the present study.

Problems involving velocity-dependent wave systems usually fall into one of two categories. The first class of problems involves single moving scatterers. The simplest problem of scattering by a moving plane mirror has been discussed by Einstein.[2,3] More complicated geometries have been considered by Le Vine[4] and Censor,[5] who also cite earlier work. Characteristically, in problems of this kind, the scattered fields are time dependent, such that in the far field the Doppler effect can be identified, but there are also other effects, e.g., the time-dependent amplitude due to the changing distance between the object and the observer. The other class of problems involves moving media, bounded by surfaces whose position in space is fixed, for example, channel flows. Problems of this kind are usually tackled by inserting the Minkowski constitutive relations[6,7] into Maxwell's equations for time harmonic variation. The pertinent differential equations can be solved for special cases, e.g., Refs. 8–11. For irrotational motion $\nabla \times \mathbf{v} = 0$, Censor[12] derived a general transformation which reduces the equations to the original form of the Maxwell equations for media at rest. For rotating systems Van Bladel[13] gives special solutions and cites previous work. The velocity effects can also be interpreted in terms of electric and magnetic sources, as done by Van Bladel.[14] In these problems there are no Doppler effects, although the velocity enters into the results.

Of course, the two categories are based on the same physical model—Maxwell's equations coupled with special relativity.[7] Therefore, a transition between the two classes of problems should exist, and criteria for the vanishing of Doppler effects are expected. To investigate this problem, we consider the medium as a collection of randomly positioned scatterers. The collective effects are computed, and the significance of the Doppler effect[1] is considered.

We start by summarizing the two fundamental approaches: (1) Maxwell–Minkowski equations for moving media and (2) scattering by moving objects. To be specific, the special case of the slab region is discussed in some detail. Following this, conditions are discussed for which the Maxwell–Minkowski model is valid, and the circumstances when

 0022-2488/84/020309-08$02.50

it is inapplicable. This points out to the way more complicated geometries should be handled. For cases where the Maxwell–Minkowski formalism applies, theoretical results are derived for scattering from nonuniform channel flows and cylindrically and spherically stratified flows.

THE TWO BASIC APPROACHES

In the "laboratory" or "unprimed" inertial frame of reference, Maxwell's equations for sourceless domains are given by

$$\begin{aligned}&\nabla\times\mathbf{H}-\frac{\partial\mathbf{D}}{\partial t}=0,\quad \nabla\cdot\mathbf{D}=0,\\ &\nabla\times\mathbf{E}+\frac{\partial\mathbf{B}}{\partial t}=0,\quad \nabla\cdot\mathbf{B}=0,\end{aligned}\tag{1}$$

in the conventional MKS notation.[7] The equations for the "comoving" or "primed" frame are given by the same structure (1), with primed symbols. Special relativity prescribes

$$\begin{aligned}&\mathbf{r}'=\widetilde{U}\cdot(\mathbf{r}-\mathbf{v}t),\qquad t'=\gamma(t-\mathbf{r}\cdot\mathbf{v}/c^2),\\ &\widetilde{U}=\widetilde{I}+(\gamma-1)\hat{\mathbf{v}}\hat{\mathbf{v}},\quad \gamma=(1-v^2/c^2)^{-1/2},\\ &v=\mathbf{v}/|\mathbf{v}|=\mathbf{v}/v,\qquad c=(\mu_0\epsilon_0)^{-1/2},\\ &\mathbf{E}'=\widetilde{V}\cdot(\mathbf{E}+\mathbf{v}\times\mathbf{B}),\\ &\mathbf{B}'=\widetilde{V}\cdot(\mathbf{B}-\mathbf{v}\times\mathbf{E}/c^2),\\ &\mathbf{D}'=\widetilde{V}\cdot(\mathbf{D}+\mathbf{v}\times\mathbf{H}/c^2),\\ &\mathbf{H}'=\widetilde{V}\cdot(\mathbf{H}-\mathbf{v}\times\mathbf{D}),\\ &\widetilde{V}=\gamma\widetilde{I}+(1-\gamma)\hat{\mathbf{v}}\hat{\mathbf{v}},\end{aligned}\tag{2}$$

where $\widetilde{I}$ is the idemfactor dyadic and $\mathbf{v}$ is the velocity of the primed system as observed from the unprimed one. Minkowski[6,7] assumed $\mathbf{D}'=\epsilon\mathbf{E}'$, $\mathbf{B}'=\mu\mathbf{H}'$ in the comoving system of reference, in which the medium is at rest, and used (2) to express the constitutive relations in terms of the laboratory frame fields. To the first order in v/c and harmonic time variations $e^{-i\omega t}$, (1) becomes

$$\begin{aligned}&\nabla\times\mathbf{H}+i\omega\epsilon\mathbf{E}=-i\omega\mathbf{\Lambda}\times\mathbf{H},\\ &\nabla\times\mathbf{E}-i\omega\mu\mathbf{H}=-i\omega\mathbf{\Lambda}\times\mathbf{E},\\ &\nabla\cdot\mathbf{H}=\mu^{-1}\nabla\cdot(\mathbf{\Lambda}\times\mathbf{E}),\\ &\nabla\cdot\mathbf{E}=-\epsilon^{-1}\nabla\cdot(\mathbf{\Lambda}\times\mathbf{H}),\\ &\mathbf{\Lambda}=(\mu\epsilon-\mu_0\epsilon_0)\mathbf{v},\end{aligned}\tag{3}$$

referred to as the Maxwell–Minkowski equations. Obviously, (3) applies only to cases where the Doppler effects are absent or negligible. This is a very strong condition because we should expect the Doppler effect to appear or vanish as a result of the geometry and other parameters of a specific problem, not by axiomatically imposing a constraint on the solution. Usually it is argued that the assumption of harmonic time variation is valid for media and boundaries whose properties are time-independent, although velocity is present. As shown in the following, this statement is too vague and might be misleading.

Strictly speaking, (3) applies to constant velocities only. This, however, rules out cases of interest involving nonuniform motion, such as rotation. Many authors heuristically stipulate that (3) applies to nonuniform motion as well.[3]

A general method for solving (3) is to assume the velocity terms to act as sources.[14,15] Since (3) is correct to first order in v/c, the zero order terms $\mathbf{E}_0,\mathbf{H}_0$ can be found by taking $\mathbf{\Lambda}=0$ in (3). The terms $-i\omega\mathbf{\Lambda}\times\mathbf{H}_0$, $-i\omega\mathbf{\Lambda}\times\mathbf{E}_0$ are then treated as sources to derive the particular solution $\mathbf{E}_\Lambda,\mathbf{H}_\Lambda$. Correct to the first order in v/c, the total field is given by $\mathbf{E}=\mathbf{E}_0+\mathbf{E}_\Lambda$, $\mathbf{H}=\mathbf{H}_0+\mathbf{H}_\Lambda$. Accordingly, (3) is recast in the form

$$\begin{aligned}&\nabla\times\mathbf{H}+i\omega\epsilon\mathbf{E}=-i\omega\mathbf{\Lambda}\times\mathbf{H}_0\equiv\mathbf{j}_e,\\ &\nabla\times\mathbf{E}-i\omega\mu\mathbf{H}=-i\omega\mathbf{\Lambda}\times\mathbf{E}_0\equiv-\mathbf{j}_m,\\ &\nabla\cdot\mathbf{H}=\mu^{-1}\nabla\cdot(\mathbf{\Lambda}\times\mathbf{E}_0)\equiv\mu^{-1}\rho_m,\\ &\nabla\cdot\mathbf{E}=-\epsilon^{-1}\nabla\cdot(\mathbf{\Lambda}\times\mathbf{H}_0)\equiv\epsilon^{-1}\rho_e.\end{aligned}\tag{4}$$

The electric and magnetic sources satisfy the continuity equations

$$\begin{aligned}&\nabla\cdot\mathbf{j}_e-i\omega\rho_e=0,\\ &\nabla\cdot\mathbf{j}_m-i\omega\rho_m=0.\end{aligned}\tag{5}$$

The particular solution of (4) (see, for example, Papas[16]) is given by

$$\begin{aligned}&\mathbf{E}_\Lambda=\int_{V(\mathbf{r}')}dV\ (\boldsymbol{\rho})\{\widetilde{\Gamma}(\mathbf{r},\boldsymbol{\rho})\cdot\mathbf{j}_e(\boldsymbol{\rho})i\omega\mu-\nabla\times\widetilde{\Gamma}(\mathbf{r},\boldsymbol{\rho})\cdot\mathbf{j}_m(\boldsymbol{\rho})\},\\ &\mathbf{H}_\Lambda=\int_{V(\mathbf{r}')}dV\ (\boldsymbol{\rho})\{\widetilde{\Gamma}(\mathbf{r},\boldsymbol{\rho})\cdot\mathbf{j}_m i\omega\epsilon+\nabla\times\widetilde{\Gamma}(\mathbf{r},\boldsymbol{\rho})\cdot\mathbf{j}_e(\rho)\},\\ &\widetilde{\Gamma}(r,\rho)=(\widetilde{I}+k^{-2}\nabla\nabla)G(\mathbf{r},\boldsymbol{\rho}),\quad k^2=\omega^2\mu\epsilon,\\ &\nabla\times\nabla\times\widetilde{\Gamma}-k^2\widetilde{\Gamma}=\widetilde{I}\delta(\mathbf{r}-\boldsymbol{\rho}),\quad (\nabla^2+k^2)G=-\delta(\mathbf{r}-\boldsymbol{\rho}),\end{aligned}\tag{6}$$

where $\widetilde{\Gamma}$ is the dyadic Green function corresponding to the scalar Green function G. Since the integrals (6) involve the source regions $\mathbf{r}=\boldsymbol{\rho}$, the question of convergence naturally arises. We obviate such problems and controversial issues currently discussed in the literature,[17] by treating (6) as a symbolic form. Actually the integration on G, (6), is performed first and the differential operators $\nabla\nabla$ and ∇ are applied later. This is tantamount to solving the problem in terms of a vector potential, for which the convergence of the integral in the presence of regular $\mathbf{j}_e,\mathbf{j}_m$ can be safely assumed.

The second approach is more general, starting with the analysis of scattering by a single moving object.[5] Media are then constructed by considering ensembles of such objects. The medium properties are obtained by averaging, and for some cases the equivalence to the Maxwell–Minkowski model can be established.

The incident wave exciting the system is

$$\mathbf{E}_i=\mathbf{e}_i e^{i\mathbf{k}_i\cdot\mathbf{r}-i\omega_i t}.\tag{7}$$

The local coordinate system of the object is defined by $\bar{\mathbf{r}}=\mathbf{r}-\boldsymbol{\rho}$, where $\boldsymbol{\rho}$ locates its origin. Translated to this coordinate system, (7) becomes

$$\mathbf{E}_i=\mathbf{e}_i e^{i\mathbf{k}_i\cdot\boldsymbol{\rho}}e^{i\mathbf{k}_i\cdot\bar{\mathbf{r}}-i\omega_i\bar{t}},\tag{8}$$

where $t = \bar{t}$. For a moving scatterer we define a system $\bar{r},\bar{t}'$ such that at $\bar{t}' = \bar{t} = 0$, the origins $\bar{\mathbf{r}}' = \bar{\mathbf{r}} = 0$, coincide. In the comoving system $\bar{r}',\bar{t}'$ we have according to (2), to first order in v/c,

$$
\begin{aligned}
&\mathbf{E}_i' = \mathbf{e}_i' e^{i\mathbf{k}_i\cdot\boldsymbol{\rho}} e^{i\mathbf{k}_i'\cdot\bar{\mathbf{r}}' - i\omega_i'\bar{t}'},\\
&\omega_i' = \omega_i - \mathbf{k}_i\cdot\mathbf{v},\\
&\mathbf{k}_i' = \mathbf{k}_i - \omega_i\mathbf{v}/c^2,\\
&\bar{\mathbf{r}}' = \bar{\mathbf{r}} - \mathbf{v}t,\\
&\bar{t}' = \bar{t} - \bar{\mathbf{r}}\cdot\mathbf{v}/c^2,\\
&\mathbf{e}_i' = \mathbf{e}_i + \mathbf{v}\times\hat{\mathbf{k}}\times\mathbf{e}_i/c.
\end{aligned}
\tag{9}
$$

This plane wave excites the object, such that in the far field in the $\bar{\mathbf{r}}',\bar{t}'$ frame of reference we have

$$
\mathbf{E}_s' = e_i' e^{i\mathbf{k}_i\cdot\boldsymbol{\rho}} \frac{e^{ik_i'\bar{r}' - i\omega_i'\bar{t}'}}{4\pi\bar{r}'}\,\mathbf{f}(\hat{\bar{\mathbf{r}}}',\hat{\mathbf{k}}_i',\omega_i',\hat{\mathbf{e}}_i'),
\tag{10}
$$

where $e_i' = |\mathbf{e}_i'|$, and the Green function is chosen as $e^{ik_i'\bar{r}_i'}/4\pi\bar{r}_i'$; hence $\mathbf{f}$ has the dimension of distance. The scattering amplitude $\mathbf{f}$ depends on the directions of incidence and observation $\hat{\mathbf{k}}_i',\hat{\bar{\mathbf{r}}}'$, respectively, on the frequency of excitation ω_i', and on the direction of polarization $\hat{\mathbf{e}}_i'$. The corresponding scattered field $\mathbf{E}_s$ in $\bar{\mathbf{r}},\bar{t}$ is given,[2] correct to first order in the velocity, as

$$
\mathbf{E}_s = \mathbf{E}_s' - \mathbf{v}\times\hat{\bar{\mathbf{r}}}'\times\mathbf{E}_s'/c.
\tag{11}
$$

In order to deal with a collection of scatterers and use (10), assume that the objects are in the far field with respect to each other, and that single scattering only is involved. After expressing (11) in terms of $\mathbf{r},t$ coordinates, we derive an explicit expression in $\boldsymbol{\rho}$. Now the ensemble properties must be introduced, e.g., the density $\alpha(\boldsymbol{\rho})$. The average field is obtained according to

$$
\langle\mathbf{E}_s\rangle = \int dV(\boldsymbol{\rho})\mathbf{E}_s(\boldsymbol{\rho})\alpha(\boldsymbol{\rho}).
\tag{12}
$$

Similarly, other statistical moments can be computed.[18,19] It will be shown that in certain cases the Doppler effect vanishes, facilitating a direct comparison with the results of the Maxwell–Minkowski model (6). In other cases the Doppler effect does not vanish, demonstrating the inadequacy of the Maxwell–Minkowski model for such problems.

SCATTERING FROM A SLAB REGION

Quantitative results for comparing the two methods are obtained by analyzing the relatively simple problem of a slab region geometry. The slab region is in the xz plane, and a thin slab of thickness Δy is considered, in order to simplify the analysis. The time-independent velocity field $\mathbf{v}(\boldsymbol{\rho})$ is defined in the xz plane. The integration involves the position vector $\boldsymbol{\rho}(\xi,\zeta)$, and $\mathbf{E}_A,\mathbf{H}_A$ are replaced by $\mathbf{E}_A/\Delta y,\mathbf{H}_A/\Delta y$ to account for the integration in the y direction. The incident wave provides the excitation $\mathbf{E}_0 = e_i e^{i\mathbf{k}_i\cdot\boldsymbol{\rho} - i\omega t}$. Assuming $\mu = \mu_0$ everywhere, we have from the Maxwell–Minkowski model (4)

$$
\begin{aligned}
\frac{\mathbf{E}_A}{\Delta y} &= (\tilde{I} + k^{-2}\nabla\nabla)\cdot\int d\xi\, d\zeta\, \frac{e^{ik_i|\mathbf{r}-\boldsymbol{\rho}|}}{4\pi|\mathbf{r}-\boldsymbol{\rho}|} i\omega\mu_0\,\mathbf{j}_e\\
&\quad - \nabla\times\int d\xi\, d\zeta\, \frac{e^{ik_i|\mathbf{r}-\boldsymbol{\rho}|}}{4\pi|\mathbf{r}-\boldsymbol{\rho}|}\mathbf{j}_m,\\
\mathbf{j}_e &= -i(\epsilon-\epsilon_0)\mathbf{v}(\boldsymbol{\rho})\times\mathbf{k}_i\times\mathbf{E}_0,\\
\mathbf{j}_m &= i\omega\mu_0(\epsilon-\epsilon_0)\mathbf{v}(\boldsymbol{\rho})\times\mathbf{E}_0,
\end{aligned}
\tag{13}
$$

and the corresponding expression for $\mathbf{H}_A/\Delta y$ is given by inspection of (6). Throughout space (13) yields the harmonic time variation $e^{-i\omega t}$, with the frequency ω of the incident wave. If $\mathbf{v}(\boldsymbol{\rho})$ is a rapidly changing function in the xz plane, the integral cannot be expected to yield transmitted and reflected plane waves in the forward and specular reflection directions. For such a case one would expect Doppler effects, discussed in the context of the second method mentioned above. On the other hand, if $\mathbf{v}(\boldsymbol{\rho})$ is a slowly varying function and k_i is large, the stationary phase approximation can be used to compute (13) (e.g., see Twersky[19]):

$$
\begin{aligned}
&\int G(\boldsymbol{\rho})e^{i\kappa g(\boldsymbol{\rho})}\,d\xi\, d\zeta \sim \frac{2\pi i}{\kappa}\,\frac{G(\boldsymbol{\rho}_s)e^{i\kappa g(\boldsymbol{\rho}_s)}}{[g_{\xi\xi}(\boldsymbol{\rho}_s)g_{\zeta\zeta}(\boldsymbol{\rho}_s) - g_{\xi\zeta}^2(\boldsymbol{\rho}_s)]^{1/2}},\\
&\frac{\partial}{\partial\xi}g(\boldsymbol{\rho}_s) = g_\xi(\boldsymbol{\rho}_s) = 0,\quad g_\zeta(\boldsymbol{\rho}_s) = 0,
\end{aligned}
\tag{14}
$$

where $\boldsymbol{\rho}_s$ are the stationary points. For the present case κg in (14) is given by

$$
\begin{aligned}
&\kappa g(\boldsymbol{\rho}) = k_x\xi + k_z\zeta + k_i[(x-\xi)^2 + (z-\zeta)^2]^{1/2},\\
&\kappa g(\boldsymbol{\rho}_s) = k_x x + k_z z \pm k_y y = \mathbf{k}_\pm\cdot\mathbf{r},\\
&k_i^2 = k_x^2 + k_y^2 + k_z^2 = k_\rho^2 + k_y^2,\\
&\frac{k_y}{k_\rho} = \pm\frac{y}{[(x-\xi_s)^2 + (z-\zeta_s)^2]^{1/2}},\quad \frac{\mathbf{k}_\rho}{k} = \frac{\mathbf{r}-\boldsymbol{\rho}_s}{|\mathbf{r}-\boldsymbol{\rho}_s|},\\
&[g_{\xi\xi}g_{\zeta\zeta} - g_{\xi\zeta}^2]^{1/2}|_{\rho s} = \frac{k_y^2}{ky},\quad \frac{2\pi i}{4\pi|\mathbf{r}-\boldsymbol{\rho}_s|}\,\frac{k_y^2}{ky} = \frac{i}{2ky}.
\end{aligned}
\tag{15}
$$

Returning to (14), we now have

$$
\begin{aligned}
\frac{\mathbf{E}_A}{\Delta y} &= \frac{i\omega\mu_0(\epsilon-\epsilon_0)}{2k_y}e^{-i\omega t}\\
&\quad\times\left\{\left(\tilde{I} + \frac{\nabla\nabla}{k^2}\right)\cdot e^{i\mathbf{k}_\pm\cdot\mathbf{r}}\mathbf{v}\times\mathbf{k}_i\times\mathbf{e}_i - i\nabla e^{i\mathbf{k}_\pm\cdot\mathbf{r}}\times\mathbf{v}\times\mathbf{e}_i\right\}\\
&= \frac{i\omega\mu_0(\epsilon-\epsilon_0)}{2k_y}e^{i\mathbf{k}_\pm\cdot\mathbf{r} - i\omega t}\\
&\quad\times\left\{\left(\tilde{I} - \frac{\mathbf{k}_\pm\mathbf{k}_\pm}{k^2}\right)\cdot\mathbf{v}\times\mathbf{k}_i\times\mathbf{e}_i + \mathbf{k}_\pm\times\mathbf{v}\times\mathbf{e}_i\right\}.
\end{aligned}
\tag{16}
$$

Similarly (6) becomes

$$
\begin{aligned}
\frac{\mathbf{H}_A}{\Delta y} &= \frac{i(\epsilon-\epsilon_0)}{2k_y}e^{i\mathbf{k}_\pm\cdot\mathbf{r} - i\omega t}\left\{-\left(\tilde{I} - \frac{\mathbf{k}_\pm\mathbf{k}_\pm}{k^2}\right)\right.\\
&\quad\left.\cdot k^2\mathbf{v}\times\mathbf{e}_i + \mathbf{k}_\pm\times\mathbf{v}\times\mathbf{k}_i\times\mathbf{e}_i\right\}.
\end{aligned}
\tag{17}
$$

As expected, we have forward propagation and specular reflection. The results (16) and (17) show that no depolarization occurs in this problem. For normal incidence the expression in braces in (16) and (17) vanishes, and with $k_y \to 0$ the fields become indeterminate. Note that (15) prescribes

$$\mathbf{v}(\boldsymbol{\rho}) = \mathbf{v}(\mathbf{r} - (\mathbf{k}_\rho / k)|\mathbf{r} - \boldsymbol{\rho}_s|); \tag{18}$$

hence the field at $\mathbf{r}$ is due to the velocity at $\boldsymbol{\rho}$, such that the ray $\mathbf{k}_\pm$ and $\mathbf{r} - \boldsymbol{\rho}_s$ are on the same line.

Treating the problem as a collection of moving scatterers and solving by successive transformations and integration, as indicated by (8)–(12), (12) now becomes

$$\langle \mathbf{E}_s \rangle = \int d\xi \, d\zeta \, e_i \, \frac{e^{i(\mathbf{k}_i \cdot \boldsymbol{\rho} + k_i' \bar{r}' - \omega_i' \bar{t}')}}{4\pi \bar{r}'} \times \alpha(\boldsymbol{\rho}) \left(\mathbf{f} - \mathbf{v}(\boldsymbol{\rho}) \times \hat{\bar{\mathbf{r}}}' \times \frac{\mathbf{f}}{c} \right), \tag{19}$$

which we have to express in terms of $\mathbf{r},t$ coordinates before integrating. First, the transformation of the phase (19) is considered. For an observer along the $\hat{\bar{\mathbf{r}}}'$ direction, we have

$$\begin{aligned} k_i' \hat{\bar{\mathbf{r}}}' \cdot \bar{\mathbf{r}}' - \omega_i' \bar{t}_i' &= \mathbf{k}_s \cdot \bar{\mathbf{r}} - \omega_s \bar{t}, \\ \mathbf{k}_s &= k_i' \hat{\bar{\mathbf{r}}}' + \omega_i' \mathbf{v} / c^2, \\ \omega_s &= \omega_i' (1 + (\mathbf{v} \cdot \hat{\bar{\mathbf{r}}}) / c) \\ &= \omega_i (1 - (\mathbf{v} \cdot \hat{\mathbf{k}}_i) / c + (\mathbf{v} \cdot \hat{\bar{\mathbf{r}}}) / c), \end{aligned} \tag{20}$$

to the first order in v/c. This result displays the direction-dependent Doppler effect, which in general conflicts with the result of the Maxwell–Minkowski model. In order to contrast (19) with (16), we assume that the stationary phase approximation applies to (19). Using $t = \bar{t}$, $\bar{\mathbf{r}} = \mathbf{r} - \boldsymbol{\rho}$, and (2) yields, to first order in v/c,

$$\bar{t}' = \bar{t} - (\bar{\mathbf{r}} \cdot \mathbf{v}) / c^2 = t - (\mathbf{r} \cdot \mathbf{v}) / c^2 + (\boldsymbol{\rho} \cdot \mathbf{v}) / c^2 = t' + (\boldsymbol{\rho} \cdot \mathbf{v}) / c^2, \tag{21}$$

and since $\mathbf{k}_i = \mathbf{k}_i' + (\omega_i' \mathbf{v}) / c^2$, we obtain in (19)

$$\mathbf{k}_i \cdot \boldsymbol{\rho} - \omega_i' \bar{t}' = \mathbf{k}_i' \cdot \boldsymbol{\rho} - \omega_i' t'. \tag{22}$$

Consequently, instead of κg in (14) we now deal with

$$k_x' \xi + k_z' \zeta + k_i' [(x' - \xi)^2 + (z' - \zeta)^2 + y'^2]^{1/2}. \tag{23}$$

In view of the identical structure, all results through (15) follow for the present stationary phase approximation, with properly primed symbols. For velocities $\mathbf{v}(\boldsymbol{\rho})$ contained in the xz plane there is no Doppler effect in the forward and specular directions. The velocity effect on the incident wave's amplitude (9) is cancelled by the velocity effect on the outgoing wave, in (19). However, not all velocity effects vanish, because in its comoving frame of reference, the object is excited by a wave (9), whose various parameters are velocity-dependent. To the first order in the velocity, in (19)

$$\mathbf{f}(\hat{\bar{\mathbf{r}}}', \hat{\mathbf{k}}_i', \omega_i', \hat{\mathbf{e}}_i') = \mathbf{f}|_{\mathbf{v}=0} + \left.\frac{\partial \mathbf{f}}{\partial \mathbf{v}}\right|_{\mathbf{v}=0} \cdot \mathbf{v}, \tag{24}$$

denoting an expression about $\mathbf{v} = 0$. The details of the symbolic form (24) may be very complicated. A simple manifestation of the velocity effect is the change in the excitation frequency; hence (24) will contain a term $\partial \mathbf{f} / \partial \omega_i' |_{\mathbf{v}=0} \mathbf{k}_i \cdot \mathbf{v}$. Since the present result contains no Doppler frequency shifts in the spectrum of the scattered field, the results can be compared with those of the Maxwell–Minkowski formalism (16),(17). However, in general, (19) contains more information, displaying the Doppler effects for various geometries and directions of observation. Nevertheless, it is shown subsequently that analytical solutions derived from the Maxwell–Minkowski formalism are relatively easy to obtain. The method is therefore of interest, as long as the criteria for its validity are met. In the next section these criteria are considered.

DISCUSSION AND GENERAL OBSERVATIONS

For many decades the subject of electrodynamics in the presence of moving systems has been considered to be of purely academic interest. Nowadays the probing of motion by means of electromagnetic radiation is becoming one of the most important tools for remote sensing, e.g., for measurement of wind velocity in the atmosphere, a key parameter in meteorology. With the advent of remote sensing methods from space platforms this subject becomes even more timely. It is well known that Doppler frequency shifts are amenable to accurate measurements, and the wide use of this method is evident from the literature. It is therefore important to know when the effect is detectible, and under what circumstances it is expected to vanish. On the other hand, detection of velocity effects in the absence of Doppler frequency shifts have been practically neglected thus far, but might become an important engineering method in the future. Recently,[20] one such method for wind velocity measurement has been proposed. The implementation of this particular method depends on the availability of lasers in space for remote sensing purposes. It is therefore worthwhile to derive solutions for various canonical problems, using the Maxwell–Minkowski model. Such problems are considered in subsequent sections.

The analysis of the slab region, given above, would suggest that Doppler effects vanish at high frequencies, since the scattered waves are in the forward and specular directions. However, this argument is somewhat oversimplified. Many problems can be defined where the motion is not parallel to the boundaries; yet the boundaries are at rest with respect to the observer. The idea of such flows seems to be nonphysical at a first glance. As an example, consider a perfectly conducting plane with an aperture, backed by a moving medium. For all practical purposes, the motion can be considered to terminate on the boundaries of the aperture. Such problems with a jump discontinuity in the velocity have been considered previously.[21] Depending on the geometry of the aperture and the frequency in a given direction, a far field radiation pattern exists for the energy scattered by the aperture. If the dimensions of the aperture are on the order of a wavelength, the stationary phase method cannot be used, as in the above discussion. Will there be Doppler effects, or can we discuss the problem in terms of the Maxwell–Minkowski formalism? Or consider the problem of a system with spherically stratified rotational velocity field—which method should be used? Doppler spectra are usually present when particles or irregularities are in motion. Intuition therefore suggests that Doppler effects will be present when there are strong fluctuations in the dielectric constant of the system, provided the typical length scales are not too small in comparison to wavelength, such that we have sufficient resolution. To put these ideas in a more concise form, scattering by random ensembles and the associated Wiener–Khintchine relations[18,22] must be considered. Essentially we define

$$\chi = \langle \chi \rangle + \chi_f, \tag{25}$$

where $\langle \chi \rangle$ is the average value of the susceptibility χ and χ_f is the fluctuating part, such that $\langle \chi_f \rangle = 0$. Hence

$$\langle \chi_f^2 \rangle = \langle \chi^2 \rangle - \langle \chi \rangle^2 \tag{26}$$

is proportional to the power scattering coefficient, describing the directional properties of the scattering process; according to the Wiener–Khintchine relations the power scattering coefficient is defined as the Fourier transform of the cross correlation function $(1/V) \int dV(\boldsymbol{\rho}) \chi(\boldsymbol{\rho}) \chi^*(\boldsymbol{\rho} + \mathbf{r})$. For long wavelength the scattering pattern is that of a dipole, and the power is proportional to $\langle \chi_f^2 \rangle$. However, there appears also a proportionality factor k^4 (typical to Rayleigh scattering); hence the scattered power is small and in this regime the Doppler effect is negligible, as assumed in the analysis of Van Bladel.[14] In the limit of short wavelengths the power still depends on $\langle \chi_f^2 \rangle$, and is now independent of frequency, but the scattering pattern is sharply peaked in the forward direction, for which the Doppler effect vanishes. This leaves a large domain where the effect of wavelength, geometry, and mean square fluctuation $\langle \chi_f^2 \rangle$ combine to produce significant Doppler effects. For such problems the Maxwell–Minkowski method is inapplicable. This is the regime where velocity profiles can be measured by means of the Doppler broadening spectrum.[23]

SIMPLE NONUNIFORM FLOW PROBLEMS

For the range of parameters where the Maxwell–Minkowski formalism applies, the velocity effects can be found by solving (6). Simple examples are given in the present and next sections. The simplest example is provided by a plane stratified nonuniformly moving region, where ϵ is identical throughout space, for which a scalar formalism is adequate. Later we consider a circular cylindrical region with circularly stratified motion along the generator. The simple case of a rigidly moving circular cylinder has been considered before.[4,24,25] The case of nonuniform motion is conveniently analyzed by solving (6). The presence of cross polarization effects, which has been noticed in the past,[25] is verified in the present solution. The problem of rotating cylindrical strata, involving nonuniform angular velocity is considered too. Special cases of uniform rotation are discussed and cited by Van Bladel.[13] The problem of the spherically stratified nonuniform flows, which is mathematically more complicated, is discussed in a separate section.

Plane stratified motion

The incident wave (7) is polarized in the $\hat{\mathbf{z}}$ direction, $\mathbf{e}_i = e_i \mathbf{z}$, and $\mathbf{k}_i$ is in the xy plane. Since there is no Doppler effect, $e^{-i\omega t}$ is suppressed. Throughout space, ϵ, μ_0 are constants. The motion $\mathbf{v} = \hat{\mathbf{x}} v(y)$ is confined to the region $-a \leqslant y \leqslant a$. For the present case (6) involves integrals of the form

$$\int_{-\infty}^{\infty} d\xi\, d\zeta\, d\eta\, \frac{e^{ikR}}{4\pi R} e^{ik_x \xi + ik_y \eta} v(\eta) = \frac{i}{2k_y} e^{ik_x x} \int_{-\infty}^{\infty} d\eta\, v(\eta) e^{ik_y(\eta + |y - \eta|)}. \tag{27}$$

It is noted that the integration of (27) is identical to the result obtained by using the leading term of the stationary phase approximation, as in (14). For the present case (6),(13) yield

$$\begin{aligned}\mathbf{E}_A &= -\mathbf{e}_i \omega \mu_0 (\epsilon - \epsilon_0) e^{ik_x x} \frac{ik_x}{k_y} \int_{-a}^{a} d\eta\, v(\eta) e^{ik_y(\eta + |y - \eta|)} \\ &= \mathbf{e}_A \left\{ e^{ik_x x + ik_y y} \int_{-a}^{y} d\eta\, v(\eta) \right. \\ &\quad \left. + e^{ik_x x - ik_y y} \int_{y}^{a} d\eta\, v(\eta) e^{i2k_y \eta} \right\},\end{aligned} \tag{28}$$

and $\mathbf{k}_{\pm} \times \mathbf{E}_A$ yields $\mathbf{H}_A$. The velocity-induced scattering produces two waves, propagating in the direction of incidence and specular reflection. In the region $y \geqslant a$ the first integral contributes $\int_{-a}^{a} d\eta\, v(\eta)$, and the second integral vanishes. The reflected wave contains the extra phase factor $e^{i2k_y \eta}$, and in the region $y \leqslant a$ the first integral vanishes, while the lower limit of the second integral is $-a$. Note that for a symmetrical flow $v(\eta) = v(-\eta)$, for example, the Poiseuille flow, there is no velocity effect in the $y \geqslant a$ region.

For practical applications, the question of remote sensing of the velocity profile must be considered. Obviously in the forward direction only an average effect is measured. The backscattered wave will be easier to measure, because of the absence of the incident wave in this direction. Furthermore, by extending the limits of the second integral to infinity it becomes a Fourier transform of the velocity field. By probing the moving region for a range of k_y, either by changing frequency or the direction of incidence, $v(\eta)$ can be found by performing the inverse Fourier transformation.

Cylindrical motion along the axis

Consider a radially symmetrical flow directed along the cylindrical axis

$$\mathbf{v} = \hat{\mathbf{z}} v(r). \tag{29}$$

The corresponding problems for rigidly moving circular cylinders have been analyzed[4,25] by using the relativistic transformations. The incident, scattered, and the zero order in v/c internal waves are given by

$$\begin{aligned}&\hat{\mathbf{z}} e_i e^{ikx} = \hat{\mathbf{z}} e_i \sum_{m=-\infty}^{\infty} i^m J_m(kr) e^{im\phi}, \\ &\hat{\mathbf{z}} e_i \sum_m i^m a_m H_m(kr) e^{im\phi}, \\ &\mathbf{E}_0 = \mathbf{z} e_i \sum_m i^m b_m J_m(\kappa r) e^{im\phi},\end{aligned} \tag{30}$$

where J_m denotes the Bessel functions, H_m are the Hankel functions of the first kind, and κ characterizes the internal domain. The coefficients a_m, b_m are computed by applying the boundary conditions to the zero order fields, i.e.,

$$J_m(ka) + a_m H_m(ka) = b_m J_m(\kappa a),$$

$$kJ'_m(ka) + kb_m H'_m(ka) = \kappa b_m J'_m(\kappa a), \tag{31}$$

where the prime denotes differentiation with respect to the argument. Hence a_m, b_m are considered to be known quantities. The two-dimensional Green function for this case[26]

$$G = \frac{i}{4\pi} H_0(\kappa|\mathbf{r} - \boldsymbol{\rho}|) = \sum_m e^{im(\phi - \nu)} \begin{cases} J_m(\kappa r) H_m(\kappa\rho) & r<\rho, \\ H_m(\kappa r) J_m(\kappa\rho) & r>\rho, \end{cases}$$

$$(\nabla^2 + \kappa^2) G = -\delta(\mathbf{r} - \boldsymbol{\rho}), \tag{32}$$

where $\boldsymbol{\rho} = \boldsymbol{\rho}(\rho,\nu), \mathbf{r} = \mathbf{r}(r,\phi)$. For the present case $\mathbf{j}_m = 0$ and $\mathbf{j}_e$ is in the cross-sectional plane. By inspection of (6) it becomes clear that $\mathbf{H}_A$ is in the $\hat{\mathbf{z}}$ direction; hence it is easier to compute. On the boundary $r = a, \mathbf{H}_A$ must be equal to the external $\mathbf{H}_A^e$ field in the $\hat{\mathbf{z}}$ direction, induced by the velocity. Since the incident wave has no field along the axis, the following boundary value problem is very simple. The velocity induced $\mathbf{E}_A^e$ field in the external domain can be derived from $\nabla \times \mathbf{H}_A^e = -i\omega\epsilon_e \mathbf{E}_A^e$, where ϵ_e pertains to this domain. The internal field $\mathbf{E}_A$ can be computed directly or by exploiting $\mathbf{E}_A^e$ and the Maxwell equation $\nabla \times \mathbf{H}_A + i\omega\epsilon E_A = \mathbf{j}_e$ inside the moving medium. From the definition of $\mathbf{j}_e$ and $\nabla \times \mathbf{E}_0 = i\omega\mu_0 \mathbf{H}_0$ we have

$$\mathbf{j}_e(\boldsymbol{\rho}) = -v(\rho) e_i (\epsilon - \epsilon_0) \times \sum_n \left(\frac{\hat{\mathbf{v}}}{\rho} \, in J_n(\kappa\rho) + \hat{\boldsymbol{\rho}} \kappa J'_n(\kappa\rho) \right) i^n b_n e^{in\nu}. \tag{33}$$

Before attempting to integrate (6), it must be noted that $\hat{\mathbf{v}}, \hat{\boldsymbol{\rho}}$ are functions of ν; hence they must be first expressed in terms of the constant Cartesian unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\sin\nu$, $\cos\nu$. By recasting $\sin\nu$, $\cos\nu$ in terms of $e^{i\nu}, e^{-i\nu}$ and adjusting the summation index of (33), we again derive a series in $e^{in\nu}$. In (6) we use the orthogonality relation $\int_0^{2\pi} d\nu \, e^{i(n-m)\nu} = \delta_{nm} 2\pi$ and then work backwards from $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ to $\hat{\mathbf{r}}, \hat{\boldsymbol{\phi}}$ representation. The total effect of this detour is identical to assuming $\hat{\mathbf{r}} = \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\phi}} = \hat{\mathbf{v}}$ in (33). Finally we obtain

$$\mathbf{H}_A(\mathbf{r}) = -\hat{\mathbf{z}} e_i \tfrac{1}{2} (\epsilon - \epsilon_0) \kappa \sum_m i^{m+1} m b_m F_m(r) e^{im\phi}, \quad 0 \leqslant r \leqslant a,$$

$$F_m(r) = J'_m(\kappa r) I_1(r) + H'_m(\kappa r) I_2(r) + J_m(\kappa r) I_3(r) + H_m(\kappa r) I_4(r),$$

$$I_1(r) = \int_r^a H_m(\kappa\rho) J_m(\kappa\rho) v(\rho) \, d\rho,$$

$$I_2(r) = \int_0^r J_m^2(\kappa r) v(\rho) \, d\rho, \tag{34}$$

$$I_3(r) = -\frac{1}{r} \int_r^a H_m(\kappa\rho) J'_m(\kappa\rho) v(\rho) \rho \, d\rho,$$

$$I_4(r) = -\frac{1}{r} \int_0^r J_m(\kappa\rho) J'_m(\kappa\rho) v(\rho) \rho \, d\rho.$$

On the boundary we have $I_1 = I_3 = 0$ and

$$\mathbf{H}_A(a) = -\mathbf{z} e_i \tfrac{1}{2} (\epsilon - \epsilon_0) \kappa \sum_m i^{m+1} m b_m F_m(a) e^{im\phi},$$

$$F_m(a) = H'_m(\kappa a) I_2(a) + H_m(\kappa a) I_4(a), \tag{35}$$

and I_2, I_4 follow from (34) by extending the upper limit to a. The continuity of the tangential $\mathbf{H}$ field at the boundary prescribes in the external domain

$$\mathbf{H}_A^e = \hat{\mathbf{z}} \sum_m i^m c_m H_m(kr) e^{im\phi}, \tag{36}$$

and by equating (35),(36) at $r = a$ the coefficients c_m are found. The mate $\mathbf{E}_A^e$ can be found from Maxwell's equations in the external domain; hence the solution is complete. From the point of view of measurements (35) is interesting because the velocity induced field is cross polarized with respect to the zero order fields; hence its detection and measurement is easier to perform.

Cylindrically stratified rotation

Another interesting class of problems is suggested by

$$\mathbf{v} = v(r) \hat{\boldsymbol{\phi}}, \tag{37}$$

describing a cylindrically stratified rotating system. Inasmuch as the rotation is not uniform, the problem cannot be approached by transforming to rotating coordinates.[13] For the present case we have

$$\mathbf{j}_e(\boldsymbol{\rho}) = \hat{\mathbf{z}} (\epsilon - \epsilon_0) e_i v(\rho) \sum_m \frac{im}{\rho} J_m(\kappa\rho) i^m b_m e^{im\nu},$$

$$\mathbf{j}_m(\boldsymbol{\rho}) = \hat{\boldsymbol{\rho}} i\omega\mu_0 (\epsilon - \epsilon_0) e_i v(\rho) \sum_m i^m b_m J_m(\kappa\rho) e^{im\nu}. \tag{38}$$

Using (32),(38) in (6) yields

$$\mathbf{E}_A(\mathbf{r}) = \hat{\mathbf{z}} \, i\omega\mu_0 (\epsilon - \epsilon_0) \frac{e_i}{2} \sum_m i^{m+1} m b_m F_m(r) e^{im\phi},$$

$$F_m(r) = J_m(\kappa r) I_1 + H_m(\kappa r) I_2,$$

$$I_1 = \int_r^a H_m(\kappa\rho) J_m(\kappa\rho) v(\rho) \left(1 + \frac{\rho}{r} \right) d\rho, \tag{39}$$

$$I_2 = \int_0^r J_m(\kappa r) v(\rho) \left(1 + \frac{\rho}{r} \right) d\rho.$$

At the boundary $\mathbf{E}_A(a)$ is obtained, and, by equating to the corresponding external field

$$\mathbf{E}_A^e = \hat{\mathbf{z}} e_i \sum_m i^m d_m H_m(kr) e^{im\phi} \tag{40}$$

at $r = a$, the coefficients d_m are determined. The mate $\mathbf{H}_A^e$ follows from the Maxwell equations. The direct direction of polarization is conserved in this problem.

SPHERICALLY STRATIFIED ROTATING MEDIA

We consider now the Mie theory for scattering by spherically stratified rotating media. A special case for a rigidly rotating sphere has been recently discussed by De Zutter.[27] In general the velocity is given by

$$\mathbf{v} = \Omega(r) \sin\theta \hat{\boldsymbol{\phi}}, \tag{41}$$

where $\Omega(r)$ is the angular velocity of a shell of radius r rotating about the polar axis $\hat{\mathbf{z}}$ of a spherical coordinate system. Here θ is the polar angle, measured off the $\hat{\mathbf{z}}$ axis, and $\hat{\boldsymbol{\phi}}$ is the azimuthal direction, with $\hat{\mathbf{r}},\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\phi}}$ defining a local right-handed Cartesian system. Again, this problem cannot be solved by transition to rotating coordinates because the rotation is nonuniform. Scattering by a sphere is a classical problem,[28] the notation used here is essentially as in Twersky,[29] who relates his to Morse and Feshbach[20] and others. The Green function for this geometry is

$$\tilde{\Gamma} = \frac{ik}{4\pi} \sum_{n=1}^{\infty} \sum_{m=-n}^{n} [\mathbf{M}_{nm}(\mathbf{r})\mathbf{M}^1_{n,-m}(\boldsymbol{\rho}) + \mathbf{N}_{nm}(\mathbf{r})N^1_{n,-m}(\boldsymbol{\rho})](-1)^m \frac{2n+1}{n(n+1)}, \quad r>\rho, \tag{42}$$

where $\mathbf{M},\mathbf{N}$ are defined in the Appendix, superscript 1 indicates that the vector spherical waves involve j_n the regular spherical Bessel functions; otherwise, the spherical Hankel functions h_n are used. For $r<\rho$ $\mathbf{r}$ and $\boldsymbol{\rho}$ in (42) are interchanged. The solution of the velocity independent part of the problem is described in detail by Stratton,[28] for example. In the interior domain $r<a$ the fields are given by

$$\mathbf{E}_0 = \sum_{n,m} [c_{nm}\mathbf{M}^1_{nm} + b_{nm}\mathbf{N}^1_{nm}],$$
$$\mathbf{H}_0 = \frac{k}{i\omega\mu_0} \sum_{n,m} [b_{nm}\mathbf{M}^1_{nm} + c_{nm}\mathbf{N}^1_{nm}], \tag{43}$$

where b_{nm},c_{nm} are assumed to be known coefficients and $j_n(\kappa r)$ involves κ, the parameter characterizing the interior of the sphere. In order to evaluate the source terms $\mathbf{j}_e,\mathbf{j}_m$, we need to know $\sin\hat{\boldsymbol{\phi}}\times\mathbf{N}^1_{nm}$. It is easy to see from the Appendix that $\sin\theta\hat{\boldsymbol{\phi}}\times\mathbf{C}^m_n = -im\,\mathbf{P}^m_n$; hence

$$\mathbf{v}\times\mathbf{M}^1_{nm} = -r\Omega(r)j_n(\kappa r)im\mathbf{P}^m_n(\hat{\mathbf{r}}) \equiv \alpha_1(n,m,r)\mathbf{P}^m_n(\hat{\mathbf{r}}). \tag{44}$$

Similarly, we wish to recast

$$\mathbf{v}\times\mathbf{N}^1_{nm} = [\Omega(r)/\kappa]\{n(n+1)j_n(\kappa r)\hat{\boldsymbol{\theta}}\sin\theta Y^m_n(\hat{\mathbf{r}}) - \partial_{\kappa r}[\kappa r j_n(\kappa r)]\hat{\mathbf{r}}\sin\theta\,\partial_\theta Y^m_n(\hat{\mathbf{r}})\}, \tag{45}$$

in terms of vector spherical harmonics. For the $\hat{\mathbf{r}}$ component we use recurrence relations for associated Legendre polynomials,[28] deriving

$$\hat{\mathbf{r}}\hat{\mathbf{r}}\cdot\mathbf{v}\times\mathbf{N}^1_{nm} = \alpha_2(n,m,r)\hat{\mathbf{r}}\cdot\mathbf{P}^m_{n-1} + \alpha_3(n,m,r)\hat{\mathbf{r}}\cdot\mathbf{P}^m_{n+1},$$
$$\alpha_2 = [\Omega(r)/\kappa]\,\partial_{\kappa r}[\kappa r j_n(\kappa r)](n+m)(n+1)/(2n+1),$$
$$\alpha_3 = -[\Omega(r)/\kappa]\,\partial_{\kappa r}[\kappa r j_n(\kappa r)](n-m+1)n/(2n+1). \tag{46}$$

The term $-\hat{\boldsymbol{\phi}}\sin\theta\,Y^m_n(\hat{\mathbf{r}})$ has been expressed in terms of $\mathbf{C}^m_n,\mathbf{B}^m_n$ before.[30] This is used to express the $\hat{\boldsymbol{\theta}}$ component of (45) in terms of vector spherical harmonics. The result is readily verified by using the definitions of the Appendix, recurrence relations for P^m_n and the differential equation satisfying p^m_n:

$$\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}\cdot\mathbf{v}\times\mathbf{N}^1_{nm} = \alpha_4(n,m,r)\mathbf{B}^m_{n-1} + \alpha_5(n,m,r)B^m_{n+1} + \alpha_6(n,m,r)\mathbf{C}^m_n,$$
$$\alpha_4 = [\Omega(r)/\kappa]n(n+1)j_n(\kappa r)(n+1)(n+m)/(2n+1),$$
$$\alpha_5 = -[\Omega(r)/\kappa]n(n+1)j_n(\kappa r)(n-m+1)n/(2n+1), \tag{47}$$
$$\alpha_6 = -[\Omega(r)/\kappa]n(n+1)j_n(\kappa r)im.$$

Therefore, the sources are given by

$$\mathbf{j}_m = i\omega\mu_0(\epsilon-\epsilon_0)\sum_{n,m}[\beta_1\mathbf{C}^m_n + \beta_2\mathbf{B}^m_n + \beta_3\mathbf{p}^m_n],$$
$$\beta_1(n,m,r) = b_{nm}\alpha_6(n,m,r),$$
$$\beta_2(n,m,r) = b_{n+1,m}\alpha_4(n+1,m,r) + b_{n-1,m}\alpha_5(n-1,m,r),$$
$$\beta_3(n,m,r) = c_{nm}\alpha_1(n,m,r) + b_{n+1,m}\alpha_2(n+1,m,r) + b_{n-1,m}\alpha_3(n-1,m,r). \tag{48}$$

Similarly

$$\mathbf{j}_e = -\kappa(\epsilon-\epsilon_0)\sum_{n,m}[\gamma_1\mathbf{C}^m_n + \gamma_2\mathbf{B}^m_n + \gamma_3\mathbf{P}^m_n], \tag{49}$$

where $\gamma_1,\gamma_2,\gamma_3$ are obtained from β_1,β_2,β_3 (48) by interchanging b_{nm} and c_{nm}.

We are now ready to represent $\mathbf{E}_\Lambda,\mathbf{H}_\Lambda$ in a closed, although highly compacted manner. In preparation for this we define

$$\mathbf{W}_{nm} = (-1)^m\frac{2n+1}{n(n+1)}(\mathbf{M}^1_{n,m} + \mathbf{N}^1_{n,m}),$$
$$\mathbf{U}^m_n = \gamma_1\mathbf{C}^m_n + \beta_2\mathbf{B}^m_n + \beta_3\mathbf{p}^m_n,$$
$$\mathbf{V}^m_n = \beta_1\mathbf{C}^m_n + \gamma_2\mathbf{B}^m_n + \gamma_3\mathbf{p}^m_n. \tag{50}$$

We now exploit the orthogonality relations (Appendix), and $\nabla\times\mathbf{M} = \kappa\mathbf{N}$, $\nabla\times\mathbf{N} = \kappa\mathbf{M}$, to obtain

$$\mathbf{E}_\Lambda = \frac{k^2\omega\mu_0}{4\pi}(\epsilon-\epsilon_0) \times\left\{\sum_{n,m}\mathbf{M}_{nm}(\mathbf{r})\int \mathbf{W}_{n,-m}(\boldsymbol{\rho})\cdot\mathbf{U}^m_n(\boldsymbol{\rho})\,dV(\boldsymbol{\rho}) + \mathbf{N}_{nm}(\mathbf{r})\int \mathbf{W}_{n,-m}(\boldsymbol{\rho})\cdot\mathbf{V}^m_n\,dV(\boldsymbol{\rho})\right\}. \tag{51}$$

By extending the integration limit to $\rho = a$, the field at $r=a$ is obtained. The detailed manipulations of (51), using orthogonality relations of the Appendix, reduce it to an integration over ρ. The details are not shown here, in the interest of saving space. A similar expression is obtained for $\mathbf{H}_\Lambda$. The solution (51) can be rewritten in the form

$$\mathbf{E}_\Lambda = \sum_{n,m}(e_{nm}\mathbf{M}_{nm} + f_{nm}\mathbf{N}_{nm}), \tag{52}$$

where the coefficients e_{nm},f_{nm} are derived by computing the integrals (51). The field at $r=a$ in directions $\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\phi}}$ is equated to the corresponding tangential components of $\mathbf{E}^e_\Lambda$,

$$\mathbf{E}^e_\Lambda = \sum_{n,m}(g_{nm}\mathbf{M}_{nm} + h_{nm}\mathbf{N}_{nm}), \tag{53}$$

in the external domain at the boundary. From this the coefficients g_{nm},h_{nm} are found. Using Maxwell's equation $\nabla\times\mathbf{E}^e_\Lambda = i\omega\mu_0\mathbf{H}^e_\Lambda$, the mate $\mathbf{H}^e_\Lambda$ field is derived. Therefore, the problem is solved.

Although (41) specializes the problem to rotation about the polar axis, arbitrary directions can be considered, using addition theorems for spherical vector waves, as given by Edmonds[31] and Stein.[32]

ACKNOWLEDGMENTS

This study was carried out while one author (D. Censor) was a National Academy of Sciences, NRC Associate at

NASA's Goddard Space Flight Center. He would like to express his gratitude for the hospitality he received there.

APPENDIX

Vector spherical waves and harmonics are defined:

$$\mathbf{M}_{nm}(\mathbf{r}) = h_n(kr)\mathbf{C}_n^m(\hat{\mathbf{r}}),$$

$$\mathbf{C}_n^m(\hat{\mathbf{r}}) = -\mathbf{r}\times\nabla Y_n^m(\hat{\mathbf{r}}) = \left(\hat{\boldsymbol{\theta}}\frac{\partial_\phi}{\sin\theta} - \hat{\phi}\partial_\theta\right)Y_n^m(\hat{\mathbf{r}}),$$

$$\mathbf{N}_{nm}(\mathbf{r}) = \{n(n+1)h_n(kr)\mathbf{P}_n^m(\mathbf{r}) + \partial_{kr}[krh_n(kr)]\mathbf{B}_n^m(\hat{\mathbf{r}})\}/kr,$$

$$\mathbf{P}_n^m(\hat{\mathbf{r}}) = \hat{\mathbf{r}}Y_n^m(\hat{\mathbf{r}}),$$

$$\mathbf{B}_n^m(\hat{\mathbf{r}}) = \hat{\mathbf{r}}\times\mathbf{C}_n^m(\hat{\mathbf{r}}) = \left(\hat{\boldsymbol{\phi}}\frac{\partial_\phi}{\sin\theta} + \hat{\boldsymbol{\theta}}\partial_\theta\right)Y_n^m(\hat{\mathbf{r}}),$$

$$Y_n^m(\hat{\mathbf{r}}) = P_n^m(\cos\theta)e^{im\phi},$$

$$Y_n^{-m} = (-1)^m[(n-m)!/(n+m)!]P_n^m(\cos\theta)e^{im\phi},$$

where ∂_ϕ denotes $\partial/\partial\phi$, etc., P_n^m are the associated Legendre functions, and h_m are the spherical Hankel functions.

Orthogonality relations:

$$\mathbf{P}_n^m\cdot\mathbf{B}_n^m = \mathbf{P}_n^m\cdot\mathbf{C}_n^m = \mathbf{B}_n^m\cdot\mathbf{C}_n^m = 0,$$

$$\int \mathbf{C}_n^{-m}\cdot\mathbf{C}_\nu^\mu d\Omega = \int \mathbf{B}_n^{-m}\cdot\mathbf{B}_\nu^\mu\, d\Omega = n(n+1)\int \mathbf{P}_n^{-m}\cdot\mathbf{P}_\nu^\mu\, d\Omega = (-1)^m 4\pi\delta_{n\nu}\delta_{m\mu}\frac{n(n+1)}{2n+1},$$

$$\int d\Omega = \int_0^{2\pi} d\phi \int_0^\pi \sin\theta\, d\theta.$$

[1] C. J. Doppler, "Über das farbige Licht der Doppelsterne und einiger anderer Gestirne des Himmels," Abh. Königlich Böhmischen Ges. Wiss. **2**, 467–482 (1842).

[2] A. Einstein, "Zur Elektrodynamik bewegter Körper," Ann. Phys. (Lpz.), **17**, 891–921 (1905).

[3] D. Censor, "Theory of the Doppler effect—fact, fiction and approximation," Proceedings of the URSI Conference on Electromagnetic Theory, Santiago de Compostela, Spain, 1983.

[4] D. M. Le Vine, "Scattering from a moving cylinder, Oblique incidence," Radio Sci. **8**, 497–504 (1973).

[5] D. Censor, "Scattering in velocity-dependent systems," Radio Sci. **7**, 331–7 (1972).

[6] H. Minkowski, "Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegter Körper," Göttingen Nachr. 53–116 (1908).

[7] A. Sommerfeld, *Electrodynamics* (Academic, New York, 1952).

[8] C. T. Tai, "A study of electrodynamics of moving media," Proc. IEEE **52**, 685–9 (1964).

[9] J. R. Collier and C. T. Tai, "Plane waves in a moving medium," Am. J. Phys. **33**, 166–7 (1965).

[10] D. Censor, "Propagation and scattering in radially flowing media," IEEE Trans. Microwave Theory Tech. **MTT-17**, 374–8 (1969).

[11] K. Nakagawa, "Scattering of a dipole field by a moving plasma column," IEEE Trans. Antennas Propagation **AP-30**, 76–82 (1982).

[12] D. Censor, "Interaction of electromagnetic waves with irrotational fluids," J. Franklin Inst. **293**, 117–29 (1972).

[13] J. Van Bladel, "Electromagnetic fields in the presence of rotating bodies," Proc. IEEE **64**, 301–318 (1976).

[14] J. Van Bladel, "Rotating dielectric sphere in low-frequency field," Proc. IEEE **67**, 1654–55 (1979).

[15] D. Censor, "Scattering of electromagnetic waves by nonuniform flows," Publications of the Department of Environmental Sciences, Tel Aviv University, ES-74-029, 1974.

[16] C. H. Papas, *Theory of Electromagnetic Wave Propagation* (McGraw-Hill, New York, 1965).

[17] C. T. Tai, reply to D. Yaghjian, "Comments on 'Electric dyadic Green's functions in the source region'," Proc. IEEE **69**, 282–5 (1981).

[18] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).

[19] V. Twersky, "On propagation in random media of discrete scatterers," Proc. Symp. Appl. Math. **16**, 84–116 (1964).

[20] D. Censor and D. M. Le Vine, "A proposed method for wind velocity measurement from space," NASA Technical Memorandum #82053, Nov. 1980.

[21] D. Censor, "Scattering of a plane wave at a plane interface separating two moving media," Radio Sci. 1079–88 (1969).

[22] A. M. Portis, *Electromagnetic Fields: Sources and Media* (Wiley, New York, 1978).

[23] D. Censor, "On Doppler broadening in velocity dependent random media," Israel J. Technol. **8**, 395–406 (1970).

[24] H. N. Kritikos, K. S. H. Lee, and C. H. Papas, "Electromagnetic reflectivity of non-uniform jet streams," Radio Sci. **2**, 991–5 (1967).

[25] D. Censor, "Scattering of electromagnetic waves by a cylinder moving along its axis," IEEE Trans. Microwave Theory Tech. **MTT-17**, 154–8 (1969).

[26] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).

[27] D. De Zutter, "Scattering by rotating dielectric sphere," IEEE Trans. Antennas Propagation **AP-28**, 643–51 (1980).

[28] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941).

[29] V. Twersky, "Multiple Scattering of Electromagnetic Waves by Arbitrary Configurations," J. Math. Phys. **8**, 589–610 (1967).

[30] D. Censor, "Scattering in velocity dependent systems," thesis (in Hebrew), Israel Institute of Technology, Haifa, 1967. See also Ref. 5 above.

[31] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, NJ, 1957).

[32] S. Stein, "Addition theorems for spherical wave functions," Q. Appl. Math. **19**, 15–24 (1961).

The J-matrix reproducing kernel: Numerical weights at the Harris energy eigenvalues

Hashim A. Yamani[a)]
Saudi Arabian National Center for Science and Technology, P.O. Box 6086, Riyadh—11442, Saudi Arabia

(Received 5 April 1983; accepted for publication 11 August 1983)

The restriction of the J-matrix scattering wave function to the subspace where the potential is nonzero is used to define a reproducing kernel in the energy parameters. The values of the kernel at the positive Harris energy eigenvalues are shown to be related to the numerical weights at these eigenvalues.

PACS numbers: 03.65.Nk, 02.60. + y

I. INTRODUCTION

It has long been known[1–4] that the partial wave kinetic energy operator $H^0 = -\frac{1}{2}d^2/dr^2 + l(l+1)/2r^2$ has a tridiagonal representation in the complete Slater[5] or oscillator basis set $\{|\phi_n\rangle\}_{n=0}^{\infty}$ with fixed scale parameter. This has meant that when the eigenvector solution $|\chi^0\rangle$ to the Schrödinger equation $J(E)|\chi^0\rangle = (H^0 - E)|\chi^0\rangle = 0$ is expanded in terms of the basis, i.e., $|\chi^0\rangle = \Sigma_{n=0}^{\infty} z_n|\phi_n\rangle$, the set $\{z_n\}_{n=0}^{\infty}$ satisfies the three-term recursion relation $J_{n,n-1}z_{n-1} + J_{n,n}z_n + J_{n,n+1}z_{n+1} = 0$. Due to the fact that this relation is one basic starting point of the theory of orthogonal polynomials, the tools of this theory have been exploited[4] to characterize the solution $|\chi^0\rangle$.

One immediate result is that two independent solutions $z_n = \tilde{s}_n$ and $z_n = \tilde{c}_n$ exist such that $\mathscr{S}(r) = \langle r|\chi_1^0\rangle = \Sigma_{n=0}^{\infty}\tilde{s}_n\langle r|\phi_n\rangle$ and $\mathscr{C}(r) = \langle r|\chi_2^0\rangle = \Sigma_{n=0}^{\infty}\tilde{c}_n\langle r|\phi_n\rangle$ behave asymptotically sinelike and cosinelike, respectively. Another result has been that if H^0 is restricted to the subspace spanned by the finite set $\{|\phi_n\rangle\}_{n=0}^{N-1}$, then a set of eigenvectors may be taken to be $\{|\widetilde{\Psi}(E_q^0)\rangle\}_{q=0}^{N-1}$ such that $|\widetilde{\Psi}^0(E_q^0)\rangle = \Sigma_{n=0}^{N-1}\tilde{s}_0(E_q^0)|\phi_n\rangle$, where $\{E_q^0\}_{q=0}^{N-1}$ are the N zeros of the function $\tilde{s}_N(E)$. Further, it has been shown[4] that $\{\langle\Psi^0(E_q^0)|\Psi^0(E_q^0)\rangle^{-1}\}$ is the set of numerical weights associated with the set of abscissa $\{E_q^0\}$. This fact has been utilized with success in conjunction with the Fredholm technique to do scattering calculations using a finite matrix representation of both the given potential V as well as the unperturbed Hamiltonian H^0.

On the other hand, the J-matrix[6,7] method works with a complete set of basis vectors $\{|\phi_n\rangle\}_{n=0}^{\infty}$, since H^0 can be solved exactly in the entire space. Only the given potential, however, needs to be restricted to the finite subspace U_N spanned by $\{|\phi_n\rangle\}_{n=0}^{N-1}$. If P_N is the projection operator[8] on U_N, then the eigenvector solution to the model Schrödinger equation, $(H^0 + P_N^\dagger VP_N)|\chi\rangle = E|\chi\rangle$, can be written as $|\chi\rangle = \Sigma_{n=0}^{N-1}T_n(E)|\phi_n\rangle + \Sigma_{n=N}^{\infty}(\tilde{s}_n + t\tilde{c}_n)/(1+t^2)^{1/2}|\phi_n\rangle$ with t being interpreted as the exact tangent of the phase shift caused by the potential $(P_{N_q}^\dagger VP_N)$. If $T_n(E)$ is defined to be $(\tilde{s}_n + t\tilde{c}_n)/(1+t^2)^{1/2}$ for $n \geqslant N$, then the N Harris[9] energy eigenvalues of the restricted full Hamiltonian $P_N^\dagger(H^0 + V)P_N$ are the N zeros, $\{E_q\}_{q=0}^{N-1}$, of $T_N(E)$. With this similarity between T_n and $\tilde{s}_n$ and the fact that $T_n(E) = \tilde{s}_n(E)$ when $V=0$, the question is asked [with $|\widetilde{\Psi}(E_q)\rangle = \Sigma_{n=0}^{N-1}T(E_q)|\phi_n\rangle$] if the set $\{\langle\widetilde{\Psi}(E_q)|\widetilde{\Psi}(E_q)\rangle^{-1}\}_{q=0}^{N}$ can be interpreted as the numerical weights associated with the set $\{E_q\}_{q=0}^{N-1}$.

In this paper we will answer in the positive the question posed above. In Sec. II we give a brief review of the J-matrix method to define the terms used subsequently. In Sec. III we establish that $\langle\widetilde{\Psi}(E)|\widetilde{\Psi}(E')\rangle$ is a reproducing kernel and investigate some of its special values and limit behavior. In particular, with the help of a suitably defined spectral function we show that $\langle\widetilde{\Psi}(E_q)|\widetilde{\Psi}(E_q)\rangle^{-1}$ can indeed be interpreted as the numerical weights associated with the positive energy subset of $\{E_q\}_{q=0}^{N-1}$. Finally, in conjunction with the "inverse" Fredholm determinant, we present in Sec. IV a numerical example using a one-term separable Yukawa potential[10] that possesses a bound state. We compare the weight computed using the J-matrix reproducing kernel with those obtained by the Heller derivative rule.[11] We compare the phase shift using the two sets of weights with the exact answer.

II. REVIEW OF THE J-MATRIX METHOD

This method employs[6,7] as a complete set of basis vectors, $\{|\phi_n\rangle\}_{n=0}^{\infty}$, either the Slater set

$$\phi_n(r) = \langle r|\phi_n\rangle = \zeta^{l+1}e^{-\zeta/2}L_n^{2l+1}(\zeta), \quad n = 0, 1, \ldots \tag{2.1}$$

or the oscillator set

$$\phi_n(r) = \langle r|\phi_n\rangle = \zeta^{l+1}e^{-\zeta^2/2}L_n^{l+1/2}(\zeta^2), \quad n = 0,1,\ldots, \tag{2.2}$$

where $\zeta = \lambda r$ and λ is a free scaling parameter. The set $\{|\bar{\phi}_n\rangle\}_{n=0}^{\infty}$ is defined as the dual basis, i.e.,

$$\langle\phi_n|\bar{\phi}_m\rangle = \langle\bar{\phi}_n|\phi_m\rangle = \delta_{nm}. \tag{2.3}$$

Given a short-range radial potential V, the J-matrix method finds the exact scattering solution to the model potential $\overline{V}$ derived by restricting V to the subspace U_N spanned by the finite set $\{|\phi_n\rangle\}_{n=0}^{N-1}$:

$$\overline{V}(r) = P_N^\dagger VP_N, \tag{2.4}$$

where

$$P_N = \sum_{n=0}^{N-1}|\phi_n\rangle\langle\bar{\phi}_n| \tag{2.5}$$

is the projection operator[8] on the subspace U_N. The adjoint operator has a similar interpretation in the dual space.

[a)] Present address: Physics Department, University of Petroleum and Minerals, Dhahran, Saudi Arabia.

 0022-2488/84/020317-06$02.50

The energy eigenvector $|\chi(E)\rangle$ which satisfies the Schrödinger equation

$$(H_0+\overline{V}-E)|\chi(E)\rangle=0 \tag{2.6}$$

has the representation

$$|\chi(E)\rangle=|\widetilde{\Psi}(E)\rangle+\sum_{n=N}^{\infty}\frac{1}{(1+t^2)^{1/2}}(\tilde{s}_n+t\tilde{c}_n)|\phi_n\rangle. \tag{2.7}$$

Here $\tilde{s}_n$ and $\tilde{c}_n$ are the components of the eigenvector of the free Hamiltonian which behave asymptotically sinelike and cosinelike, respectively; i.e.,

$$\mathscr{S}(r)\equiv\sum_{n=0}^{\infty}\tilde{s}_n(E)\phi_n(r)\underset{r\to\infty}{\sim}\sqrt{\frac{2}{\pi k}}\sin\Big(kr-\frac{l\pi}{2}\Big), \tag{2.8a}$$

$$\mathscr{C}(r)\equiv\sum_{n=0}^{\infty}\tilde{c}_n(E)\phi_n(r)\underset{r\to\infty}{\sim}\sqrt{\frac{2}{\pi k}}\cos\Big(kr-\frac{l\pi}{2}\Big). \tag{2.8b}$$

The vector $|\widetilde{\Psi}(E)\rangle$ is written as

$$|\widetilde{\Psi}(E)\rangle=\sum_{n=0}^{N-1}|\phi_n\rangle T_n(E) \tag{2.9}$$

with restriction on $T_n(E)$ so as to make $|\chi(E)\rangle$ normalized in the sense that

$$\langle\chi(E)|\chi(E')\rangle=\delta(E-E'). \tag{2.10}$$

With $t(E)$ as the tangent of the phase shift caused by $\overline{V}$, $\langle r|\chi(E)\rangle$ behaves asymptotically as

$$\langle r|\chi(E)\rangle\underset{r\to\infty}{\sim}\sqrt{\frac{2}{\pi k}}\sin\Big(kr-\frac{l\pi}{2}+\delta\Big). \tag{2.11}$$

The J-matrix method solves for the set of quantities $\{t(E), T_n(E)\}_{n=0}^{N-1}$ with the following explicit results[8,12,13]:

$$\text{(i)}\ t(E)=-\frac{\tilde{s}_{N-1}(E)+\bar{g}_{N-1,N-1}(E)J_{N-1,N}(E)\tilde{s}_N(E)}{\tilde{c}_{N-1}(E)+\bar{g}_{N-1,N-1}(E)J_{N-1,N}(E)\tilde{c}_N(E)}, \tag{2.12}$$

$$\text{(ii)}\ T_n(E)=-\bar{g}_{n,N-1}(E)J_{N-1,N}(E)T_N(E), \tag{2.13a}$$

where $J=H^0-E$ is the J-matrix, and

$$\bar{g}_{n,m}=\langle\bar{\phi}_n|[P_N^\dagger(J+V)P_N]^{-1}|\bar{\phi}_m\rangle, \tag{2.13b}$$

$$T_N(E)=(1+t^2(E))^{-1/2}(\tilde{s}_N(E)+t(E)\tilde{c}_N(E)). \tag{2.14}$$

More explicitly, if we define

$$|\Psi_q\rangle=\sum_{n=0}^{N-1}|\phi_n\rangle\Gamma_{nq} \tag{2.15}$$

such that

$$\langle\Psi_q|P_N^\dagger(J+V)P_N|\Psi_{q'}\rangle=E_q\delta_{qq'}, \tag{2.16}$$

then

$$\bar{g}_{nm}(E)=\sum_{q=0}^{N-1}\frac{\Gamma_{nq}\Gamma_{mq}}{E_q-E}. \tag{2.17}$$

We note in passing that T_N vanishes at the Harris energy eigenvalues, i.e.,

$$T_N(E_q)=0. \tag{2.18}$$

It can be seen that $|\widetilde{\Psi}(E)\rangle$ is the infinite-dimensional analog (with continuous energy parameter) to the finite-dimensional $|\Psi_q\rangle$. Furthermore, there is a corresponding relationship between $T_n(E)$ and $\Gamma_{n,l}$ which can be seen clearly by using Eq. (2.15) twice to first obtain

$$\frac{T_n(E)}{T_m(E)}=\frac{\bar{g}_{n,N-1}(E)}{\bar{g}_{m,N-1}(E)}, \tag{2.19}$$

then by taking the limit as $E\to E_q$, and using Eq. (2.13) to get

$$\frac{T_n(E_q)}{T_m(E_q)}=\frac{\Gamma_{n,q}}{\Gamma_{m,q}}. \tag{2.20}$$

This means that

$$\frac{\Gamma_{n,q}}{T_n(E_q)}=\frac{\Gamma_{m,q}}{T_m(E_q)}=d_q \tag{2.21}$$

is dependent only on q. This result has the interesting consequence that

$$|\Psi_q\rangle=d_q|\widetilde{\Psi}(E_q)\rangle. \tag{2.22}$$

It will be shown in Sec. III that d_q is related to the numerical weights at $E=E_q$. It is useful to finally note that the Green's matrix elements $\bar{g}_{nm}(E)$ of Eq. (2.17) can now be written as

$$\bar{g}_{nm}(E)=\sum_{q=0}^{N-1}d_q^2\frac{T_n(E_q)T_m(E_q)}{E_q-E}. \tag{2.23}$$

III. THE REPRODUCING KERNEL

Before we proceed to define the reproducing kernel, discuss its properties, and find its values in special cases, it is useful to introduce some quantities in the dual space. Corresponding to $T_n(E)$, which can be written

$$T_n(E)=\langle\bar{\phi}_n|\widetilde{\Psi}(E)\rangle=\langle\bar{\phi}_n|P_N|\chi(E)\rangle, \tag{3.1}$$

we define

$$\overline{T}_n(E)=\langle\widetilde{\Psi}(E)|\phi_n\rangle=\langle\chi(E)|P_N^\dagger|\phi_n\rangle. \tag{3.2}$$

It is then clear that

$$\overline{T}_n(E)=\sum_{m=0}^{N-1}T_m(E)S_{mn}, \tag{3.3}$$

where the overlap matrix S_{mn} is given by

$$S_{mn}=\langle\phi_m|\phi_n\rangle. \tag{3.4}$$

Also, corresponding to $\Gamma_{n,q}$, we define

$$\gamma_{mq}=\sum_{n=0}^{N-1}\Gamma_{nq}S_{mn}. \tag{3.5}$$

It then follows that

$$\sum_{q=0}^{N-1}\gamma_{mq}\Gamma_{nq}=\delta_{mn} \tag{3.6}$$

and

$$\sum_{m=0}^{N-1}\gamma_{mq}\Gamma_{mq'}=\delta_{qq'}, \tag{3.7}$$

signifying the orthogonality of the matrix which diagonalizes $[P_N^\dagger(J+V)P_N]$ in the subspace U_N.

A. Definition and properties of the reproducing kernel

The reproducing kernel $K_N(E,E')$ is defined as

$$K_N(E,E')=\langle\widetilde{\Psi}(E)|\widetilde{\Psi}(E')\rangle. \tag{3.8}$$

From the definition it is clear that $K_N(E, E')$ is symmetric; i.e.,

$$K_N(E, E') = K_N(E', E). \tag{3.9}$$

More explicitly, we have

$$K_N(E, E') = \sum_{n=0}^{N-1} \overline{T}_n(E) T_n(E'). \tag{3.10}$$

Using the definition for $\overline{T}_n$ and T_n, and the fact that P_N is idempotent, we immediately have

$$K_N(E, E') = \langle \chi(E) | P_N^\dagger P_N | \chi(E') \rangle, \tag{3.11}$$

which with the help of the orthogonality relation (2.10) yields the result

$$\lim_{N\to\infty} K_N(E, E') = \delta(E - E'). \tag{3.12}$$

To obtain a closed form expression for $K_N(E, E')$, we start with the expression (3.10) and use the explicit form of $\overline{T}_n(E)$ and relation (2.13a) to get

$$K_N(E, E') = [J_{N-1,N}(E)T_N(E)]\,[J_{N-1,N}(E')T_N(E')] \times \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \bar{g}_{m,N-1}(E) S_{mn} \bar{g}_{n,N-1}(E'). \tag{3.13}$$

Further simplification of Eq. (3.13) results from using the explicit expression (2.17) for $g_{n,m}(E)$ and the orthogonality relation (3.7) to obtain

$$K_N(E, E') = [J_{N-1,N}(E)T_N(E)]\,[J_{N-1,N}(E')T_N(E')] \times \sum_{q=0}^{N-1} \frac{\Gamma^2_{N-1,q}}{(E_q - E)(E_q - E')}. \tag{3.14}$$

The denominator in the sum can be factored, resulting in

$$K_N(E, E') = [J_{N-1,N}(E)T_N(E)]\,[J_{N-1,N}(E')T_N(E')] \times [\bar{g}_{N-1,N-1}(E) - \bar{g}_{N-1,N-1}(E')]/(E - E').$$

This is the first closed form expression for $K_N(E, E')$. Another form results from using Eq. (2.13a). Thus

$$K_N(E, E') = [T_{N-1}(E')J_{N-1,N}(E)T_N(E) - T_{N-1}(E)J_{N-1,N}(E')T_N(E')]/[E - E']. \tag{3.15}$$

This is the equivalent expression to the Christoffel–Darboux formula usually encountered in the theory of orthogonal polynomials.

B. Special cases of $K_N(E, E')$

Using the expression derived in the previous subsection and taking the appropriate limit when necessary we can easily derive the following special values for $K_N(E, E')$:

(i) $K_N(E, E) = W(T_{N-1}(E), J_{N-1,N}(E)T_N(E))$, $\quad$ (3.16)

where W is the Wronskian. On the other hand, Eq. (3.14) yields

$$K_N(E, E) = [J_{N-1,N}(E)T_N(E)]^2 \sum_{q=0}^{N-1} \frac{\Gamma^2_{N-1,q}}{(E - E_q)^2} \tag{3.17}$$

which is a positive definite quantity. Furthermore, noting Eq. (2.13a), we may write

$$K_N(E, E) = \left[\frac{T_{N-1}(E)}{g_{N-1,N-1}(E)}\right]^2 \sum_{q=0}^{N-1} \frac{\Gamma^2_{N-1,q}}{(E - E_q)^2}; \tag{3.18}$$

(ii) $K_N(E, E_q) = T_{N-1}(E_q) J_{N-1,N}(E) T_N(E)/(E - E_q)$; $\quad$ (3.19)

(iii) $K_N(E_q, E_{q'}) = 0 \quad \text{if } E_q \neq E_{q'}$; $\quad$ (3.20)

(iv) $K_N(E_q, E_q) = T_{N-1}(E_q) J_{N-1,N}(E_q) \times \left(\frac{\partial}{\partial E} T_N(E)\right)_{E=E_q}$, $\quad$ (3.21)

which results from (3.19) by taking the appropriate limit. Furthermore by taking the limit of (3.18) as $E \to E_q$ we obtain

$$K_N(E_q, E_q) = T^2_{N-1}(E_q)/\Gamma^2_{N-1,q}, \tag{3.22}$$

which can be cast in the form

$$K_N(E_q, E_q) = |\langle \bar{\phi}_{N-1} | \tilde{\Psi}(E_q) \rangle|^2 / |\langle \bar{\phi}_{N-1} | \Psi_q \rangle|^2. \tag{3.23}$$

This has the interesting interpretation that $K_N(E_q, E_q)$ is the ratio squared of the component of $|\tilde{\Psi}(E_q)\rangle$ to that of $|\Psi_q\rangle$ at the "boundary" of the subspace U_N. Result (3.22) can be made more general due to relation (2.20). In fact,

$$K_N(E_q, E_q) = \frac{T_n(E_q)}{\Gamma_{n,q}} \frac{T_m(E_q)}{\Gamma_{m,q}} \tag{3.24}$$

is independent of any $m,n \leqslant N-1$.

With the definition

$$\omega_q = 1/K_N(E_q, E_q), \tag{3.25}$$

we may write the relation (2.22) as

$$|\Psi_q\rangle = \sqrt{\omega_q}\, |\tilde{\Psi}(E_q)\rangle. \tag{3.26}$$

In the Sec. IIID we will explain the sense in which $\{\omega_q\}$ can be interpreted as the numerical weight associated with the set of Harris energy eigenvalues $\{E_q\}$.

C. The spectral function and reproducing property

As a prelude to proving the reproducing property of $K_N(E, E')$ we define the spectral function[14] $\sigma_N(E)$ as a step function which is constant except at E_q, and whose jump at E_q is ω_q. More specifically,

$$\sigma_N(E) = \begin{cases} \sum_{0 < E_q < E} \omega_q & \text{for } E \geqslant 0, \\ -\sum_{E < E_q < 0} \omega_q & \text{for } E < 0. \end{cases}$$

Hence,

$$\frac{d\sigma_N(E)}{dE} = \sum_{q=0}^{N-1} \omega_q \delta(E - E_q). \tag{3.27}$$

We first prove two lemmas.

Lemma 1: The sets $\{T_n(E)\}_{n=0}^{N-1}$ and $\{\overline{T}_n(E)\}_{n=0}^{N-1}$ are mutually orthogonal in the sense that

$$\int T_n(E) \overline{T}_m(E) d\sigma_N(E) = \delta_{nm}. \tag{3.28}$$

Proof:

$$\int T_n(E)\overline{T}_m(E)d\sigma_N(E) = \sum_{q=0}^{N-1} \omega_q T_n(E_q)\overline{T}_m(E_q).$$

Writing ω_q as $\Gamma_{n,q}\Gamma_{m,q}/T_n(E_q)T_m(E_q)$, and noting the definition (3.3) of $\overline{T}_m(E)$ and the orthogonality property (3.6), we find that

$$\int T_n(E)\overline{T}_m(E)d\sigma_N = \sum_{q=0}^{N-1} \Gamma_{nq}\gamma_{mq} = \delta_{nm}.$$

Lemma 2:

$$\int K_N(E, E')d\sigma_N(E') = 1. \tag{3.29}$$

Proof: This is a special case of the reproducing property

$$\int K_N(E, E')d\sigma_N(E') = \sum_{q=0}^{N-1} \omega_q K_N(E, E_q).$$

Using the special case (3.9), we write

$$\int K_N(E, E')d\sigma_N(E')$$
$$= \sum_{q=0}^{N-1} \omega_q \frac{T_{N-1}(E_q)}{E-E_q} J_{N-1,N}(E)T_N(E)$$
$$= 1.$$

The last equality is verified in Appendix A.

Now, we are ready to prove the reproducing property.

Lemma 3:

$$\int K_N(E, E')f(E')d\sigma_N(E') = f(E) \tag{3.30}$$

provided, for some $\{c_n\}_{n=0}^{N-1}$, $f(E)$ can be represented as

$$f(E) = \sum_{n=0}^{N-1} c_n T_n(E). \tag{3.31}$$

Proof: It suffices to take $f(E) = T_m(E)$ for some $m \leqslant N-1$. Then with the help of (3.19) we have

$$\int K_N(E, E')T_m(E')d\sigma_N(E')$$
$$= \sum_{q=0}^{N-1} \omega_q K_N(E, E_q)T_m(E_n)$$
$$= \sum_{q=0}^{N-1} \left\{\frac{T_{N-1}(E_q)}{E-E_q} J_{N-1,N}(E)T_N(E)\right\} T_m(E_q).$$

Recalling the definition of $\overline{g}_{m,N-1}(E)$ and the relation (2.13a), we have

$$\int K_N(E, E')T_m(E')d\sigma_N(E')$$
$$= -g_{m,N-1}(E)J_{N-1,N}(E)T_N(E) = T_m(E).$$

D. Numerical weights associated with the Harris eigenvalues $\{E_q\}_{q=0}^{N-1}$

When $\overline{V} = 0$, the resulting set $\{\omega_q^0\}_{q=0}^{N-1}$, has already been shown[4] to be the numerical weights associated with free Hamiltonian energy eigenvalues $\{E_q^0\}_{q=0}^{N-1}$. Furthermore, the full Green's operator

$$g(z) = \sum_b \frac{|\chi_b\rangle\langle\chi_b|}{E_b - z} + \int_0^\infty \frac{|\chi(E))\langle\chi(E)|}{E-Z} dE \tag{3.32}$$

has the matrix elements

$$g_{n,m}(z) = \langle\overline{\phi}_n|g(z)|\overline{\phi}_m\rangle$$
$$= \sum_b \frac{\langle\overline{\phi}_n|\chi_b\rangle\langle\chi_b|\overline{\phi}_m\rangle}{E_b - z}$$
$$+ \int_0^\infty dE \frac{\langle\overline{\phi}_n|\chi(E))\langle\chi(E)|\overline{\phi}_m\rangle}{E-z}, \tag{3.33}$$

where the sum is over bound states of the full Hamiltonian $(H^0 + P_N^\dagger VP_N)$.

If $n, m \leqslant N-1$, then

$$g_{n,m}(z) = \sum_b \frac{\langle\overline{\phi}_n|\chi_b\rangle\langle\chi_b|\overline{\phi}_m\rangle}{E_b - z}$$
$$+ \int_0^\infty dE \frac{T_n(E)T_m(E)}{E-z} dE. \tag{3.34}$$

For z complex and away from the real energy axis the integral can be approximated by a quadrature. If the set $\{E_q\colon E_q > 0\}$ is chosen as the set of abscissas, and $\{\Omega_q\}$ is the associated set of numerical weights, then we may write

$$\int_0^\infty dE \frac{T_{n(E)}T_{m(E)}}{E-z} \simeq \sum_{E_q>0} \Omega_q \frac{T_{m(E_q)}T_{m(E_q)}}{E_q - z}. \tag{3.35}$$

On the other hand, from (2.23), (3.24), and (3.25), we have

TABLE I. Numerical weights $\{\tilde{\omega}_q\}$ associated with the abscissas $\{E_q\}$ calculated analytically compared with the weights $\{\tilde{\Omega}_q\}$ as calculated by the Heller derivative rule. Results are given for the s-wave Hamiltonian with potential $V = -2\pi|u><u|$, $u(r) = e^{-r}/r$, and Slater basis with $N = 10$ and a scale parameter $\lambda = 3.2$ is used.

q	E_q	x_q	$\tilde{\Omega}_q$	$\tilde{\omega}_q$
0	−0.113 49E + 01	—	—	—
1	0.387 04E − 01	−0.941 30E + 00	0.117 44E + 00	0.115 59E + 00
2	0.161 94E + 00	−0.775 38E + 00	0.211 69E + 00	0.211 83E + 00
3	0.395 70E + 00	−0.527 72E + 00	0.278 13E + 00	0.278 09E + 00
4	0.800 84E + 00	−0.230 27E + 00	0.311 22E + 00	0.311 25E + 00
5	0.151 45E + 01	0.839 24E − 01	0.311 95E + 00	0.311 93E + 00
6	0.287 40E + 01	0.383 72E + 00	0.283 17E + 00	0.283 11E + 00
7	0.587 17E + 01	0.642 04E + 00	0.230 03E + 00	0.230 04E + 00
8	0.145 38E + 02	0.838 16E + 00	0.160 18E + 00	0.160 06E + 00
9	0.615 75E + 02	0.959 27E + 00	0.814 83E − 01	0.813 27E − 01

TABLE II. S-wave phase shift for scattering from the separable potential $V = -2\pi|u\rangle\langle u|$ with $u(r) = e^{-r}/r$. The inverse Fredholm technique is used utilizing $\tilde{\omega}_q$ and $\tilde{\Omega}_q$ of Table I, respectively. The results are compared with the J-matrix phase shift in the same approximation, namely, Slater basis with $N = 10$ and $\lambda = 3.2$, and also with the exact result.

k (a.u.)	Phase shift using $\tilde{\Omega}_q$	Phase shift using $\tilde{\omega}_q$	J-matrix phase shift	Exact phase shift
0.25	0.256 436E + 01	0.255 988E + 01	0.255 842E + 01	0.255 838E + 01
0.50	0.203 266E + 01	0.203 534E + 01	0.203 550E + 01	0.203 550E + 01
1.00	0.126 201E + 01	0.126 257E + 01	0.126 267E + 01	0.126 263E + 01
2.00	0.520 271E + 00	0.520 419E + 00	0.520 510E + 00	0.520 456E + 00
3.00	0.245 732E + 00	0.245 805E + 00	0.245 852E + 00	0.245 815E + 00
4.00	0.130 409E + 00	0.130 410E + 00	0.130 446E + 00	0.130 415E + 00
5.00	0.758 118E − 01	0.758 446E − 01	0.758 594E − 01	0.758 509E − 01

$$\bar{g}_{nm}(z) = \sum_{E_q<0} \omega_q \frac{T_{n(E_q)} T_{m(E_q)}}{E_q - z} + \sum_{E_q>0} \omega_q \frac{T_{n(E_q)} T_{m(E_q)}}{E_q - z}; \tag{3.36}$$

comparing this with (3.35), we conclude that the set $\{\omega_q: E_q > 0\}$ is the numerical weights associated with the set of abscissas $\{E_q: E_q \geqslant 0\}$.

IV. AN EXAMPLE USING THE INVERSE FREDHOLM DETERMINANT

To test numerically the results of the previous section, we utilize the L^2-Fredholm technique.[3,4] Unlike the usual treatment which starts with the Fredholm determinant

$$D(z) = \det\left(\frac{z-H}{z-H^0}\right), \tag{4.1}$$

we here start with the inverse Fredholm determinant

$$D^{-1}(z) = \det\left(\frac{z-H^0}{z-H}\right). \tag{4.2}$$

It is easy to show that, except for the introduction of the bound states, the inverse determinant $D^{-1}(z)$ has similar properties to $D(z)$ itself. In particular,

(i) $D^{-1}(E+i\epsilon) = |D^{-1}(E)|e^{+i\delta(E)} \equiv B(E) - i\pi A(E)$, (4.3)

(ii) $$D^{-1}(z) = 1 + \sum_b \frac{a_b}{z-E_b} + \int_0^\infty \frac{A(E)}{z-E}\,dE, \tag{4.4}$$

where the sum is over the bound states of H, and a_b are the residues of $D^{-1}(z)$ at bound-state energies.

(iii) The real and imaginary parts of $D^{-1}(E+i\epsilon)$ are related by the dispersion relation

$$B(E) = 1 + \sum_b \frac{a_b}{E-E_b} + \mathscr{P}\int_0^\infty \frac{A(E')}{E-E'}\,dE'. \tag{4.5}$$

An approximate determinant $D^{-1}_{\text{approx}}(z)$ arising from diagonalization of both H^0 and H in the finite basis $\{|\phi_n\rangle\}_{n=0}^{N-1}$ (resulting in the eigenvalues $\{E_q^0\}_{q=0}^{N-1}$ and $\{E_q\}_{q=0}^{N-1}$) will have the form

$$D^{-1}_{\text{approx}}(z) = \prod_{q=0}^{N-1}\left(\frac{z-E_q^0}{z-E_q}\right) \tag{4.6}$$

$$= 1 + \sum_{E_q<0} \frac{\tilde{a}_q}{z-E_q} + \sum_{E_q>0} \frac{\tilde{b}_q}{z-E_q}, \tag{4.7}$$

where the last sum in (4.7) gives the quadrature approximation in the integral of (4.4). We now use the quadrature weights ω_q derived in Sec. III. Using the transformation

$$x = \frac{E - \lambda^2/8}{E + \lambda^2/8} \tag{4.8}$$

and making the definition

$$\tilde{\omega}_q = \left(\frac{dx}{dE}\right)_{E=E_q} \omega_q, \tag{4.9}$$

we then have

$$B(E) = \prod_{q=0}^{N-1}\left(\frac{E-E_q^0}{E-E_q}\right) + \frac{A^{\text{approx}}(E)(dE/dx)}{F(x)} \times\left(\mathscr{P}\int_{-1}^{+1} \frac{F(x')dx'}{E-E(x')} - \sum_{E_q>0} \frac{\tilde{\omega}_q F(x_q)}{E-E_q}\right), \tag{4.10}$$

and $A^{\text{approx}}(E)$ is obtained by interpolating the values $\{A(E_q): E_q > 0\}$, where

$$A^{\text{approx}}(E_q) = \tilde{b}_q/\omega_q. \tag{4.11}$$

For convenience, we have taken[15] $F(x)$ to be $(1-x^2)^{1/2}$. The example we have chosen for the potential is the one-term separable Yukawa potential

$$V = -2\pi|u\rangle\langle u|, \tag{4.12}$$

with

$$u(r) = \langle r|u\rangle = e^{-r}/r. \tag{4.13}$$

We have used the Slater basis with $\lambda = 3.2$ and $N = 10$. In Table I we give $\tilde{\omega}_q$ and compare it with the Heller derivative[11] weights $\tilde{\Omega}_q$. These are obtained by interpolating x_q (corresponding to $E_q \geqslant 0$) in q such that

$$[x(\eta)]_{\eta=q} = x_q. \tag{4.14}$$

Then

$$\tilde{\Omega}_q = \left[\frac{dx(\eta)}{d\eta}\right]_{\eta=q}. \tag{4.15}$$

We see that the agreement is excellent.

In Table II we calculate $\delta(E)$ for several values of the energy using both $\tilde{\omega}$ and $\tilde{\Omega}$, and compare them with the J-matrix $\delta(E)$ of Eq. (2.12) and the exact result.[10]

APPENDIX A

Since

$$T_{N-1}(E) = -g_{N-1,N-1}(E)J_{N-1,N}(E)T_N(E) \tag{A1}$$

and

$$\bar{g}_{N-1,N-1}(E)=\sum_{q=0}^{N-1}\omega_q\,\frac{T_{N-1}^2(E_q)}{E_q-E}, \qquad \text{(A2)}$$

we can write

$$\bar{g}_{N-1,N-1}(E)=T_{N-1}(E)\left[\sum_{q=0}^{N-1}\frac{f_q}{E_q-E}\right]. \qquad \text{(A3)}$$

To solve for f_q, we cast (A3) in the form

$$T_{N-1}(E)=\bar{g}_{N-1,N-1}(E)\left[\sum_{q=0}^{N-1}\frac{f_q}{E_q-E}\right], \qquad \text{(A4)}$$

and then take the limit $E\rightarrow E_q$ to obtain

$$T_{N-1}(E_q)=\omega_q T_{N-1}^2(E_n)/f_q$$

or

$$f_q=\omega_q T_{N-1}(E_q). \qquad \text{(A5)}$$

Inserting the result back into (A3), we have

$$\bar{g}_{N-1,N-1}(E)=T_{N-1}(E)\sum_{q=0}^{N-1}\omega_q\,\frac{T_{N-1}(E_q)}{E_q-E}, \qquad \text{(A6)}$$

and using (A1) again for $T_{N-1}(E)$, we finally get

$$1=J_{N-1,N}(E)T_N(E)\sum_{q=0}^{N-1}\omega_q\,\frac{T_{N-1}(E_q)}{E-E_q}. \qquad \text{(A7)}$$

[1]C. Schwartz, Ann. Phys. (N.Y.) **16**, 36 (1961).
[2]M. B. Hidalgo, J. Nuttall, and R. W. Stagat, J. Phys. B **6**, 1364 (1973).
[3]E. J. Heller, T. N. Rescigno, and W. P. Reinhardt, Phys. Rev. A **8**, 2946 (1973).
[4]H. A. Yamani and W. P. Reinhardt, Phys. Rev. A **11**, 1144 (1975).
[5]The reference Hamiltonian H^0 containing the Coulomb term z/r is tridiagonal in the Slater basis. See Ref. 4.
[6]E. J. Heller and H. A. Yamani, Phys. Rev. A **9**, 1201 (1974).
[7]H. A. Yamani and L. Fishman, J. Math. Phys. **16**, 1144 (1975).
[8]H. A. Yamani, J. Math. Phys. **23**, 83 (1982).
[9]F. E. Harris, Phys. Rev. Lett. **19**, 173 (1967).
[10]K. Gottfried, *Quantum mechanics* (Benjamin, New York, 1966), Vol. I, p. 161.
[11]E. J. Heller, thesis, Harvard University, 1973 (unpublished). A discussion of Heller's derivative rule is given in Ref. 4; a useful treatment of the rule as it applies both to the set $\{E_q^0\}$ and $\{E_q\}$ is given in Ref. 12.
[12]J. Broad, Phys. Rev. A **18**, 287 (1978).
[13]E. J. Heller, Phys. Rev. A **12**, 1222 (1975).
[14]F. V. Atkinson, *Discrete and Continuous Boundary Problems* (Academic, New York, 1964), p. 106.
[15]See Ref. 3 for discussion of the condition that $F(x)$ has to satisfy.

Coulomb-modified nuclear scattering. III

B. Talukdar and D. K. Ghosh
Department of Physics, Visva-Bharati University, Santiniketan-731235, West Bengal, India
T. Sasakawa
Department of Physics, Tohoku University, Sendai 980, Japan

(Received 11 February 1983; accepted for publication 26 August 1983)

Scattering by the Coulomb plus Graz separable potential is studied by employing a coordinate space approach to the problem. Exact analytical expressions for on- and off-shell Jost functions $f_l(k)$ and $f_l(k,q)$ are constructed and certain useful checks are made with regard to their limiting behavior and on-shell discontinuity.

PACS numbers: 03.65.Nk, 25.10. + s

I. INTRODUCTION

In two recent papers[1,2] (hereafter referred as papers I and II) we developed a method to treat the problem in the title, which does not make explicit use of the two-potential theorem.[3] Concentrating on the s-wave problem we solved the associated radial Schrödinger equation for some useful Coulomb plus nucleon-nucleon interactions with regular boundary condition[4] and constructed expressions for on- and off-shell Jost functions[4–7] by making use of certain integral representations for the latter.

In the present paper we extend the results of papers I and II to higher partial waves by working with a form of the Graz potential[8] in uncoupled partial waves as used by van Haeringen and Kok.[9] In Sec. II we solve the Schrödinger equation for the Coulomb plus Graz potential and use this result in Sec. III to construct expressions for on- and off-shell Jost functions. Finally, in Sec. IV we present some concluding remarks.

II. REGULAR SOLUTION FOR COULOMB PLUS GRAZ POTENTIAL

In the representation space the rank-1 Graz separable potential[8] is written as

$$V_l(r,r') = -\lambda_l 2^{-2l}(l!)^{-2}(rr')^{l-1}e^{-\beta(r+r')}. \tag{1}$$

The radial Schrödinger equation at a center of mass energy $E = k^2$ for the Coulomb plus potential in (1) is

$$\left[\frac{d^2}{dr^2} + k^2 - \frac{2\eta k}{r} - \frac{l(l+1)}{r^2}\right]\phi_l(k,r) = d_l(k)r^l e^{-\beta r} \tag{2}$$

with

$$d_l(k) = -\lambda_l 2^{-2l}(l!)^{-2}\int_0^\infty s^l e^{-\beta s}\phi_l(k,s)ds, \tag{3}$$

and η the well-known Sommerfeld parameter. To solve (2) for the regular boundary condition we change the variable by substituting

$$\phi_l(k,r) = r^{l+1}e^{ikr}g_l(k,r),\qquad r = -z/2ik, \tag{4}$$

and get

$$z\frac{d^2g_l(z)}{dz^2} + \{2(l+1) - z\}\frac{dg_l(z)}{dz} - (l+1+i\eta)g_l(z) = -\frac{d_l(k)}{2ik}e^{\rho z} \tag{5}$$

with

$$\rho = (\beta + ik)/2ik. \tag{6}$$

Applying the Laplace transform method of paper I in (5) we obtain the regular solution $\phi_l(k,r)$ in the form

$$\phi_l(k,r) = r^{l+1}e^{ikr}\Phi(l+1+i\eta,2l+2;-2ikr) - \frac{d_l(k)}{2ik}r^{l+1}e^{ikr}\sum_{n=1}^{\infty}\frac{\rho^{n-1}}{(n-1)!}\theta_n \times (l+1+i\eta,2l+2;-2ikr). \tag{7}$$

The result for $d_l(k)$ is determined[10] by substituting (7) in (3), and we have

$$d_l(k) = -\lambda_l\frac{2^{-2l}(l!)^{-2}\Gamma(2l+2)}{D_l(k)(\beta^2+k^2)^{l+1}}e^{2\eta y} \tag{8}$$

with $y = \arctan k/\beta$. Here

$$D_l(k) = 1 - \lambda_l 2^{-2l}(l!)^{-2}(\beta^2+k^2)^{-1}(\beta-ik)^{-2l-1} \times \sum_{n=1}^{\infty}(-1)^n\frac{(2l+n)!}{n!}\left(\frac{\beta+ik}{\beta-ik}\right)^n \times {}_2F_1\left(1,l+1+n+i\eta;n+1;\frac{-2ik}{\beta-ik}\right) \tag{9}$$

is the Fredholm determinant associated with the regular solution for the Coulomb plus Graz potential. Thus

$$\phi_l(k,r) = r^{l+1}e^{ikr}\Phi(l+1+i\eta,2l+2;-2ikr) + \lambda_l\frac{2^{-2l}(l!)^{-2}\Gamma(2l+2)}{2ikD_l(k)(\beta^2+k^2)^{l+1}}e^{2\eta y}r^{l+1}e^{ikr} \times \sum_{n=1}^{\infty}\frac{\rho^{n-1}}{(n-1)!}\theta_n(l+1+i\eta,2l+2;-2ikr). \tag{10}$$

The function $\theta_\sigma\,(a,c;z)$ has been defined in (20) of paper I.

III. JOST FUNCTIONS

For the lth partial wave the on-shell Jost function is obtained from the integral representation[4–6]

$$f_l(k) = f_l^c(k) + \frac{k^l e^{-i\pi l/2}}{(2l+1)!!} \times \int_0^\infty \int_0^\infty rr'\, dr\, dr'\, f_l^c(k,r) V_l(r,r') \times \phi_l(k,r'), \quad (11)$$

where

$$f_l^c(k) = e^{\pi\eta/2}\Gamma(l+1)/\Gamma(l+1+i\eta) \quad (12)$$

stands for the Coulomb Jost function. The Coulomb Jost solution is given by[4]

$$f_l^c(k,r) = -(2kr)^{l+1} i e^{\pi\eta/2} e^{i(kr - l\pi/2)} \times \Psi(l+1+i\eta, 2l+2; -2ikr) \quad (13)$$

with $\Psi(a,c;z)$ an irregular confluent hypergeometric function. Substituting (1), (10), and (13) in (11) we get

$$f_l(k) = f_l^c(k) - \frac{k^l e^{-i\pi l}}{(2l+1)!!} d_l(k)(2k)^{l+1} i e^{\pi\eta/2} \times \int_0^\infty r^{2l+1} e^{-(\beta - ik)r} \times \Psi(l+1+i\eta, 2l+2; -2ikr) dr. \quad (14)$$

The integral in (14) can be evaluated by using[11]

$$\int_0^\infty e^{-ax} x^{s-1} \Psi(b,d;\mu x) dx = \frac{\Gamma(1+s-d)\Gamma(s)}{\Gamma(1+b+s-d)} \times a^{-s} {}_2F_1(b,s;1+b+s-d;1-\mu/a),$$
$$\mathrm{Re}\, s > 0,\ 1 + \mathrm{Re}\, s > \mathrm{Re}\, d. \quad (15)$$

Thus

$$f_l(k) = f_l^c(k) + \frac{\lambda_l i 2^{l+1} k^{2l+1} (2l+1)!! e^{-i\pi l} e^{\pi\eta/2} e^{2\eta y}}{\Gamma(l+2+i\eta) D_l(k)(\beta - ik)^{2l+2}(\beta^2 + k^2)^{l+1}} \times {}_2F_1\left(l+1+i\eta, 2l+2; l+2+i\eta; \frac{\beta+ik}{\beta-ik}\right). \quad (16)$$

The potential in (1) goes over to the Yamaguchi potential for $l=0$. Thus for the s-wave (16) should yield our result in paper I for the Coulomb plus Yamaguchi potential. We show below this is indeed the case.

For the s-wave (16) reads

$$f_0(k) = f_0^c(k) + \frac{2ik\lambda_0 e^{\pi\eta/2} e^{2\eta y}(\beta - ik)^{-2}}{\Gamma(2+i\eta) D_0(k)(\beta^2+k^2)} \times {}_2F_1\left(1+i\eta, 2; 2+i\eta; \frac{\beta+ik}{\beta-ik}\right). \quad (17)$$

The Fredholm determinant $D_0(k)$ coincides with that for the Coulomb plus Yamaguchi case. While making a comparison it should be noted that instead of the negative sign before λ_l in (1) the conventional Yamaguchi potential is written with a positive sign. We now transform the ${}_2F_1$ function occurring in (17) by the three-term recursion relation[12]

$$p(1-x){}_2F_1(a,p+1;b;x) + \{(a-p)(1-x) + (b-p-a)\} \times {}_2F_1(a,p;b;x) + (p-b){}_2F_1(a,p-1;b;x) = 0. \quad (18)$$

The relation (18) generates the set of hypergeometric functions which differ only in p-values with fixed values of a and b. The parameters a and b may be complex or real integers. In our case both these are complex. In view of (18),

$${}_2F_1\left(1+i\eta, 2; 2+i\eta; \frac{\beta+ik}{\beta-ik}\right) = -\frac{(\beta-ik)(1+i\eta)}{2ik} - i\eta\, {}_2F_1\left(1+i\eta, 1; 2+i\eta; \frac{\beta+ik}{\beta-ik}\right). \quad (19)$$

Using (19) in (17) we get

$$f_0(k) = f_0^c(k)\left[1 - \frac{\lambda_0 e^{2\eta y}}{D_0(k)(\beta^2+k^2)(\beta-ik)} \times \left\{1 - \frac{2\eta k}{(1+i\eta)(\beta-ik)} \times {}_2F_1\left(1+i\eta, 1; 2+i\eta; \frac{\beta+ik}{\beta-ik}\right)\right\}\right], \quad (20)$$

which agrees with our previous result. It is of interest to remark that the result in paper I was obtained without the use of integral (15). The simple technique described there does not appear to work for higher partial waves. Thus, use of the standard integral of Slater[11] was unavoidable for the present case.

In terms of the regular solution $\phi_l(k,r)$ the expression for the off-shell Jost function is given by[2,7]

$$f_l(k,q) = 1 + \frac{q^l e^{-i\pi l}}{(2l+1)!!}\int_0^\infty w_l^{(+)}(qr)\left[\frac{2\eta k}{r}\phi_l(k,r) - \lambda_l 2^{-2l}(l!)^{-2} r^l e^{-\beta r}\int_0^\infty s^l e^{-\beta s}\phi_l(k,s) ds\right] dr \quad (21)$$

with $w_l^{(+)}(qr)$ the Riccati–Hankel function in the phase convention of Messiah.[13] Here q is an off-shell momentum. Using in (21) the expansion of $w_l^{(+)}(qr)$ given in paper II we get $f_l(k,q)$ in the form

$$f_l(k,q) = f_l^c(k,q) - \frac{\lambda_l 2^{-l}(l!)^{-1} q^{-l} i^{-l} e^{2\eta y}}{D_l(k)(\beta^2+k^2)^{l+1}} \times \sum_{j=0}^{l} \frac{(-1)^j (l+j)!}{j!(2iq)^j}\left[(\beta - iq)^{j-l-1} - \frac{(-i)^{j-l}\eta}{(l-j)!} \times \sum_{n=1}^\infty \frac{\rho^{n-1}}{n!}\frac{(l-j+n)!}{(n+2l+1)}\left(\frac{2k}{k+q}\right)^n (k+q)^{j-l-1} \times {}_3F_2\left(1, l+1+n+i\eta, l+1+n-j; 1+n, 2l+2+n; \frac{2k}{k+q}\right)\right], \quad (22)$$

where ${}_3F_2$ is a special case of the generalized hypergeometric function defined by

$${}_mF_n(\alpha_1,\ldots,\alpha_m;\beta_1,\ldots,\beta_n;z) = \sum_{j=0}^\infty \frac{(\alpha_1)_j\cdots(\alpha_m)_j z^j}{(\beta_1)_j\cdots(\beta_n)_j\, j!}. \quad (23)$$

In (22) $f_l^c(k,q)$ is the off-shell Coulomb Jost function for the lth partial wave,[2]

$$f_l^c(k,q) = 1 + \frac{i\eta}{2^l(2l+1)!!}\left(\frac{k}{q}\right)\sum_{j=0}^{l}\frac{(l+j)!}{j!}\left(\frac{2k}{k+q}\right)^{l-j+1}$$
$$\times {}_2F_1\left(l+1+i\eta, l-j+1; 2l+2; \frac{2k}{k+q}\right). \quad (24)$$

The hypergeometric functions which occur in (24) can be generated by using the recurrence relation (18). In an earlier publication van Haeringen[14] performed some kind of regrouping of the ${}_2F_1$ functions in (24) to write a convenient formula for $f_l^c(k,q)$.

For scattering on short-range potentials[15]

$$\lim_{q\to k} f_l(k,q) = f_l(k). \quad (25)$$

The Coulomb analog[16] of (25) is

$$\lim_{q\to k} \omega f_l^c(k,q) = f_l^c(k), \quad k>0, \quad (26)$$

where

$$\omega = \left(\frac{q-k}{q+k}\right)^{i\eta}\frac{e^{\pi\eta/2}}{\Gamma(1+i\eta)}. \quad (27)$$

Using the two-potential formalism in conjunction with the integral representation of the off-shell Coulomb Jost function it has been found that[14] the general relation

$$\lim_{q\to k} \omega f_l(k,q) = f_l(k), \quad k>0 \quad (28)$$

holds good for the Coulomb plus short-range potential. It is not immediately clear if our results for $f_l^c(k,q)$ and $f_l(k,q)$ satisfy the criteria in (26) and (28). Thus in Appendix A we present a proof in respect of this.

IV. CONCLUDING REMARKS

Based on observations in two recent papers, we have constructed exact analytical expressions for the on- and off-shell Jost functions for scattering by the Coulomb plus Graz potential which is believed to provide an accurate description of the proton-proton system by means of separable interactions. We have made some useful checks on our expressions for $f_l(k,q)$ with particular emphasis on their limiting behavior and on-shell discontinuity.

APPENDIX A

From (24) and (27) we have

$$\lim_{q\to k}\omega f_l^c(k,q) = \frac{i\eta e^{\pi\eta/2}}{\Gamma(1+i\eta)2^l(2l+1)!!}$$
$$\times \lim_{q\to k}\left(\frac{q-k}{q+k}\right)^{i\eta}\left(\frac{k}{q}\right)\left[l!\left(\frac{2q}{k+q}\right)^{l+1}\right.$$
$$\times {}_2F_1\left(l+1+i\eta, l+1; 2l+2; \frac{2k}{k+q}\right)$$
$$+ \sum_{j=1}^{l}\frac{(l+j)!}{j!}\left(\frac{2q}{k+q}\right)^{l-j+1}$$
$$\left.\times {}_2F_1\left(l+1+i\eta, l-j+1; 2l+2; \frac{2k}{k+q}\right)\right]. \quad \text{(A1)}$$

The hypergeometric function inside the summation in (A1) is finite as $q\to k$. Thus

$$\lim_{q\to k}\left(\frac{q-k}{q+k}\right)^{i\eta}\left(\frac{k}{q}\right)\sum_{j=1}^{l}\frac{(l+j)!}{j!}\left(\frac{2q}{k+q}\right)^{l-j+1}$$
$$\times {}_2F_1\left(l+1+i\eta, l-j+1; 2l+2; \frac{2k}{k+q}\right) = 0. \quad \text{(A2)}$$

On the other hand, ${}_2F_1\,(l+1+i\eta, l+1; 2l+2; 2k/(k+q))$ $[={}_2F_1\,(\alpha,\beta;\gamma;z)]$ exhibits a singularity in this limit since Re $(\alpha+\beta-\gamma)=0$. We deal with this by using transformation formula

$${}_2F_1(\alpha,\beta;\gamma;z) = (1-z)^{\gamma-\alpha-\beta}{}_2F_1(\gamma-\alpha,\gamma-\beta;\gamma;z) \quad \text{(A3)}$$

and the recursion relation[17]

$$\gamma(\gamma-\beta z-\alpha){}_2F_1(\alpha,\beta;\gamma;z) - \gamma(\gamma-\alpha){}_2F_1(\alpha-1,\beta;\gamma;z)$$
$$+ \alpha\beta z(1-z){}_2F_1(\alpha+1,\beta+1;\gamma+1;z) = 0, \quad \text{(A4)}$$

and finally obtain

$$\lim_{q\to k}\left(\frac{q-k}{q+k}\right)^{i\eta}\left(\frac{k}{q}\right)\left(\frac{2q}{k+q}\right)^{l+1}$$
$$\times {}_2F_1\left(l+1+i\eta, l+1; 2l+2; \frac{2k}{k+q}\right)$$
$$= \frac{\Gamma(1+i\eta)\Gamma(2l+2)}{i\eta\Gamma(l+1+i\eta)}. \quad \text{(A5)}$$

The criterion in (26) now follows from (A1), (A2), and (A5).

While dealing with the Coulomb plus separable potential we specialize to the s-wave for simplicity of presentation. The general l-wave case can be treated similarly. For the s-wave (22) reads

$$f_0(k,q) = f_0^c(k,q)$$
$$- \frac{\lambda_0 e^{2\eta\gamma}(\beta+iq)}{D_0(k)(\beta^2+k^2)(\beta^2+q^2)}$$
$$+ \frac{\lambda_0\eta e^{2\eta\gamma}}{D_0(k)(\beta^2+k^2)(k+q)}$$
$$\times \sum_{n=1}^{\infty}\frac{\rho^{n-1}}{(n+1)}\left(\frac{2k}{k+q}\right)^n$$
$$\times {}_2F_1\left(1, 1+n+i\eta; n+2; \frac{2k}{k+q}\right). \quad \text{(A6)}$$

From (A6) and the s-wave form of (26) we can write

$$\lim_{q\to k}\omega f_0(k,q) = f_0^c(k)$$
$$+ \frac{\lambda_0\eta e^{2\eta\gamma}e^{\pi\eta/2}}{D_0(k)(\beta^2+k^2)\Gamma(1+i\eta)}\lim_{q\to k}\left(\frac{q-k}{q+k}\right)^{i\eta}$$
$$\times (k+q)^{-1}\sum_{n=1}^{\infty}\frac{\rho^{n-1}}{(n+1)}\left(\frac{2k}{k+q}\right)^n$$
$$\times {}_2F_1\left(1, 1+n+i\eta; n+2; \frac{2k}{k+q}\right). \quad \text{(A7)}$$

Transforming the hypergeometric function in (A7) by

$${}_2F_1(\alpha,\beta;\gamma;z) = (1-z)^{-\beta}$$
$$\times {}_2F_1(\beta,\gamma-\alpha;\gamma;z/(z-1)), \quad \text{(A8)}$$

we have

$$\lim_{q\to k}\left(\frac{q-k}{q+k}\right)^{i\eta}(k+q)^{-1}\sum_{n=1}^{\infty}\frac{\rho^{n-1}}{(n+1)}\left(\frac{2k}{k+q}\right)^{n}$$
$$\times{}_2F_1\left(1,1+n+i\eta;n+2;\frac{2k}{k+q}\right)$$
$$=\lim_{q\to k}(q-k)^{-1}\sum_{n=1}^{\infty}\frac{\rho^{n-1}}{(n+1)}\left(\frac{2k}{q-k}\right)^{n}$$
$$\times{}_2F_1\left(1+n+i\eta;n+1;n+2;\frac{2k}{k-q}\right). \quad \text{(A9)}$$

The transformation in (A8) analytically continues ${}_2F_1$ $(1,1+n+i\eta;n+2;2k/(k+q))$ beyond its circle of convergence.[18] Thus, evaluation of the limit in (A9) can be facilitated by the asymptotic behavior[19]

$$\begin{aligned}{}_2F_1(\alpha,\beta;\gamma;z)\to&\frac{\Gamma(\gamma)\Gamma(\beta-\alpha)}{\Gamma(\beta)\Gamma(\gamma-\alpha)}(-z)^{-\alpha}\\&+\frac{\Gamma(\gamma)\Gamma(\alpha-\beta)}{\Gamma(\alpha)\Gamma(\gamma-\beta)}(-z)^{-\beta}\end{aligned} \quad \text{(A10)}$$

of ${}_2F_1(\alpha,\beta,\gamma;z)$ for $z\to\infty$. Equations (17) and (A6)–(A10) can now be combined to show that

$$\lim_{q\to k}\omega f_0(k,q)=f_0(k). \quad \text{(A11)}$$

The treatment given above explicitly demonstrates that the criterion in (28) holds good both for Coulomb and Coulomb-like potentials.[14]

[1] B. Talukdar, D. K. Ghosh, and T. Sasakawa, J. Math. Phys. **23**, 1700 (1982).
[2] B. Talukdar, S. Saha, and T. Sasakawa, J. Math. Phys. **24**, 683 (1983).
[3] M. Gell-Mann and M. L. Goldberger, Phys. Rev. **91**, 398 (1953).
[4] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[5] V. de Alfaro and T. Regge, *Potential Scattering* (North-Holland, Amsterdam, 1965).
[6] L. G. Arnold and R. G. Seyler, Phys. Rev. C **7**, 574 (1973).
[7] M. G. Fuda, Phys. Rev. C **14**, 37 (1976).
[8] L. Crepinsek, C. B. Lang, H. Oberhummer, W. Plessas, and H. Zingl, Acta Phys. Austr. **42**, 139 (1975); L. Crepinsek, H. Oberhummer, W. Plessas, and H. Zingl, Acta Phys. Austr. **39**, 345 (1974).
[9] H. van Haeringen and L. P. Kok, Phys. Lett. A **82**, 317 (1981).
[10] B. Talukdar, U. Das, and S. Chakravarty, Phys. Rev. C **19**, 322 (1979); B. Talukdar and U. Das, Pramana **13**, 525 (1979).
[11] L. J. Slater, *Confluent Hypergeometric Functions* (Cambridge U. P., New York, 1960).
[12] C. Snow, Nat. Bur. Standards (U.S.) Applied Math. Series **19** (U.S. Govt. Printing Office, Washington D. C., 1952).
[13] A. Messiah, *Quantum Mechanics* (Wiley-Interscience, New York, 1961, 1962), Vols. 1 and 2.
[14] H. van Haeringen, J. Math. Phys. **20**, 1109 (1979).
[15] M. G. Fuda and J. S. Whiting, Phys. Rev. C **8**, 1255 (1973).
[16] H. van Haeringen, J. Math. Phys. **19**, 1379 (1978).
[17] W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Chelsea, New York, 1949).
[18] L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge U. P., New York, 1966).
[19] S. Flügge, *Practical Quantum Mechanics* (Springer-Verlag, New York, 1974).

A fully crossing symmetric vertex function

M. Z. I. Gering and W. D. Heiss
Physics Department and Nuclear Physics Research Unit, University of the Witwatersrand, Johannesburg, 2001, South Africa

(Received 2 June 1983; accepted for publication 30 September 1983)

A fully crossing symmetric vertex function was calculated in an exactly soluble model. It has branch points in the interaction strength whose positions are independent of the energy variables. The irreducible vertex parts entering the Bethe–Salpeter equations are more involved than the vertex function itself.

PACS numbers: 03.80. + r, 11.10.St, 11.50.Jg

The four-point vertex function Γ, as defined via the four-point function $\mathscr{K}$ by the equation[1]

$$\mathscr{K} = GG + iGG\Gamma GG, \tag{1}$$

plays a central role in many-body theory. The two-particle propagator $K(s)$, the particle-hole propagator $F(t)$, and the self-energy $\Sigma(\omega)$ can all be calculated from the same vertex function Γ, provided it satisfies crossing symmetry. Attempts to calculate an approximate Γ have yielded a consistent K but an approximate F,[2] or vice versa.[3] Further approaches have been only partially successful.[4] Attempts to construct a Γ that satisfies all three Bethe–Salpeter equations, thus guaranteeing crossing symmetry, result even for a simple model in an intractable power series.[5] To the best of our knowledge, no fully crossing symmetric vertex function has ever been given, and therefore knowledge about its structure is incomplete. In this paper we present novel features of an exact (hence crossing symmetric) vertex function calculated in a soluble model.

The model consists of two twice-degenerate single-particle levels at energies ϵ and δ. The interaction has matrix elements λ between the unperturbed ground state and the two-particle two-hole excited state as in the Lipkin model.[6] The two-point function G can be calculated directly from the spectral representation and the four-point function $\mathscr{K}$ is calculated from the definition

$$(i)^2\mathscr{K}_{1234}(t_1, t_2, t_3, t_4) = \langle 0|T\left[c_1(t_1)c_2(t_2)c_4^\dagger(t_4)c_3^\dagger(t_3)\right]|0\rangle. \tag{2}$$

This enables us to calculate Γ by inverting the Fourier transform of Eq. (1).

Each matrix entry of the vertex function $\Gamma_{1234}(s, t, u)$ thus calculated has a simple pole in each of the frequency variables ω_1, ω_2, ω_3, and ω_4 which correspond to the matrix labels (leg poles). The positions of those poles coincide with the corresponding entries of the self-energy matrix $\Sigma_{13}(\omega)$. A further left- and right-hand pole occurs in the variable $s = \omega_1 + \omega_2$, and corresponds to the $A-2$ and $A+2$ particle systems, respectively. The two excited states of the A particle system arise as left- and right-hand poles in the variables $t = \omega_3 - \omega_1$ and $u = \omega_3 - \omega_2$. From this vertex function, both $F(t)$ and $K(s)$ as well as $\Sigma(\omega_3)$ are obtained by performing the appropriate integrals. The exact form of the vertex function is given elsewhere.[7]

Since the essential features of our findings are found in the simpler structure of $F(t)$, we present only the matrix $F_{24,31}(t)$ where each pair of indices runs over the labels pp, hh, ph, hp:

$$F_{24,31}(t) = \begin{pmatrix} \frac{\alpha^2\beta^2}{t-2\overline{\omega}} - \frac{\alpha^2\beta^2}{t+2\overline{\omega}} & -\frac{\alpha^2\beta^2}{t-2\overline{\omega}} + \frac{\alpha^2\beta^2}{t+2\overline{\omega}} & 0 & 0 \\ -\frac{\alpha^2\beta^2}{t-2\overline{\omega}} + \frac{\alpha^2\beta^2}{t+2\overline{\omega}} & \frac{\alpha^2\beta^2}{t-2\overline{\omega}} - \frac{\alpha^2\beta^2}{t+2\overline{\omega}} & 0 & 0 \\ 0 & 0 & \frac{\alpha^2}{t-\overline{\omega}} - \frac{\beta^2}{t+\overline{\omega}} & \frac{\alpha\beta}{t-\overline{\omega}} - \frac{\alpha\beta}{t+\overline{\omega}} \\ 0 & 0 & \frac{\alpha\beta}{t-\overline{\omega}} - \frac{\alpha\beta}{t+\overline{\omega}} & \frac{\beta^2}{t-\overline{\omega}} - \frac{\alpha^2}{t+\overline{\omega}} \end{pmatrix}, \tag{3}$$

$$\overline{\omega} = ((\epsilon-\delta)^2 + \lambda^2)^{1/2}, \quad \alpha^2 = (1-\beta^2) = (\overline{\omega} + (\epsilon-\delta))/2\overline{\omega}.$$

It is obvious that unlike an RPA solution the pole positions are real for any real interaction strength. Note that the RPA sum rule no longer holds and that F has entries outside the traditional RPA corner. While this was expected, there are features which appear to be new.

In contrast to a previous conjecture,[5] Γ has singularities in the interaction strength λ which are independent of the frequency variables s, t, and u. The eigenmode frequency $\overline{\omega}$

 0022-2488/84/020327-03$02.50

as a function of λ has square root branch points at $\lambda = \pm i(\epsilon - \delta)$. These branch points appear in the vertex function and carry over to $F(t)$ [Eq. (3)] and $K(s)$.

Branch-point singularities cannot emerge as a solution of a linear integral equation of the Fredholm type, as these solutions are meremorphic functions of the strength parameter. Consequently, they cannot arise from solutions of the traditional approximations such as the RPA, the Galitskii ladder, or more sophisticated but still linear approaches.[4] While it is true that the RPA eigenmode frequency $\omega_{\rm RPA} = ((\epsilon - \delta)^2 - \lambda^2)^{1/2}$ has a branch point at $\lambda = \pm(\epsilon - \delta)$, the singularities of $F_{\rm RPA}$, as a function of λ, are all frequency dependent and $F_{\rm RPA}$ is single valued.

The appearance of the branch points underlines the nonlinear character of the equations for Γ when starting from the bare interaction.[5] When starting from the Bethe–Salpeter equations it follows that such branch points must appear in the respective inputs of these equations:

$$\begin{aligned} \Gamma &= \mathscr{V} + (\mathscr{V} GG\Gamma)_s, \\ \Gamma &= \mathscr{U} + (\mathscr{U} GG\Gamma)_t, \\ \Gamma &= \mathscr{W} + (\mathscr{W} GG\Gamma)_u. \end{aligned} \tag{4}$$

For the sake of completeness one may ask for the respective inputs $\mathscr{V}$, $\mathscr{U}$, or $\mathscr{W}$ in our model, as these are usually believed to be simpler than Γ itself and therefore used as a starting point for realistic calculations. Knowing Γ we invert Eqs. (4) to obtain the respective irreducible vertex parts. Despite the simpler form for Γ as described above, the irreducible vertex parts each have an infinite number of poles corresponding to diagrams with an arbitrary number of simultaneous particle-hole propagations. Such diagrams clearly violate the Pauli principle, yet they must occur in $\mathscr{V}$, $\mathscr{U}$, or $\mathscr{W}$. Only when the irreducible parts are combined to form the full vertex function do the offensive terms cancel.

In Fig. 1 three sixth-order diagrams, relating to our model, are drawn, each having a u-pole, a pole in ω_3, and a pole in ω_1. The first diagram is s-reducible, the second t-reducible, and the third is irreducible with respect to all three channels; they must be included in $\mathscr{V}$, $\mathscr{U}$, or $\mathscr{W}$ wherever appropriate. However, an expression of the above form does not appear in the full vertex function, since the different Feynman diagrams cancel when the subclasses are combined to form Γ. This explains why $\mathscr{V}$, $\mathscr{U}$, and $\mathscr{W}$ are more complicated than Γ, and indeed, in this simple model the "smaller" set of all irreducible diagrams (denoted by γ_0) is more complicated than Γ itself.

A fully crossing symmetric vertex function can be constructed using γ_0 as input.[5] In particular, using the simplest expression for γ_0, i.e., the bare interaction, one obtains in principle a Γ which obeys all the conditions of crossing symmetry. However, this procedure violates the Pauli principle, and an exact γ_0 would contain, as part of the input, diagrams to exactly cancel terms yet to be generated.

From the practical point of view this would scarcely be possible. However, the unphysical poles thus generated are expected to make an insignificant contribution as they occur

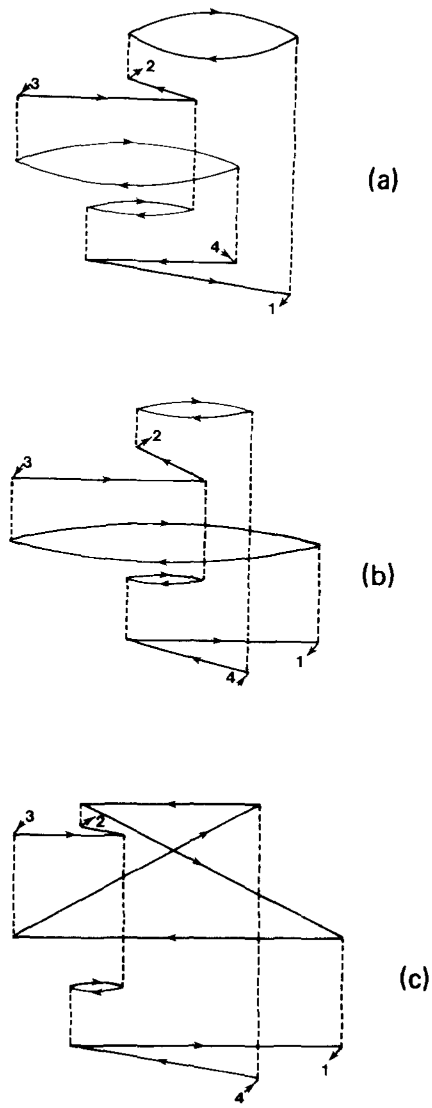

FIG. 1. Three sixth-order diagrams contributing to Γ_{phph}. Each diagram has the same pole structure. (a) is s-reducible, (b) is t-reducible, and (c) is irreducible.

at high energies and have a small residue.

Although our findings appear to cast doubt as to the usefulness of the Bethe–Salpeter equations, we feel that the physical concepts of highly collective states are valid, in particular when large particle numbers are considered. Guided by these ideas one expects higher-order contributions to γ_0 to be unimportant and Eqs. (4) to be relevant from a practical viewpoint.

The branch point in the interaction strength is probably of much greater physical relevance. We recall that the breakdown of the RPA is often associated with a phase transition. This should clearly be revealed in an exactly soluble model,[8] and could in fact be signaled by a branch point in the interaction strength. As discussed above, we see the occurrence of this particular singularity as a consequence of crossing symmetry. Whether connections to this effect can be established is subject to further investigations.

[1]C. A. Engelbrecht, F. J. W. Hahne, and W. D. Heiss, Ann. Phys. (N.Y.) **104**, 221 (1977).
[2]F. J. W. Hahne, W. D. Heiss, and C. A. Engelbrecht, Ann. Phys. (N.Y.) **104**, 251 (1977).
[3]W. D. Heiss, C. A. Engelbrecht, and F. J. W. Hahne, Ann. Phys. (N.Y.) **104**, 274 (1977).
[4]W. D. Heiss, J. Math. Phys. **21**, 848 (1980).
[5]W. D. Heiss, J. Math. Phys. **22**, 1182 (1981).
[6]H. J. Lipkin, N. Meshkov, and A. J. Glick, Nucl. Phys. **62**, 188 (1965).
[7]M. Z. I. Gering and W. D. Heiss, S. Afr. J. Phys. (to be published).
[8]D. Agassi, H. J. Lipkin, and N. Meshkov, Nucl. Phys. **86**, 321 (1966).

A note on fluctuations of weak gravitational radiation and the classical Hanbury Brown–Twiss effect

Ronald L. Mallett
Center for Theoretical Physics, University of Texas, Austin, Texas 78712 and Department of Physics, University of Connecticut, Storrs, Connecticut 06268[a]

(Received 18 January 1983; accepted for publication 9 September 1983)

We comment on the possible application of the Hanbury Brown–Twiss method to the study of the classical statistical properties of weak gravitational radiation from chaotic microscopic sources.

PACS numbers: 04.20. − q, 04.30. + x

It is well known that one of the most important predictions of general relativity is the existence of gravitational waves. Much experimental and theoretical work is being done to detect and interpret this gravitational radiation.[1] In this paper we comment on the possible application of the Hanbury Brown–Twiss method[2] to the analysis of the classical statistical properties of the radiative part of the weak gravitational field. This treatment is based on the polarized classical degree of second-order coherence.[3] For the gravitational field at space-time points (1) and (2) in terms of observable quantities this is given by

$$\langle R^*_{xoxo}(1)R^*_{xoxo}(2)R_{xoxo}(2)R_{xoxo}(1)\rangle = |\langle R^*_{xoxo}(1)R_{xoxo}(2)\rangle|^2 + \langle |R_{xoxo}(1)|^2\rangle\langle |R_{xoxo}(2)|^2\rangle, \tag{1}$$

where $\langle\ \rangle$ denotes the ensemble average. R_{xoxo} is the radiative part of the linearized Riemann curvature tensor[4] $R_{xoxo} = (\kappa/2c^2)\partial_t^2\gamma_+^{xx}$ with $\kappa = (16\pi G/c^4)^{1/2}$, G = Newtonian gravitational constant, and γ_+^{xx} is the $+$ polarization state of the positive frequency part of the analytic[5] gravitational signal. The stochastic properties of the gravitational field are described by a Fourier expansion of γ_+^{xx}. The complex Fourier amplitudes are then regarded as random variables.

As an example, consider weak gravitational radiation emitted by chaotic microscopic sources[6] in thermal motion.[7] For sources in motion along the z-axis with a Maxwellian velocity distribution, then from the statistical independence of the amplitudes and the ergodic theorem[3] we obtain

$$\langle R^*_{xoxo}(1)R_{xoxo}(2)\rangle = \frac{\mathscr{I}\kappa^2}{2(2\pi)^{1/2}\Omega c^3} \times\int_{-\infty}^{\infty}\omega_k^2\exp[-(\omega_k-\omega_0)^2/2\Omega^2 + i\omega_k\tau]\,d\omega_k, \tag{2}$$

where $\tau = t_1 - t_2 + (1/c)(z_2 - z_1)$ and the product of Fourier expansions has been replaced by an integration. $\mathscr{I}$ is the time-averaged intensity and ω_0 the rest-frame frequency of the gravitational radiation. Performing the integration in Eq. (2) we have

$$\langle R^*_{xoxo}(1)R_{xoxo}(2)\rangle = (\mathscr{I}\kappa^2/2c^3)[(\tau^2\Omega^4 - \Omega^2 - \omega_0^2) - i2\gamma\omega_0\Omega^2] \times\exp[\tfrac{1}{2}\tau^2\Omega^2 + i\omega_0\tau]. \tag{3}$$

Substituting Eq. (3) in (1) yields

$$\langle R^*_{xoxo}(1)R^*_{xoxo}(2)R_{xoxo}(2)R_{xoxo}(1)\rangle = \frac{\mathscr{I}^2\kappa^4}{4c^6}\{[(\tau^2\Omega^4 - \Omega^2 - \omega_0^2)^2 + 4\tau^2\omega_0^2\Omega^4] \times\exp(-\tau^2\Omega^2) + (\Omega^2 + \omega_0^2)^2\}. \tag{4}$$

From Eq. (4) it is clear that for $\tau\Omega < \infty$,

$$\langle R^*_{xoxo}(1)R^*_{xoxo}(2)R_{xoxo}(2)R(1)\rangle > \langle |R_{xoxo}(1)|^2\rangle\langle |R_{xoxo}(2)|^2\rangle = (\kappa^4\mathscr{I}^2/4c^6)(\Omega^2 + \omega_0^2)^2. \tag{5}$$

Consequently, gravitational radiation from sources in thermal motion show a positive HBT correlation. Hence, two detectors at two different space-time points with a response time less than the coherence time $\tau_c = 1/\Omega$ of the radiation would exhibit a correlation in gravitational field fluctuations of the form in Eq. (4).

A Gaussian spectral distribution due to the Doppler broadening of electromagnetic radiation due to sources in thermal motion leads to the result[3]

$$\langle E_x^*(1)E_x^*(2)E_x(2)E_x(1)\rangle = (4\bar{I}^2/\epsilon_0^2c^2)[\exp(-\delta^2\tau^2) + 1], \tag{6}$$

where ϵ_0 is the permittivity, $\bar{I}$ is the time average light intensity, and the coherence time is $\tau_c = 1/\delta$. Comparing Eqs. (4) and (6) shows that, unlike the electric field, the overall Gaussian profile of the gravitational HBT correlation function is modulated by a time-dependent factor. It is important to note the origin of the difference between Eqs. (4) and (6). Basically, the electromagnetic coherence function, as measured by the ensemble average of the electric field strength E_i is proportional to terms involving the first time derivatives of the vector potential A_i. On the other hand, the gravitational coherence function, as measured by the ensemble average of the tidal field strength R_{iojo}, is proportional to terms involving the second time derivatives of the gravitational potential γ_{ij}. The importance of the tidal force field R_{iojo} as the observable quantity associated with the gravitational field can be traced directly back to the principle of equivalence and the meaning of the gravitational field as measured from a free-falling frame of reference. The components R_{iojo} of the Riemann curvature tensor provide a real measure of the presence of a gravitational field through the mechanism of geodesic deviation[8] when the velocity of the detector is negligible compared to the speed of light.

[a] Permanent address.

 0022-2488/84/020330-02$02.50

A further example of gravitational radiation comes from consideration of collisions. If a microscopic source of gravitational radiation is considered to be emitting radiation continuously before and after a collision with the same frequency, then the net result of the collision is an abrupt and random change of phase.[9] The net result of multiple successive collisions is a Lorentz spectral distribution which leads to the result, after a contour integration,

$$\begin{aligned} |\langle R^*_{xoxo}(1)R_{xoxo}(2)\rangle| &= \frac{\kappa^2\bar{\mathscr{I}}\Gamma}{2\pi c^3}\left|\int_{-\infty}^{\infty}\frac{\omega_k^2 e^{i\omega_k\tau}}{(\omega_0-\omega_k)^2+\Gamma^2}d\omega_k\right| \\ &= \frac{\kappa^2\bar{\mathscr{I}}}{2c^3}\left[(\omega_0^2-\Gamma^2)^2+4\omega_0^2\Gamma^2\right]^{1/2}e^{-\Gamma|\tau|}, \end{aligned} \tag{7}$$

where $\tau = t_1 - t_2 + (1/c)(z_2 - z_1)$. Hence, substituting Eq. (7) in Eq. (1) we have

$$\langle R^*_{xoxo}(1)R^*_{xoxo}(2)R_{xoxo}(2)R_{xoxo}(1)\rangle = (\kappa^2\bar{\mathscr{I}}^2/4c^6)\left[(\omega_0^2-\Gamma^2)^2+4\omega_0^2\Gamma^2\right](e^{-2\Gamma|\tau|}+1). \tag{8}$$

From Eq. (8), for $\Gamma|\tau| < \infty$ we have

$$\begin{aligned} \langle R^*_{xoxo}(1)R^*_{xoxo}(2)R_{xoxo}(2)R_{xoxo}(1)\rangle &> \langle |R^{(1)}_{xoxo}|^2\rangle\langle |R^{(2)}_{xoxo}|^2\rangle \\ &= \frac{\kappa^4\bar{\mathscr{I}}^2}{4c^6}\left[(\omega_0^2-\Gamma^2)+4\omega_0^2\Gamma^2\right]. \end{aligned} \tag{9}$$

Hence, gravitational sources in collision exhibit a positive HBT correlation of the form given by Eq. (8). For electromagnetic sources with a Lorentz spectral distribution the electric field correlation function has the form[3]

$$\langle E^*_x(1)E^*_x(2)E_x(2)E_x(1)\rangle = \frac{4\bar{I}^2}{\epsilon_0^2c^2}(e^{-2\gamma|\tau|}+1). \tag{10}$$

Hence, unlike the electromagnetic field, the overall gravitational Lorentz correlation profile is modified by the collision frequency.

Comparing Eqs. (4) and (8) we see that Eq. (1) can serve to distinguish the coherence characteristics of the gravitational radiation associated with different spectral distributions. The application of Eq. (1) is, however, by no means limited to these cases. In principle, this analysis can be applied to the arbitrary gravitational spectral distribution associated with any collection of sources.

It should be mentioned that in the quantum-mechanical interpretation of the degree of second-order coherence for the electromagnetic field a positive Hanbury Brown–Twiss correlation manifests itself as a tendency for photons to arrive in pairs. This phenomenon is called photon bunching.[2] The quantum-mechanical interpretation of a positive gravitational HBT correlation should, by analogy, indicate the existence of graviton bunching. This bunching is, however, a characteristic of the chaotic nature of the sources. A coherent source would result in a zero HBT correlation or random graviton arrivals.

ACKNOWLEDGMENT

The author would like to thank Dr. John A. Wheeler, Director of the Center for Theoretical Physics at the University of Texas, for his hospitality.

[1]See, e.g., K. S. Thorne, Rev. Mod. Phys. **52**, 285 (1980); M. Rees, R. Ruffini, and J. A. Wheeler, *Black Holes Gravitational Waves and Cosmology* (Gordon and Breach, New York, 1974), pp. 84–142; W. H. Press and K. S. Thorne, Annu. Rev. Astron. Astrophys. **10**, 335 (1970); L. Halpern and B. Laurent, Nuovo Cimento **33**, 728 (1964).

[2]R. Hanbury Brown and R. Q. Twiss, Nature (London) **178**, 1449 (1956). For a comprehensive collection of early papers on both the theoretical and experimental work done on the classical and quantum statistical properties of light see, e.g., *Selected Papers on Coherence and Fluctuations of Light*, Vols. I and II, edited by L. Mandel and E. Wolf (Dover, New York, 1970).

[3]R. J. Glauber, in *Quantum Optics and Electronics*, edited by C. DeWitt, A. Blandin, and C. Cohen-Tannoudji (Gordon and Breach, New York, 1965), p. 63; R. London, *The Quantum Theory of Light* (Clarendon, Oxford, 1981).

[4]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 952.

[5]The representation of a real signal by an analytic one is well known in classical statistical optics. See, e.g., J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968), p. 4; H. M. Nussenzveig, *Introduction to Quantum Optics* (Gordon and Breach, New York, 1973), p. 4.

[6]L. Halpern and B. Laurent, Ref. 1.

[7]G. B. Rybicki and A. P. Lightman, *Radiative Processes In Astrophysics* (Wiley, New York, 1982), pp. 287–289.

[8]See, e.g., M. Rees, R. Ruffini, and J. A. Wheeler, Ref. 1, p. 64; W. H. Press and K. S. Thorne, Ref. 1.

[9]G. B. Rybicki and A. P. Lightman, Ref. 7, p. 290.

A class of shear-free perfect fluids in general relativity. I

A. J. White[a] and C. B. Collins
Department of Applied Mathematics, University of Waterloo, Ontario, N2L 3G1, Canada

(Received 18 August 1983; accepted for publication 23 September 1983)

It has been conjectured that shear-free perfect fluids in general relativity, with an equation of state $p = p(\mu)$ and satisfying $\mu + p \not\equiv 0$, necessarily have either zero expansion or zero vorticity. We prove that this result holds in the restricted case when the fluid's vorticity and acceleration are parallel. Specifically, we prove that if the vorticity is nonzero, the fluid's volume expansion must vanish.

PACS numbers: 04.20.Cv, 02.40. + m

1. INTRODUCTION

We shall assume in the following a familiarity with the kinematic quantities of a fluid, as given by Ellis[1]:

$$u_{i;j} = \sigma_{ij} + \tfrac{1}{3}\theta h_{ij} + \omega_{ij} - \dot{u}_i u_j,$$

where u^i is the future-pointing (timelike) unit tangent vector to the flow, $\dot{u}^i$ is the acceleration, and σ_{ij}, ω_{ij}, and θ are respectively the (rate of) shear tensor, the (rate of) vorticity tensor, and the (rate of) volume expansion scalar. Shear-free fluids are characterized by the vanishing of σ_{ij}. The tensor h_{ij} is the projection tensor, $h_{ij} = g_{ij} + u_i u_j$, into the rest space of an observer whose 4-velocity is u^i. As usual, indices are raised and lowered with the metric tensor g_{ij}; our signature, units and conventions for the Riemann and Ricci tensors coincide with those of Ellis.[1]

We shall suppose that the matter content of the space-time is a shear-free perfect fluid obeying Einstein's equations

$$R_{ij} - \tfrac{1}{2}Rg_{ij} + \Lambda g_{ij} = \mu u_i u_j + p h_{ij},$$

where μ is the energy density and p is the pressure of the fluid. We shall further suppose that μ and p are related by a barotropic equation of state $p = p(\mu)$, and that μ and p satisfy the physically reasonable requirement $\mu + p \not\equiv 0$. In fact, we could well impose the more stringent requirement $\mu + p > 0$, but for mathematical completeness we shall adopt the less restrictive condition $\mu + p \not\equiv 0$.

It is important to note that, strictly speaking, the fluid is merely assumed to have vanishing shear in some open subset of the space-time manifold, and that, as such, our results will be local and not global.

A number of existing results show that a shear-free perfect fluid, with equation of state $p = p(\mu)$ such that $\mu + p \not\equiv 0$, will have either vanishing vorticity or vanishing expansion, *provided various additional assumptions are satisfied*. The first result of this nature of which we are aware is due to Gödel,[2] who proved its validity for dust ($p \equiv 0$) space-times that are spatially homogeneous of Bianchi type IX. This result was generalized to all dust space-times (irrespective of additional symmetries) by Ellis.[3] It has also been generalized by various authors in another direction, viz., to all spatially homogeneous space-times. This was done first for specific equations of state [Schücking[4] considered dust models and Banerji[5] discussed the equations of state of the form $p = (\gamma - 1)\mu$, where γ is a constant satisfying $1 < \gamma \neq \frac{10}{9}$], and culminated in the result of King and Ellis,[6] who showed that the result was valid even when the only restriction on the quantities μ and p is $\mu + p > 0$. We have since shown[7] that the result is valid for spatially homogeneous space-times under the more general condition $\mu + p \not\equiv 0$. Finally, Treciokas and Ellis[8] proved that for shear-free radiation (in which the equation of state is $p = \frac{1}{3}\mu$) either the vorticity or the expansion must vanish. This compendium of results leads one to suspect that the following conjecture holds:

Conjecture: Any shear-free perfect fluid in general relativity with an equation of state $p = p(\mu)$, such that $\mu + p \not\equiv 0$, has either vanishing vorticity or vanishing expansion, i.e., $\sigma \equiv 0 \Rightarrow \omega\theta \equiv 0$.

Here ω and σ are the vorticity and shear scalars, respectively, defined by $\omega := (\frac{1}{2}\omega^{ij}\omega_{ij})^{1/2}$ and $\sigma := (\frac{1}{2}\sigma^{ij}\sigma_{ij})^{1/2}$. We know of exact solutions which satisfy the conditions of this conjecture (but we know of none which indicates that the conjecture is false). Among such solutions are the Friedmann–Robertson–Walker (FRW) models, and the solutions due to Wyman,[9] Gödel,[10] Krasinski,[11] and Collins and Wainwright.[12] We show in the present article that the above conjecture is true in yet another special case, viz., when the fluid's vorticity and acceleration are parallel. More precisely, we shall describe the dynamics using an orthonormal tetrad in which the timelike axis is aligned along the fluid flow, and with respect to which the (spatial) components of the vorticity and acceleration may be written as $(\omega, 0, 0)$ and $(\dot{u}, 0, 0)$. We shall prove the result by first assuming that $\omega\theta \not\equiv 0$, and then deriving a contradiction. A familiarity with the orthonormal tetrad technique (see, e.g., Refs. 3 and 13) will be assumed.

The plan of this article is as follows. In Sec. 2, we describe the orthonormal tetrad specialization, and introduce a notation which makes transparent the role of the commutation functions. In Sec. 3, we prove our result. The special case in which the flow is geodesic ($\dot{u} = 0$) was effectively treated by Ellis[3] and requires a separate proof. We provide a proof which follows closely that of Ellis,[3] but which is in our notation, thus providing a clearer understanding of the role of the intrinsic geometrical quantities, and allowing direct comparison with the proof in the case $\dot{u} \not\equiv 0$. Since we plan to investigate various features of the class of solutions presently

[a] Present address: Materials and Mechanics Branch, Atomic Energy of Canada Limited, Whiteshell Nuclear Research Establishment, Pinawa, Manitoba, R0E 1L0, Canada.

 0022-2488/84/020332-06$02.50

being considered, a treatment of the special case $\dot{u} = 0$ is also desirable. A discussion of the features which we intend to investigate is provided in Sec. 4.

2. TETRAD SPECIALIZATION

We adhere closely to the notations and conventions of MacCallum,[13] but we shall find it convenient to transcribe the commutation functions $\gamma^{\alpha}{}_{\beta\delta}$ $(\alpha,\beta,\delta = 1, 2, 3)$ as follows:

$$\gamma^{\alpha}{}_{\beta\delta} = \epsilon_{\beta\delta\epsilon} n^{\epsilon\alpha} + \delta^{\alpha}{}_{\delta} a_{\beta} - \delta^{\alpha}{}_{\beta} a_{\delta}, \quad n^{\alpha\beta} = n^{(\alpha\beta)},$$

where

$$n_{\alpha\beta} = \begin{bmatrix} n & \frac{1}{2}(d_3 + A_3) & -\frac{1}{2}(d_2 + A_2) \\ \frac{1}{2}(d_3 + A_3) & \hat{\theta}_{23} + \hat{\Omega} & -\frac{1}{2}(\hat{\theta}_{22} - \hat{\theta}_{33}) \\ -\frac{1}{2}(d_2 + A_2) & -\frac{1}{2}(\hat{\theta}_{22} - \hat{\theta}_{33}) & -(\hat{\theta}_{23} - \hat{\Omega}) \end{bmatrix}$$

and $a_{\alpha} = [-\frac{1}{2}(\hat{\theta}_{22} + \hat{\theta}_{33}), \frac{1}{2}(d_2 - A_2), \frac{1}{2}(d_3 - A_3)]$. If the timelike vector $\mathbf{e}_0$ of the tetrad is aligned along the fluid flow $\mathbf{u}$, then the commutation functions $\gamma^{0}{}_{0\alpha}$, $\gamma^{\alpha}{}_{0\beta}$, and $\gamma^{0}{}_{\alpha\beta}$ $(\alpha, \beta = 1, 2, 3)$ are directly related to the kinematic quantities of the fluid, and to the angular velocity Ω^{α} of the triad $\{\mathbf{e}_{\alpha}\}$ with respect to a set of Fermi-propagated axes: $\gamma^{0}{}_{0\alpha} = \dot{u}_{\alpha}$, $\gamma^{0}{}_{\alpha\beta} = -2\epsilon_{\alpha\beta\delta}\omega^{\delta}$, and $\gamma^{\alpha}{}_{0\beta} = -\sigma^{\alpha}{}_{\beta} - \frac{1}{3}\theta\delta^{\alpha}{}_{\beta} - \epsilon^{\alpha}{}_{\beta\gamma}(\omega^{\gamma} + \Omega^{\gamma})$, where $\dot{u}_{\alpha}$, ω_{α}, $\sigma_{\alpha\beta}$, and θ are the components of the fluid's acceleration vector, vorticity vector, shear tensor, and expansion scalar, respectively. Moreover, with the above transcription on $\gamma^{\alpha}{}_{\beta\delta}$, the quantities d_A, n, and $\hat{\theta}_{AB}$ $(A, B = 2, 3)$ may be regarded as the *spatial* components of the acceleration, vorticity, and expansion tensor of the $\mathbf{e}_1$ lines, while $\hat{\Omega}$ measures the *spatial* component of the angular velocity of the dyad $\{\mathbf{e}_2, \mathbf{e}_3\}$ along the $\mathbf{e}_1$ lines [relative to the "nonrotating" frame, in which $D\mathbf{e}_2\cdot\mathbf{e}_3 = 0$ $(\Leftrightarrow D\mathbf{e}_3\cdot\mathbf{e}_2 = 0)$, where D denotes directional differentiation along $\mathbf{e}_1$]. We can further decompose $\hat{\theta}_{AB}$ into its *spatial* trace and trace-free parts. This gives rise to an "expansion" scalar $\hat{\theta} = \hat{\theta}_{22} + \hat{\theta}_{33}$ and to a "shear" tensor, whose spatial components $\hat{\sigma}_{AB}$ are given by $\hat{\sigma}_{22} = \frac{1}{2}(\hat{\theta}_{22} - \hat{\theta}_{33}) = -\hat{\sigma}_{33}$ and $\hat{\sigma}_{23} = \hat{\theta}_{23}$; then $n_{22} = \hat{\sigma}_{23} + \hat{\Omega}$, $n_{33} = -(\hat{\sigma}_{23} - \hat{\Omega})$, $n_{23} = -\hat{\sigma}_{22} = \hat{\sigma}_{33}$, and $a_1 = -\frac{1}{2}\hat{\theta}$. Note that this decomposition into "kinematic" quantities associated with the $\mathbf{e}_1$-congruence is similar to, but not identical with, that of previous authors.[3,14–16]

Since we shall be choosing both $\mathbf{e}_0$ and $\mathbf{e}_1$ in a geometrically invariantly defined manner, the mathematical expressions that we shall encounter will consist almost entirely of "kinematic" quantities associated with the $\mathbf{e}_0$ and $\mathbf{e}_1$ directions, whose role will thereby be accentuated. The quantities A_2 and A_3 have not yet been interpreted; they can be characterized in terms of components of the acceleration and expansion of the $\mathbf{e}_2$ lines.

We now proceed to specify an orthonormal tetrad to be used in obtaining the results of Sec. 3. This tetrad is for a shear-free perfect fluid with an equation of state $p = p(\mu)$, such that $\mu + p \not\equiv 0$, and with the timelike axis $(\mathbf{e}_0)$ aligned along the fluid flow $\mathbf{u}$. For this development, the following proposition is required:

Proposition 2.1: For any shear-free perfect fluid with an equation of state $p = p(\mu)$, such that $\mu + p \not\equiv 0$, the vorticity vector $\boldsymbol{\omega}$ satisfies the propagation equation

$$\partial_0\omega^{\alpha} = (p' - \tfrac{2}{3})\omega^{\alpha}\theta + \epsilon^{\alpha\beta\gamma}\omega_{\beta}\Omega_{\gamma}, \tag{2.1}$$

where $\mathbf{e}_0$ defines the direction of the fluid flow and $p' := dp/d\mu$.

Proof: By the contracted Bianchi identities,

$$T^{ij}{}_{;j} = 0 \Leftrightarrow \partial_0\mu + (\mu + p)\theta = 0 \text{ and } \partial_{\alpha}p + (\mu + p)\dot{u}_{\alpha} = 0. \tag{2.2}$$

We define a function F by

$$F(\mu) := -\int \frac{p'}{\mu + p}\, d\mu, \quad \text{where } p' = \frac{dp}{d\mu} \tag{2.3}$$

(apart from a multiplicative constant, e^F is then the thermodynamic "enthalpy" of the fluid; the quantity $-F$ coincides with the "index function" of Ref. 14). We observe that, by (2.2),

$$\partial_0 F = p'\theta \quad \text{and} \quad \partial_{\alpha}F = \dot{u}_{\alpha} \tag{2.4}$$

and that F is uniquely defined, up to an additive constant. Applying the $[\mathbf{e}_{\alpha}, \mathbf{e}_{\beta}]$ commutator to F, we obtain

$$\epsilon^{\gamma\alpha\beta}\partial_{\alpha}\dot{u}_{\beta} = -2\omega^{\gamma}p'\theta + n^{\gamma\alpha}\dot{u}_{\alpha} + \epsilon^{\gamma\alpha\beta}a_{\alpha}\dot{u}_{\beta}, \tag{2.5}$$

and hence by the Jacobi identities [Eq. (79) of Ref. 13], Eq. (2.1) results. □

Remarks: Note that the proof of Proposition 2.1 holds even in the case when p is identically constant [and then $\dot{u}_{\alpha} = 0$ by (2.2)] and in particular in the case of dust $(p \equiv 0)$. Note also that Eq. (2.1) is simply the tetrad form of the usual vorticity conservation equation for a perfect fluid with an equation of state, restricted to the shear-free case (see Ref. 1 and references cited therein).

Since we shall be considering models in which the vorticity and acceleration are parallel, we choose $\mathbf{e}_1$ to be aligned along their common direction. Thus $\boldsymbol{\omega} = (\omega, 0, 0)$ and $\dot{\mathbf{u}} = (\dot{u}, 0, 0)$. For our present purposes, we shall assume $\omega \not\equiv 0$, but for later investigations we shall also wish to consider the cases $\omega \equiv 0$, $\dot{u} \not\equiv 0$ and $\omega \equiv 0$, $\dot{u} \equiv 0$. The latter case is exceptional, since the conditions $\sigma \equiv 0$, $\omega \equiv 0$, $\dot{u} \equiv 0$ characterize the spatially homogeneous and isotropic FRW models[1], and we shall exclude this from our discussion. Thus we shall in any event be considering solutions in which $\dot{u}^2 + \omega^2 \not\equiv 0$. The tetrad is then fixed to within a reflection $\mathbf{e}_1 \to \pm\mathbf{e}_1$ and an arbitrary position-dependent rotation, possibly combined with a reflection:

$$\begin{aligned} &\mathbf{e}_2 \to \mathbf{e}_2 \cos\Theta + \mathbf{e}_3 \sin\Theta, \\ &\mathbf{e}_3 \to \delta(-\mathbf{e}_2 \sin\Theta + \mathbf{e}_3 \cos\Theta), \quad \delta = \pm 1. \end{aligned} \tag{2.6}$$

We now show that we may choose a tetrad in which $\boldsymbol{\omega} + \boldsymbol{\Omega} = \mathbf{0}$. By applying the $[\mathbf{e}_0, \mathbf{e}_{\alpha}]$ commutator to F and using (2.3) and (2.4), we obtain

$$\begin{aligned} \partial_0\dot{u}_{\alpha} &= p'\,\partial_{\alpha}\theta + \theta\,\partial_{\alpha}\,p' + (p' - \tfrac{1}{3})\dot{u}_{\alpha}\theta \\ &\quad + \epsilon_{\alpha\beta\gamma}\dot{u}^{\beta}(\omega^{\gamma} + \Omega^{\gamma}). \end{aligned} \tag{2.7}$$

We first show that $\Omega_2 = \Omega_3 = 0$. If $\boldsymbol{\omega} \not\equiv \mathbf{0}$, preservation of the conditions $\omega_2 = \omega_3 = 0$ along the $\mathbf{e}_0$ lines implies $\Omega_2 = \Omega_3 = 0$, by (2.1). If $\boldsymbol{\omega} \equiv \mathbf{0}$ and $\dot{\mathbf{u}} \not\equiv \mathbf{0}$, then preservation of the conditions $\dot{u}_2 = \dot{u}_3 = 0$ along the $\mathbf{e}_0$ lines implies $p'\partial_2\theta - \dot{u}\Omega_3 = p'\,\partial_3\theta + \dot{u}\Omega_2 = 0$, by (2.7), where use is made of (2.2) and of the fact that $p' \not\equiv 0$, to deduce that $\partial_2 p' = \partial_3 p' = 0$. Now the $(0\,\alpha)$ field equations, with $\sigma = \omega = 0$, imply $\partial_{\alpha}\theta = 0$; hence $\Omega_2 = \Omega_3 = 0$. Under a transformation (2.6), with $\delta = 1$, $\gamma^3{}_{02} = \omega_1 + \Omega_1 \to \omega_1 + \Omega_1 + \partial_0\Theta$, so by applying a transformation in which Θ is propagated according to the requirement $\partial_0\Theta = -(\omega_1 + \Omega_1)$ we may arrange for $\omega_1 + \Omega_1 = 0$. The subsequent freedom of the tetrad is then $\mathbf{e}_1 \to \pm \mathbf{e}_1$ and (2.6) with $\partial_0\Theta = 0$.

It is possible to specialize the tetrad even further. We could, for example, arrange for $\gamma^3{}_{12} = n_{33} = -(\hat{\theta}_{23} - \hat{\Omega})$ to vanish everywhere. Under a transformation (2.6) with $\delta = 1$, $\gamma^3{}_{12} = n_{33} \to n_{33}\cos^2\Theta + n_{22}\sin^2\Theta - 2n_{23}\sin\Theta\cos\Theta + \partial_1\Theta$, which can be made to vanish on a hypersurface transverse to the fluid flow, by appropriate propagation of Θ. Subsequent application of the Jacobi identities [Eq. (81) of Ref. 13, which requires that $\partial_0 n_{33} + \frac{1}{3}n_{33}\theta = 0$] shows that $n_{33} \equiv 0$. This tetrad choice corresponds to that of Ellis.[3] We prefer to make the choice $\gamma^2{}_{31} - \gamma^3{}_{12} = n_{22} - n_{33} = 2\hat{\sigma}_{23} = 0$ instead. Under a transformation (2.6) with $\delta = 1$, $\gamma^2{}_{31} - \gamma^3{}_{12} = n_{22} - n_{33} \to (n_{22} - n_{33})\cos 2\Theta + 2n_{23}\sin 2\Theta$, and so we can arrange for $n_{22} - n_{33}$ to vanish on a hypersurface transverse to the fluid flow, and then apply the Jacobi identities [Eq. (81) of Ref. 13, which requires that $\partial_0(n_{22} - n_{33}) + \frac{1}{3}(n_{22} - n_{33})\theta = 0$] to deduce that $n_{22} - n_{33} = 2\hat{\sigma}_{23} \equiv 0$. Our preference for this tetrad is predicated on a choice which favors neither $\mathbf{e}_2$ nor $\mathbf{e}_3$ over the other, and diagonalizing the matrix $\hat{\sigma}_{AB}$ provides a simple way of doing this. With this more convenient tetrad choice, the remaining freedom is then $\mathbf{e}_1 \to \pm \mathbf{e}_1$ and (2.6) with $\Theta = k\pi/2$, where $k = 0, 1, 2$, or 3 in the general case where $\hat{\sigma}_{AB} \not\equiv 0$, whereas if $\hat{\sigma}_{AB} \equiv 0$ (so that $n_{22} - n_{33} = 0$ in all allowed tetrads) the freedom is unchanged, i.e., it is still $\mathbf{e}_1 \to \pm \mathbf{e}_1$ and (2.6) with $\partial_0\Theta = 0$. Since the basis vectors $\mathbf{e}_2$ and $\mathbf{e}_3$ are treated on an equal footing, our choice of tetrad allows the various components of the Jacobi identities and field equations to be checked against each other, by means of the allowed transformations (e.g., under the rotation $\mathbf{e}_1 \to \mathbf{e}_1$, $\mathbf{e}_2 \to \mathbf{e}_3$, and $\mathbf{e}_3 \to -\mathbf{e}_2$, the quantities $\dot{u}$, θ, $\hat{\Omega}$, $\hat{\theta}$, ω, and n are invariant, whereas $d_2 \to d_3$, $d_3 \to -d_2$, $A_2 \to A_3$, $A_3 \to -A_2$, and $\hat{\sigma}_{22} \to -\hat{\sigma}_{22}$). Indeed, use of this technique in Sec. 3 allows us to avoid some tedious calculations.

The Jacobi identities, Einstein field equations, Bianchi identities, and commutation relations in our tetrad and present notation are written out in the Appendix.

3. THE MAIN RESULT

In this section, we prove the main theorem, viz., that a shear-free perfect fluid with equation of state $p = p(\mu)$ (and with $\mu + p \not\equiv 0$), in which the vorticity (ω) and acceleration ($\dot{u}$) are parallel, has either zero vorticity or zero expansion (θ). The proof of this result is obtained by assuming that $\omega\theta \not\equiv 0$, and then deriving a contradiction. The proof of the theorem divides naturally into two cases, according as $\dot{u} \not\equiv 0$ or $\dot{u} \equiv 0$. The latter alternative is equivalent to p being identically constant. For the Bianchi identities (2.2) imply that $\dot{u} = 0 \Leftrightarrow \partial_\alpha p = 0$. Hence if p is identically constant, clearly $\dot{u} \equiv 0$, whereas if $\dot{u} \equiv 0$, the $[\mathbf{e}_\alpha, \mathbf{e}_\beta]$ commutator acting on p implies that $\omega\partial_0 p \equiv 0$, and so (assuming $\omega \not\equiv 0$) p is identically constant. This case is therefore equivalent to the situation treated by Ellis[3], who investigated shear-free dust ($p \equiv 0$), since any nonzero constant pressure component can be absorbed into the cosmological term (with a corresponding adjustment of the energy density). For reasons explained in the introduction, we shall provide a proof of the result both when $\dot{u} \not\equiv 0$ and when $\dot{u} \equiv 0$.

Theorem 3.1: Consider a shear-free perfect fluid in general relativity, with an equation of state $p = p(\mu)$ satisfying $\mu + p \not\equiv 0$. Suppose that the vorticity and acceleration are nonzero and parallel. Then the fluid's volume expansion scalar is zero.

Proof: We shall suppose that, under the conditions of the theorem, the expansion scalar θ is nonzero, and derive a contradiction. By applying the commutation relations to the function F, we obtain propagation equations for $\dot{u}$, viz., (2.5) and (2.7). It follows from (2.7) that since $\dot{u}_2 \equiv \dot{u}_3 \equiv 0$,

$$\partial_2\theta = \partial_3\theta = 0, \tag{3.1}$$

where use is made of the Bianchi identities (A26) and (A27) and of the fact that p is not identically constant to deduce that $\partial_2 p' = \partial_3 p' = 0$. Hence from the field equations (A16) and (A17),

$$\partial_2\omega = d_2\omega \quad \text{and} \quad \partial_3\omega = d_3\omega. \tag{3.2}$$

Applying the $[\mathbf{e}_2, \mathbf{e}_3]$ commutator (A32) to θ, and using (3.1) and (A15), we obtain

$$\partial_0\theta = \tfrac{3}{4}n^2, \tag{3.3}$$

since $\omega \not\equiv 0$. Similarly, the $[\mathbf{e}_1, \mathbf{e}_2]$ and $[\mathbf{e}_3, \mathbf{e}_1]$ commutators (A31) and (A33) applied to θ require, using (3.1), (3.2), (A15), and the fact that $\omega \not\equiv 0$,

$$\partial_2 n = \partial_3 n = 0. \tag{3.4}$$

Using the $[\mathbf{e}_2, \mathbf{e}_3]$ commutator (A32) on F, we obtain from (2.5) that

$$2\omega\, p'\theta = n\dot{u}, \tag{3.5}$$

from which $n \not\equiv 0$, by our assumptions that $\omega\theta \not\equiv 0$ and $\dot{u} \not\equiv 0$. Now applying the $[\mathbf{e}_2, \mathbf{e}_3]$ commutator (A32) to ω and simplifying with the aid of (3.2), (3.5), (A1), (A2), and (A5), we obtain

$$\partial_1 n = -\tfrac{2}{3}\omega\theta. \tag{3.6}$$

The next step is to obtain a purely algebraic relation from the (00) field equation (A14), by substituting for $\partial_0\theta$ from (3.3) and by obtaining an equation for $\partial_1\dot{u}$ from differentiation of Eq. (3.5). This differentiation yields

$$\partial_1\dot{u} = 3p'\omega^2 - \dot{u}\hat{\theta} - (1/p')[(p''/p')(\mu + p) - p' - \tfrac{1}{3}]\dot{u}^2, \tag{3.7}$$

where use is made of (3.5), (3.6), (A1), (A15), and (A25). The (00) field equation (A14) becomes

$$\begin{aligned}\tfrac{3}{4}n^2 + \tfrac{1}{3}\theta^2 - 2\omega^2(1 + \tfrac{3}{2}p') \\ + (1/p')[(p''/p')(\mu + p) - 2p' - \tfrac{1}{3}]\dot{u}^2 \\ + \tfrac{1}{2}(\mu + 3p - 2\Lambda) = 0.\end{aligned} \tag{3.8}$$

We next show that the requirement that (3.8) be propagated along the $\mathbf{e}_2$ and $\mathbf{e}_3$ directions implies that $d_2 = d_3 = 0$. For this purpose, we shall need the equations

$$\partial_2 \dot{u} = d_2 \dot{u} \quad \text{and} \quad \partial_3 \dot{u} = d_3 \dot{u}, \tag{3.9}$$

obtained by applying the $[\mathbf{e}_1, \mathbf{e}_2]$ and $[\mathbf{e}_3, \mathbf{e}_1]$ commutators to the function F [cf. (2.5)].

Differentiating (3.8) along $\mathbf{e}_2$ and using (3.1), (3.2), (3.4), (3.9), (A26), and (A27), we have

$$d_2\{2\omega^2(1 + \tfrac{3}{2}p') - (1/p')[(p''/p')(\mu + p) - 2p' - \tfrac{1}{3}]\dot{u}^2\} = 0, \tag{3.10a}$$

and similarly

$$d_3\{2\omega^2(1 + \tfrac{3}{2}p') - (1/p')[(p''/p')(\mu + p) - 2p' - \tfrac{1}{3}]\dot{u}^2\} = 0. \tag{3.10b}$$

Note that we may regard (3.10b) as being derived by differentiation of (3.8) along the $\mathbf{e}_3$ direction, but that it can also be readily obtained from (3.10a) using the symmetry of the tetrad, and a rotation $\mathbf{e}_1 \to \mathbf{e}_1$, $\mathbf{e}_2 \to \mathbf{e}_3$, $\mathbf{e}_3 \to -\mathbf{e}_2$. If in (3.10)

$$2\omega^2(1 + \tfrac{3}{2}p') - (1/p')[(p''/p')(\mu + p) - 2p' - \tfrac{1}{3}]\dot{u}^2 = 0, \tag{3.11}$$

then the propagation of (3.8) along $\mathbf{e}_0$ implies, using (3.3) and (A11), that $(1 + 3p')(\mu + p)\theta = 0$. Thus $p' = -\tfrac{1}{3}$, since $(\mu + p)\theta \not\equiv 0$ by assumption. But then Eq. (3.11) requires $\omega^2 + \dot{u}^2 = 0$, which is a contradiction. Hence Eq. (3.11) is false, and so Eqs. (3.10) imply that

$$d_2 = d_3 = 0. \tag{3.12}$$

In view of (3.12), the $[\mathbf{e}_2, \mathbf{e}_3]$ commutator (A32) applied to ω yields, on using (3.2), (3.5), (A1), and (A5),

$$n\hat{\theta} = \tfrac{4}{3}\omega\theta, \tag{3.13}$$

and, hence, by (3.5) and the fact that $n \not\equiv 0$,

$$\dot{u} = \tfrac{3}{2}p'\hat{\theta}. \tag{3.14}$$

Now recalling (3.1), (3.2), (3.4), and (3.12), we see that propagation of Eq. (3.13) along $\mathbf{e}_2$ and $\mathbf{e}_3$ implies that

$$\partial_2\hat{\theta} = \partial_3\hat{\theta} = 0 \tag{3.15}$$

(where again we make use of the fact that $n \not\equiv 0$). If we apply the $[\mathbf{e}_2, \mathbf{e}_3]$ commutator (A32) to $\hat{\theta}$ and use (3.13), (3.14), (3.15), (A6), (A15), and the fact that $n \not\equiv 0$, we obtain

$$\partial_1\hat{\theta} = 2\omega^2 + \tfrac{1}{2}(3p' - 1)\hat{\theta}^2. \tag{3.16}$$

However, if we eliminate $\partial_1\dot{u}$ from the field equations (A14) and (A18), and use Eqs. (3.3), (3.14), and (3.16), we find that

$$\mu + p + 2\hat{\sigma}_{22}{}^2 = 0. \tag{3.17}$$

Since Eq. (3.17) requires $\mu + p \leqslant 0$, it could be regarded as a contradiction. However, the conditions of the theorem involve the simpler assumption that $\mu + p \not\equiv 0$, and, as such, (3.17) is potentially valid. If (3.17) is to hold, then differentiating in the $\mathbf{e}_0$ direction gives us $p' = -\tfrac{1}{3}$, using (3.17), (A13), (A24), and the assumption that $(\mu + p)\theta \not\equiv 0$. In this case, Eq. (3.8) implies

$$\tfrac{3}{4}n^2 + \tfrac{1}{3}\theta^2 - \omega^2 - \dot{u}^2 + \tfrac{1}{2}(\mu + 3p - 2\Lambda) = 0,$$

whose propagation along $\mathbf{e}_0$ implies [using (2.7), (3.3), (3.5), (A5), (A11), and (A24)] that $\omega^2 + \dot{u}^2 = 0$, a contradiction. Hence, under the conditions of the theorem, our assumption that $\theta \not\equiv 0$ is shown to be false. □

Remark: An alternative proof of this theorem involves using (3.5) as a starting point and investigating its propagation along the $\mathbf{e}_0$ direction.[7]

The special case in which $\dot{u} \equiv 0$ is considered in the next theorem.

Theorem 3.2 (Ellis[3]): Consider a shear-free perfect fluid in general relativity, in which the fluid flow is geodesic and in which $\mu + p \not\equiv 0$. Suppose that the vorticity is nonzero. Then the fluid's volume expansion scalar vanishes.

Proof: As previously observed, under the conditions of the theorem, the pressure p is identically constant. Since the acceleration must vanish, the field equations (A14) and (A15) provide equations for propagating θ along $\mathbf{e}_0$ and $\mathbf{e}_1$, and for compatibility it follows that

$$\partial_1\mu = -8\omega^2\hat{\theta}, \tag{3.18}$$

by virtue of the commutator (A28) and Eqs. (A1), (A5), (A11), and (A15). Similarly, the Bianchi identity (A24) and Eq. (3.18) provide equations for propagating μ along $\mathbf{e}_0$ and $\mathbf{e}_1$, and the commutator (A28) applied to μ yields the algebraic relation

$$8\omega(\tfrac{1}{3}\theta\hat{\theta} - n\omega) + \tfrac{3}{2}(\mu + p)n = 0, \tag{3.19}$$

where use is made of (3.18), (A5), (A6), (A15), and the assumption that $\omega \not\equiv 0$. We now differentiate Eq. (3.19) along $\mathbf{e}_0$, and use (A5), (A6), (A11), (A14), (A15), and (A24), together with Eq. (3.19) and the assumption that $\omega \not\equiv 0$. We obtain

$$2n\omega\theta + \hat{\theta}[2\omega^2 - \tfrac{1}{2}(\mu + 3p - 2\Lambda)] = 0. \tag{3.20}$$

Differentiation of this along $\mathbf{e}_0$ yields, in a similar manner,

$$\theta\hat{\theta}[\tfrac{2}{3}\omega^2 + p - \Lambda] = 3n\omega[2\omega^2 - \tfrac{1}{2}(\mu + 3p - 2\Lambda)]. \tag{3.21}$$

We shall now assume that $\theta \not\equiv 0$, and arrive at a contradiction. We first show that Eqs. (3.19)–(3.21) imply that

$$n = \hat{\theta} = 0. \tag{3.22}$$

Multiply (3.21) by ω and substitute (3.19) to obtain

$$n[4\omega^4 - \tfrac{3}{8}\omega^2(3\mu + 19p - 16\Lambda) + \tfrac{9}{16}(\mu + p)(p - \Lambda)] = 0. \tag{3.23}$$

If $n \not\equiv 0$, then (3.20) shows that $\hat{\theta} \not\equiv 0$, since otherwise $\omega\theta \equiv 0$. Moreover, Eq. (3.23) shows that ω and μ are functionally dependent, and hence, using (3.18), (A1), (A5), and (A24), we obtain $\omega^2 = \tfrac{3}{16}(\mu + p)$, whose propagation along $\mathbf{e}_0$ requires that $\omega\theta \equiv 0$, a contradiction. Thus $n \equiv 0$, and so, by (3.19), $\hat{\theta} \equiv 0$, and Eq. (3.22) results.

Next, we apply the $[\mathbf{e}_0, \mathbf{e}_2]$ commutator to ω, to yield

$$\partial_0(\partial_2\omega) = -\theta\,\partial_2\omega + \omega\,\partial_3\omega - d_3\omega^2, \tag{3.24a}$$

where use is made of Eq. (A16). Similarly, invoking a rotation $\mathbf{e}_1 \to \mathbf{e}_1$, $\mathbf{e}_2 \to \mathbf{e}_3$, $\mathbf{e}_3 \to -\mathbf{e}_2$, we have

$$\partial_0(\partial_3\omega) = -\theta\,\partial_3\omega - \omega\,\partial_2\omega + d_2\omega^2. \tag{3.24b}$$

We now apply the $[\mathbf{e}_0, \mathbf{e}_2]$ commutator to θ and use Eqs. (3.24b), (A5), (A8), (A14), and (A16), thereby obtaining

$$\partial_2\mu = \tfrac{13}{2}\omega\,\partial_2\omega + \tfrac{3}{2}d_2\omega^2, \tag{3.25a}$$

and similarly

$$\partial_3\mu = \tfrac{13}{2}\,\omega\,\partial_3\omega + \tfrac{3}{2}\,d_3\omega^2. \qquad (3.25b)$$

The $[\mathbf{e}_0, \mathbf{e}_2]$ commutator applied to μ yields

$$3(\partial_3\omega - d_3\omega)[29\omega^2 - 6(\mu + p)] - \omega\theta\,(26\partial_2\omega + 6d_2\omega) = 0, \qquad (3.26a)$$

where we have used Eqs. (3.24a), (A5), (A7), (A16), and (A24). Similarly,

$$3(\partial_2\omega - d_2\omega)[29\omega^2 - 6(\mu + p)] + \omega\theta\,(26\partial_3\omega + 6d_3\omega) = 0. \qquad (3.26b)$$

If we regard Eqs. (3.26) as linear algebraic equations for the unknowns $29\omega^2 - 6(\mu + p)$ and $\omega\theta$, we see that, since $\omega\theta \not\equiv 0$, the determinant of the coefficients must vanish, i.e.,

$$26[(\partial_2\omega)^2 + (\partial_3\omega)^2] - 20\omega[d_2\partial_2\omega + d_3\partial_3\omega] - 6(d_2{}^2 + d_3{}^2)\omega^2 = 0. \qquad (3.27)$$

Differentiation of (3.27) along $\mathbf{e}_0$ now requires

$$d_2\partial_3\omega - d_3\partial_2\omega = 0, \qquad (3.28)$$

where we have used Eqs. (3.24), (A5), (A7), (A8), (A16), and (A17).

Finally, we apply the $[\mathbf{e}_2, \mathbf{e}_3]$ commutator to μ, using (3.25), (3.28), (A2), (A5), and (A24), to deduce that

$$29\omega^2 = 6(\mu + p), \qquad (3.29)$$

whose propagation along $\mathbf{e}_0$ requires, using (A5) and (A24), that $\omega\theta \equiv 0$, a contradiction.

It therefore follows that under the conditions of the theorem, the expansion scalar θ must be zero.

Theorems 3.1 and 3.2 show that under certain circumstances, the condition $\sigma \equiv 0$ implies $\omega\theta \equiv 0$. This result is trivially true in the case when $\omega \equiv 0$, so we can amalgamate the results of Theorems 3.1 and 3.2 to deduce the following:

Theorem 3.3: Any shear-free perfect fluid in general relativity with an equation of state $p = p(\mu)$, such that $\mu + p \not\equiv 0$, has either vanishing vorticity or vanishing expansion (i.e., $\sigma \equiv 0 \Rightarrow \omega\theta \equiv 0$), provided that the vorticity and acceleration are parallel (and possibly zero).

4. DISCUSSION

If the conjecture of Sec. 1 were true, it would be natural to investigate the space-times arising in the three cases

(i) $\sigma \equiv \omega \equiv 0, \quad \theta \not\equiv 0,$
(ii) $\sigma \equiv \omega \equiv \theta \equiv 0,$
(iii) $\sigma \equiv \theta \equiv 0, \quad \omega \not\equiv 0,$

where the matter content is a perfect fluid with an equation of state $p = p(\mu)$ satisfying $\mu + p \not\equiv 0$. Case (i) is that considered recently by Collins and Wainwright,[12] in which all resulting space-times were obtained. The procedure used was to specialize the results of Barnes[17] on shear-free normal flows of a perfect fluid to the situation when there is an equation of state $p = p(\mu)$, with $\mu + p \not\equiv 0$. The resulting space-times were first shown to be necessarily "locally rationally symmetric,"[3,18] and it was then deduced that either they were FRW models (in the case $\dot{u} \equiv 0$), or that they were spherically symmetric or plane symmetric (and hypersurface-homogeneous). We have found a somewhat more satisfactory derivation of this result which uses the orthonormal tetrad formalism and a more fundamental starting point (rather than assuming the specific coordinate system and the results derived by Barnes[17]).

The space-times resulting in cases (ii) and (iii) are less clearly understood. In case (ii), the space-times are static, and the flow is irrotational. In the case when $\dot{u} \equiv 0$, the solutions are FRW and hence are Einstein static models, generalized to include a pressure term. In the case when $\dot{u} \not\equiv 0$, and the solutions are *algebraically degenerate*, the space-times appear in the set of exact solutions given by Barnes.[17,19] The existence and nature of algebraically general solutions in case (ii) does not appear to have been investigated yet. We have considered this question, together with the question of the isometry group, in both the most general case and in various special cases. We have also made some progress toward understanding the nature of the solutions in case (iii) when $\boldsymbol{\omega}$ and $\dot{\mathbf{u}}$ are parallel, and specifically with regard to allowed Petrov types and isometry groups (cf. the results of Ellis[3] in the case of dust). It is our intention to present these results in a future article, using a procedure which allows simultaneous discussion of cases (i), (ii), and (iii) in a uniform manner.

It is interesting to note that the conjecture of Sec. 1 is entirely *general relativistic* in nature, for, as noted by Ellis,[1] there are many corresponding Newtonian solutions, even when the acceleration is zero, in which $\sigma \equiv 0$ yet $\omega\theta \not\equiv 0$. It is also of interest to note that the conclusion of the conjecture, viz., that $\omega\theta \equiv 0$, requires that either the fluid-flow vector or the vorticity vector be hypersurface-orthogonal, since if $\omega \not\equiv 0$ then $\theta \equiv 0$, and so by choosing a tetrad in which $\boldsymbol{\omega} = (\omega, 0, 0)$ and $\boldsymbol{\omega} + \boldsymbol{\Omega} = \mathbf{0}$ as in Sec. 2, the (01) field equation [Eq. (83) of Ref. 13) implies that $n = 0$ and hence the hypersurface orthogonality of $\boldsymbol{\omega}$.

Ellis[3] has discussed the physics of shear-free dust, and similar remarks apply for shear-free perfect fluids. For a *null* congruence, the propagation equations for the expansion and vorticity contain terms involving the expansion, the shear, the vorticity, and the Ricci tensor (but no Weyl tensor terms); on the other hand, the propagation equations for the shear contain not only terms involving the expansion, shear, vorticity, and the Ricci tensor, but *also* the Weyl tensor. Thus the free gravitational field (or the "news") enters the evolution of the congruence by way of the shear, which then plays a role in driving the other "kinematic" quantities. We can say exactly the same for a *timelike* congruence associated with a perfect fluid. Thus, in either case, for a *shear-free* congruence, the free gravitational field is prevented from exerting an influence on the evolution of the congruence. Furthermore, by the well-known theorem of Goldberg and Sachs,[20] any *vacuum* space-time containing a shear-free *null geodesic* congruence is restricted, since it must be algebraically degenerate. The theorem of Ellis[3] (cf. our Theorem 3.2) provides an analogous result: any *geodesic* shear-free perfect fluid space-time is restricted, since the flow must have $\omega\theta \equiv 0$.

ACKNOWLEDGMENTS

This research was partially supported by an Individual Operating Grant #A3978 (C. B. C.) and by a Postgraduate Scholarship (A. J. W.), both from the Natural Sciences and Engineering Research Council of Canada. We are grateful to Helen Warren for her careful typing of the manuscript.

APPENDIX

Jacobi identities

$$\partial_1 n = \omega[\dot{u} - \hat{\theta}], \tag{A1}$$

$$\partial_1 n + \partial_2 d_3 - \partial_3 d_2 - \tfrac{2}{3}\omega\theta + n\hat{\theta} - d_2 A_3 + d_3 A_2 = 0, \tag{A2}$$

$$\partial_1 A_2 + \partial_2 \hat{\sigma}_{22} - \tfrac{1}{2}\partial_2\hat{\theta} - \partial_3\hat{\Omega} + \tfrac{1}{2}(d_2 + A_2)\hat{\theta} - (d_2 - A_2)\hat{\sigma}_{22} + \hat{\Omega}(d_3 - A_3) = 0, \tag{A3}$$

$$\partial_1 A_3 - \partial_3 \hat{\sigma}_{22} - \tfrac{1}{2}\partial_3\hat{\theta} + \partial_2\hat{\Omega} + \tfrac{1}{2}(d_3 + A_3)\hat{\theta} + (d_3 - A_3)\hat{\sigma}_{22} - \hat{\Omega}(d_2 - A_2) = 0, \tag{A4}$$

$$\partial_0\omega = (p' - \tfrac{2}{3})\omega\theta, \tag{A5}$$

$$\partial_0\hat{\theta} - \tfrac{2}{3}\partial_1\theta + \tfrac{1}{3}\theta(\hat{\theta} - 2\dot{u}) = 0, \tag{A6}$$

$$\partial_0 d_2 + \tfrac{1}{3}\partial_2\theta + \tfrac{1}{3}d_2\theta = 0, \tag{A7}$$

$$\partial_0 d_3 + \tfrac{1}{3}\partial_3\theta + \tfrac{1}{3}d_3\theta = 0, \tag{A8}$$

$$\partial_0 A_2 - \tfrac{1}{3}\partial_2\theta + \tfrac{1}{3}A_2\theta = 0, \tag{A9}$$

$$\partial_0 A_3 - \tfrac{1}{3}\partial_3\theta + \tfrac{1}{3}A_3\theta = 0, \tag{A10}$$

$$\partial_0 n + \tfrac{1}{3}n\theta = 0, \tag{A11}$$

$$\partial_0\hat{\Omega} + \tfrac{1}{3}\hat{\Omega}\theta = 0, \tag{A12}$$

$$\partial_0\hat{\sigma}_{22} + \tfrac{1}{3}\hat{\sigma}_{22}\theta = 0. \tag{A13}$$

Equation (A5) has been obtained from (2.1), i.e., by invoking (2.5).

Field equations

(00):
$$\partial_0\theta + \tfrac{1}{3}\theta^2 - 2\omega^2 - \partial_1\dot{u} - \dot{u}^2 - \dot{u}\hat{\theta} + \tfrac{1}{2}(\mu + 3p - 2\Lambda) = 0. \tag{A14}$$

(0α):
$$\tfrac{2}{3}\partial_1\theta - n\omega = 0, \tag{A15}$$

$$\tfrac{2}{3}\partial_2\theta + \partial_3\omega - d_3\omega = 0, \tag{A16}$$

$$\tfrac{2}{3}\partial_3\theta - \partial_2\omega + d_2\omega = 0. \tag{A17}$$

$(\alpha\beta)$:
$$\partial_1\hat{\theta} - \partial_2 d_2 - \partial_3 d_3 + 2\hat{\sigma}_{22}{}^2 + \tfrac{1}{2}\hat{\theta}^2 + d_2{}^2 + d_3{}^2 - d_2 A_2 - d_3 A_3 - \tfrac{1}{2}n^2 = \tfrac{1}{3}\partial_0\theta - \partial_1\dot{u} - \dot{u}^2 + \tfrac{1}{3}\theta^2 - \tfrac{1}{2}(\mu - p) - \Lambda, \tag{A18}$$

$$\partial_1 A_2 - \partial_2\hat{\sigma}_{22} + \tfrac{1}{2}\partial_2\hat{\theta} - \partial_3\hat{\Omega} + \partial_3 n - 2nd_3 + (d_3 - A_3)\hat{\Omega} + \tfrac{1}{2}(d_2 + A_2)\hat{\theta} - (d_2 + 3A_2)\hat{\sigma}_{22} = -\partial_2\dot{u} - d_2\dot{u}, \tag{A19}$$

$$\partial_1 A_3 + \partial_3\hat{\sigma}_{22} + \tfrac{1}{2}\partial_3\hat{\theta} + \partial_2\hat{\Omega} - \partial_2 n + 2nd_2 - (d_2 - A_2)\hat{\Omega} + \tfrac{1}{2}(d_3 + A_3)\hat{\theta} + (d_3 + 3A_3)\hat{\sigma}_{22} = -\partial_3\dot{u} - d_3\dot{u}, \tag{A20}$$

$$\partial_1\hat{\sigma}_{22} + \tfrac{1}{2}\partial_1\hat{\theta} - \partial_2 d_2 + \partial_2 A_2 + \partial_3 A_3 + \hat{\sigma}_{22}\hat{\theta} + \tfrac{1}{2}\hat{\theta}^2 - d_3 A_3 + A_2^2 + A_3^2 + d_2^2 + \tfrac{1}{2}n^2 - n\hat{\Omega} = -\dot{u}(\hat{\sigma}_{22} + \tfrac{1}{2}\hat{\theta}) - \tfrac{1}{2}(\mu - p) - \Lambda + \tfrac{1}{3}\theta^2 + \tfrac{1}{3}\partial_0\theta + 2\omega^2, \tag{A21}$$

$$-\partial_1\hat{\sigma}_{22} + \tfrac{1}{2}\partial_1\hat{\theta} - \partial_3 d_3 + \partial_2 A_2 + \partial_3 A_3 - \hat{\sigma}_{22}\hat{\theta} + \tfrac{1}{2}\hat{\theta}^2 - d_2 A_2 + A_2^2 + A_3^2 + d_3^2 + \tfrac{1}{2}n^2 - n\hat{\Omega} = \dot{u}(\hat{\sigma}_{22} - \tfrac{1}{2}\hat{\theta}) - \tfrac{1}{2}(\mu - p) - \Lambda + \tfrac{1}{3}\theta^2 + \tfrac{1}{3}\partial_0\theta + 2\omega^2, \tag{A22}$$

$$\partial_2 d_3 + \partial_3 d_2 - 4\hat{\Omega}\hat{\sigma}_{22} - 2d_2 d_3 - d_2 A_3 - d_3 A_2 + 2n\hat{\sigma}_{22} = 0. \tag{A23}$$

Bianchi identities

$$\partial_0\mu + (\mu + p)\theta = 0, \tag{A24}$$

$$\partial_1 p + (\mu + p)\dot{u} = 0, \tag{A25}$$

$$\partial_2 p = 0, \tag{A26}$$

$$\partial_3 p = 0. \tag{A27}$$

Commutation relations

$$\begin{array}{lccccr}
[\mathbf{e}_0, \mathbf{e}_1] = & & \dot{u}\mathbf{e}_0 - \tfrac{1}{3}\theta\mathbf{e}_1, & & & \text{(A28)}\\
[\mathbf{e}_0, \mathbf{e}_2] = & & & -\tfrac{1}{3}\theta\mathbf{e}_2, & & \text{(A29)}\\
[\mathbf{e}_0, \mathbf{e}_3] = & & & & -\tfrac{1}{3}\theta\mathbf{e}_3, & \text{(A30)}\\
[\mathbf{e}_1, \mathbf{e}_2] = & & -d_2\mathbf{e}_1 & -(\hat{\sigma}_{22} + \tfrac{1}{2}\hat{\theta})\mathbf{e}_2 & +\hat{\Omega}\mathbf{e}_3, & \text{(A31)}\\
[\mathbf{e}_2, \mathbf{e}_3] = & -2\omega\mathbf{e}_0 & +n\mathbf{e}_1 & +A_3\mathbf{e}_2 & -A_2\mathbf{e}_3, & \text{(A32)}\\
[\mathbf{e}_3, \mathbf{e}_1] = & & d_3\mathbf{e}_1 & +\hat{\Omega}\mathbf{e}_2 & -(\hat{\sigma}_{22} - \tfrac{1}{2}\hat{\theta})\mathbf{e}_3. & \text{(A33)}
\end{array}$$

[1] G. F. R. Ellis, "Relativistic Cosmology," in *General Relativity and Cosmology*, Proceedings of the International School of Physics "Enrico Fermi" Course XLVII, 1969, edited by R. K. Sachs (Academic, London and New York, 1971), p. 104.
[2] K. Gödel, Proc. Int. Cong. Math. **1**, 175 (1950).
[3] G. F. R. Ellis, J. Math. Phys. **8**, 1171 (1967).
[4] E. Schücking, Naturwiss. **19**, 507 (1957).
[5] S. Banerji, Prog. Theor. Phys. **39**, 365 (1968).
[6] A. R. King and G. F. R. Ellis, Commun. Math. Phys. **31**, 209 (1973).
[7] A. J. White, "Shear-Free Perfect Fluids in General Relativity," M. Math. thesis, University of Waterloo, Waterloo, Ontario, Canada, 1981.
[8] R. Treciokas and G. F. R. Ellis, Commun. Math. Phys. **23**, 1 (1971).
[9] M. Wyman, Phys. Rev. **70**, 396 (1946).
[10] K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[11] A. Krasinski, Acta Phys. Polon. B **5**, 411 (1974); Acta Phys. Polon. B **6**, 223 (1975).
[12] C. B. Collins and J. Wainwright, Phys. Rev. D **27**, 1209 (1983).
[13] M. A. H. MacCallum, "Cosmological Models from a Geometric Point of View," in *Cargèse Lectures in Physics*, Vol. 6, Lectures at the International Summer School of Physics, Cargèse, Corsica, 1971, edited by E. Schatzman (Gordon and Breach, New York, 1973), p. 61.
[14] P. J. Greenberg, J. Math. Anal. Appl. **30**, 128 (1970).
[15] D. A. Szafron, "Inhomogeneous Cosmologies," Ph.D. thesis, University of Waterloo, Waterloo, Ontario, Canada, 1978.
[16] M. Tsamparlis and D. P. Mason, J. Math. Phys. **24**, 1577 (1983).
[17] A. Barnes, Gen. Rel. Grav. **4**, 105 (1973).
[18] J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).
[19] A. Barnes, J. Phys. A **5**, 374 (1972).
[20] J. N. Goldberg and R. K. Sachs, Acta Phys. Polon. **22**, 13 (1962).

The gravitational fields of some rotating and nonrotating cylindrical shells of matter

John Stachel[a)]

Einstein Project, Princeton University Press, 41 William Street, Princeton, New Jersey 08540 and Department of Physics, Princeton University, Princeton, New Jersey 08540

(Received 5 October 1982; accepted for publication 28 January 1983)

The exterior and flat interior metrics of an infinite cylindrical shell (surface layer) of matter are studied, for nonrotating and rigidly rotating shells. Relations between the parameters characterizing the exterior metric and the components of the shell's stress-energy tensor are established. It is shown that one of these parameters, characterizing the "conicality" of the exterior field, is related to the energy density of the source; while another, characterizing local nonflatness, is related to certain components of the stress tensor. A one-parameter family of locally flat but conical exterior metrics generated by a particular type of massive cylindrical shell is exhibited. The globally stationary but locally static exterior field of a rigidly rotating shell is studied. The (nonlocal) parameter of the exterior field characterizing the rotation is related to the rate of rotation of the shell as defined by the flat interior metric.

PACS numbers: 04.20.Jb

I. NONROTATING SHELL

The exterior gravitational field of an infinite rotating cylinder of matter in general relativity is globally stationary but locally static,[1] just as the analogous electromagnetic field of a rotating charged cylinder is locally electrostatic but shows nonlocal effects of the rotation which may be verified through the Aharonov–Bohm effect.[1] The verification of an analogous gravitational effect may be carried out using classical light waves which are entirely confined to the region outside the cylinder of matter.[1] However, to study the relationship between the exterior nonlocal parameter characterizing the rotational effects (the period or line integral of a certain one-form, closely related to the globally timelike Killing vector field of the metric, taken around any closed curve encircling the cylinder of matter) and the actual rotation of the source, some model of the latter is needed.

In this paper, I shall study a very simple source: an infinite cylindrical shell (i.e., surface layer) of matter inside of which there is a flat space-time. Since the interior is flat, special-relativistic considerations are all that are needed to define the rate of rotation of the shell. In the rest of this section, I shall discuss a nonrotating shell of matter, giving rise to a globally static exterior gravitational field. Discussion of the matching of the exterior static field to the shell stress-energy tensor will enable relationships to be established between the components of that tensor and the parameters which characterize such a static exterior metric. In the second section, I shall discuss how the static solutions of this section may be used to generate the globally stationary but locally static exterior metric of rigidly rotating shells; and the relationship between the nonlocal exterior parameter mentioned above and the rate of rotation of the shell.

I shall use units in which $\mathbf{c} = 8\pi G = 1$, and signature $+ - - -$, so that timelike vectors have positive norm.

Let an infinite cylindrical shell of matter separate a flat interior region from an exterior gravitational field created by the shell. The empty-space exterior and (flat) interior fields must be matched at the shell surface in such a way that the conditions for a surface layer of matter in general relativity[2] are satisfied on the shell. I shall first consider the case of a nonrotating shell of matter, and then discuss how these results can be modified for the case of a rigidly rotating shell.

In the nonrotating case, the exterior metric will be static with whole cylinder symmetry. The metric of such a field may always be put in the form[3]

$$\begin{aligned} ds^2 = {} & \exp[2(\gamma - \psi)][(dx^0)^2 - (dx^1)^2] \\ & - (x^1)^2 \exp(-2\psi)(dx^3)^2 \\ & - \exp[2(\psi + \mu)](dx^2)^2, \\ & -\infty < x^0 < \infty, \quad x_0^1 \leqslant x^1 < \infty, \\ & -\infty < x^2 < \infty, \quad 0 \leqslant x^3 < 2\pi. \end{aligned} \tag{1.1}$$

Here x^0 is a timelike coordinate, x^1 is analogous to the cylindrical radial coordinate, x^2 is analogous to the cylindrical axial coordinate, and x^3 is analogous to the cylindrical angular coordinate; ψ, γ, and μ are functions of (x^0, x^1) only; and $(x^1)\exp\mu$ obeys the (flat) two-dimensional wave equation in (x^0, x^1). In the static case, this requires that

$$\exp\mu = A + B/x^1. \tag{1.2}$$

For the flat interior region of the cylinder, assumed to be of radius x_0^1, we may choose $A = 1$, $B = 0$, and $\psi = \gamma = 0$, to get the usual form of the flat cylindrical metric:

$$ds_-^2 = (dx^0_-)^2 - (dx^1_-)^2 - (x^1_-)^2(dx^3_-)^2 - (dx^2_-)^2. \tag{1.3}$$

(From now on, minus signs will denote quantities having to do with the the region interior to the shell, and plus signs will denote quantities connected with the exterior of the shell.) However, as we shall see, it is simpler not to make the same assumption about the exterior values of A and B if we want to match with a shell having nonvanishing mass density. Consideration of the matching conditions for the shell (agreement of the first fundamental forms of the interior and exterior metrics on the shell, and difference of the second fundamental forms proportional to the surface stress-energy tensor)[2]

[a)] On leave from Boston University.

 0022-2488/84/020338-04$02.50

shows that in this case all the components of the metric may be made continuous across the shell on which $x^1_+ = x^1_0 > 0$; however, their derivatives will not be continuous across the shell, but undergo finite discontinuities related to the shell's stress-energy tensor. By requiring that $A^+ + B^+/x^1_0 = 1$ (i.e., that μ be continuous across the shell), we can assure that $x^2_+ = x^2_-$ for points on the shell; and it is similarly possible to identify the values of the other interior and exterior coordinates at each point on the shell. Note that continuity of ψ and γ across the shell requires that $\psi^+ = \gamma^+ = 0$ on the shell.

The form of exterior vacuum metric is well known[4] for $\mu = 0$. Transforming it to the form (1.1), with arbitrary A^+ and B^+, gives

$$\psi^+ = \ln\left[\left(\frac{x^1_+}{x^1_0}\right)\left\{A^+\left(\frac{x^1_+}{x^1_0}\right) + \frac{B^+}{x^1_0}\right\}^{a-1}\right],$$
$$\gamma^+ = \ln\left[\left(\frac{x^1_+}{x^1_0}\right)\left\{A^+\left(\frac{x^1_+}{x^1_0}\right) + \frac{B^+}{x^1_0}\right\}^{a^2-1}\right], \quad (1.4)$$
$$\mu^+ = \ln\left[A^+ + \left(\frac{B^+}{x^1_0}\right)\left(\frac{x^1_0}{x^1_+}\right)\right].$$

(Numerical constants in ψ^+ and γ^+ have been chosen to assure continuity of ψ and γ at x^1_0.) The metric (1.4) depends on two independent parameters, which may be taken as a and A^+. (B^+/x^1_0) is then equal to $1 - A^+$. x^1_0 itself is a scaling parameter: rescaling all the exterior coordinates except x^3 by a factor $1/x^1_0$ multiples the entire line element by a constant conformal factor. It is important to note that this rescaling does not affect x^3. As will be seen shortly, the range of this angular coordinate is related to an important global property of the exterior gravitational field. We may use the coordinates x^0, x^2, x^3 on the surface $x^1 = x^1_0$ as intrinsic coordinates for the shell (plus or minus signs are superfluous here, since both sets of coordinates have the same values at each point on this surface). The three-dimensional line element of the shell is given by

$$d\sigma^2 = (dx^0)^2 - (x^1_0)^2(dx^3)^2 - (dx^2)^2. \quad (1.5)$$

If we denote the components of the second fundamental forms of the surface on each side by $K^{\pm}_{ab}$ in this coordinate system ($a,b = 0, 2, 3$), then the surface stress-energy tensor S_{ab} is given by[2]

$$S_{ab} = (K^+_{ab} - g_{ab}K^+) - (K^-_{ab} - g_{ab}K^-). \quad (1.6)$$

(Plus or minus signs are not needed for the metric tensor g_{ab} of the shell, since the components are the same.) A short computation (see the Appendix) shows that the only nonvanishing components of S^b_a are

$$\sigma = S^0_0 = B^+/(x^1_0)^2, \quad (1.7a)$$
$$p_2 = -S^2_2 = [(a^2 - 2a)(1 - B^+/x^1_0)/x^1_0] - B^+/(x^1_0)^2, \quad (1.7b)$$
$$p_3 = -S^3_3 = a^2(1 - B^+/x^1_0)/x^1_0. \quad (1.7c)$$

Here σ is the surface energy density, and p_2 and p_3 are the partial pressures along the axis of the cylinder and tangential to it, respectively. In order to discuss the behavior of this stress-energy tensor most simply, it is convenient to set $x^1_0 = 1$. (As mentioned above, this results in no loss of generality, but rather to a rescaling of the x^0, x^1, x^2 coordinates). In order to have a positive mass density, B^+ must be greater than 0, and in order to keep A^+ positive, B^+ (really B^+/x^1_0) must remain less than 1. I shall therefore restrict discussion for the present to values of B^+ between 0 and 1. For a fixed value of B^+ in this range, p_2 and p_3 may be considered as functions of a. Both are given by parabolas, with p_3 remaining positive for all values of $a \neq 0$: the shell is always subject to compression tangentially. The longitudinal p_2 is positive for $a > 1 + \sqrt{1/(1 - B^+)}$ and $a < 1 - \sqrt{1/(1 - B^+)}$, turning negative (tension) for values of a between these limits. The dominant energy condition[5] ($\sigma \geqslant 0$, $-\sigma \leqslant p_2, p_3 \leqslant \sigma$) requires that a be negative and greater than $-\sqrt{B^+/(1 - B^+)}$. (Both positive and negative values of p_2 are possible within this range.) It is gratifying that the dominant energy condition requires a to be negative, since a study of the geodesic equation of a test particle outside the shell shows that it is attracted by shells with negative a, and repelled by shells with positive a.

In a previous discussion of this problem,[4] I neglected the role of the second parameter characterizing the exterior gravitational field, which may be taken as either B^+ or A^+ (related by $A^+ = 1 - B^+$, for $x^1_0 = 1$). While a characterizes the local behavior of a test particle, A^+ is essential for understanding its global behavior. The case $a = 0, A^+ = 1$ corresponds to Minkowski space-time, as noted earlier,[4,6] and is not compatible with the existence of a shell of matter. The case $a = 0, A^+ \neq 1$, however, is compatible with certain shells: p_2 must equal $-\sigma$ and p_3 must vanish. The exterior space-time is still locally flat (vanishing Riemann tensor), but there is an effective global gravitational field, as Marder has shown.[7]

Substitution of $a = 0$ into (1.4) gives the exterior line element:

$$ds^2_+ = (dx^0_+)^2 - (dx^1_+)^2 - (A^+x^1_+ + B^+)^2(dx^3_+)^2 - (dx^2_+)^2, \quad (1.8)$$

which is indeed a locally flat space-time, but it corresponds to a globally nontrivial gravitational field. Making the coordinate substitutions

$$\rho = x^1_+ + B^+/A^+, \quad \phi = A^+x^3_+, \quad (1.9)$$

(1.10) may be put into the form

$$ds^2_+ = (dx^0_+)^2 - (d\rho)^2 - \rho^2(d\phi)^2 - (dx^2_+)^2, \quad (1.10)$$

which looks like the ordinary cylindrical form of the flat space-time metric. However, x^3_+ had a range $0 \leqslant x^3_+ < 2\pi$ (because x^3_+ and x^3_- coincide on the surface x^1_0, and x^3_- is the usual angular cylindrical coordinate in the flat interior of the shell). This means that ϕ has a range $0 \leqslant \phi < 2\pi A^+$. For $0 < B^+ < x^1_0$, the range of ϕ will therefore be less than 2π: the exterior metric is a flat conical space-time. A discussion of the null geodesics (light rays) in such a flat conical space-time has been given by Marder.[7] It shows that null geodesics initially diverging from the same point are reconverged due to the conicality if they pass on opposite sides of the cylinder. They may even meet again, if they do not initially diverge too much relative to the degree of conicality, measured by the extent by which A^+ differs from 1. This converging effect is enhanced for timelike geodesics (test particles). Thus, the shell of matter with positive energy density σ acts as an attractive gravitational source in a global sense, even though it fails to produce a locally attractive gravitational field. A shell with negative σ produces a flat "anticonical" space-time ($A^+ > 1$), which can be shown to act as a globally repulsive gravitational source.

Similar results hold if a does not vanish. The exterior metric with $a \neq 0$ takes the form

$$\begin{aligned} ds_+^2 &= [(A^+ x_+^1 + B^+)/x_0^1]^{2a(1-a)}[(dx^0)^2 - (dx^1)^2] \\ &\quad - (A^+ x_+^1 + B^+)^2[(x_0^1/(A^1 x_+^1 + B^+)]^{2a}(dx^3)^2 \\ &\quad - [(A^+ x_+^1 + B^+)/x_0^1)]^{2a}(dx^2)^2. \end{aligned} \tag{1.11}$$

By making the coordinate transformation (1.9), it can be put into a canonical form for an empty cylindrically symmetric space-time[8]:

$$\begin{aligned} ds_+^2 &= (A^+ \rho/x_0^1)^{2a(1-a)}[(dx^0)^2 - (d\rho)^2] \\ &\quad - \rho^2(x_0^1/A^+\rho)^{2a}(d\phi)^2 \\ &\quad - (A^+\rho/x_0^1)^{2a}(dx^2)^2. \end{aligned} \tag{1.12}$$

Thus, (1.12) also represents a conical space-time (the range of ϕ is 0 to $2\pi A^+ < 2\pi$ if B^+ is positive), but one which is nonflat locally if $a \neq 0$. A study of diverging null geodesics (light rays) in this exterior space-time shows that a negative a (locally attractive gravitational field) serves to enhance the convergence effect due to the degree of conicality ($0 < A^+ < 1$), as was to be expected.

These static results are also in accord with the calculations of Fierz[9] and Marder[8] showing that an outgoing pulse of cylindrical gravitational radiation, which is expected to decrease the mass of the source, results in a decrease in the degree of conicality of the exterior field.

Another interesting limiting case is $B^+ = 0$, $a < 0$. The local energy density then vanishes on the shell, and hence there is no conicality in (1.12). The local curvature of space-time is entirely produced by the positive stresses p_2 and p_3.

II. RIGIDLY ROTATING SHELL

As shown in Ref. 1, any globally static space-time on a manifold with nonvanishing first Betti number R_1 can be used to generate an R_1-parameter family of globally stationary but locally static space-times which coincide locally with the initial static space-time. The manifold of the exterior metric ds_+^2 is R^4 minus an infinite four-dimensional cylinder, the surface of which forms the shell of matter. The exterior manifold is thus a manifold with a boundary. Any closed curve encircling the cylindrical boundary cannot be continuously shrunk to a point, while any other closed curve may be so shrunk. The first Betti number of the manifold with boundary is thus 1. So a one-parameter family of globally stationary but locally static space-times can be generated from the exterior metric given by (1.4), (1.1). To exhibit these metrics explicitly, one may perform the formal coordinate transformation

$$\bar{x}^0 = x^0 - cx^3, \quad c = \text{const.} \tag{2.1}$$

Since the range of x^0 is $-\infty < x < +\infty$, while x^3 is limited to the range $0 < x^3 < 2\pi$ (or must be considered a periodic coordinate), this is not a proper global coordinate transformation; therefore the resulting line element is globally distinct from the initial one, even though the coordinate transformation (2.1) makes them locally equivalent. This line element has the form (dropping the bar over x^0)

$$\begin{aligned} ds^2 &= \exp[2(\gamma - \psi)][(dx^0 + c dx^3)^2 - (dx^1)^2] \\ &\quad - (x^1)^2 \exp(-2\psi)(dx^3)^2 - \exp[2(\psi + \mu)](dx^2)^2, \end{aligned} \tag{2.2}$$

with the same range of the coordinates as in (1.1), and ψ, γ, and μ given by (1.4). Since the local transformation (2.1) makes (2.2) coincide with (1.1), and a similar one may be carried out locally on the flat interior metric, the local matching of interior and exterior metrics on the shell of matter at $x^1 = x_0^1$ proceeds in exactly the same way as in Sec. I. The difference is that, with respect to the flat interior region, the shell is now rotating rigidly.

To investigate the motion of the shell with respect to the flat interior space-time, the Minkowski metric written in cylindrical coordinates may be used to describe the interior region:

$$ds_-^2 = dt^2 - d\rho^2 - \rho^2 d\phi^2 - dz^2. \tag{2.3}$$

By carrying out the coordinate transformation $\bar{\phi} = \phi + \omega t$, this may be put into the form (dropping the bar over ϕ)[10]

$$\begin{aligned} ds_-^2 &= [1 - (\omega\rho)^2]dt^2 + 2\omega\rho^2 d\phi dt \\ &\quad - d\rho^2 - \rho^2 d\phi^2 - dz^2. \end{aligned} \tag{2.4}$$

In this form, the curves to which the vector field $v^\mu = \delta_0^\mu$ is tangent are timelike worldlines rotating with the angular velocity with respect to the interior inertial frame. On a surface $\rho = \rho_0$, the metric (2.4) reduces to

$$d\sigma^2 = [1 - (\omega\rho_0)^2]dt^2 + 2\omega\rho_0^2 d\phi\, dt - \rho_0^2 d\phi^2 - dz^2. \tag{2.5}$$

On the surface $x^1 = x_0^1$, (2.2) reduces to

$$\begin{aligned} d\sigma^2 &= (dx^0)^2 + c\, dx^0\, dx^3 \\ &\quad - [(x_0^1)^2 - c^2](dx^3)^2 - (dx^2)^2. \end{aligned} \tag{2.6}$$

Setting

$$\begin{aligned} &[1 - (\omega\rho_0)^2]^{1/2} t = x^0, \quad (\omega\rho_0)^2 t = cx^0, \\ &(\rho_0)^2 = (x_0^1)^2 - c^2, \quad \phi = x^3, \quad z = z^2 \end{aligned} \tag{2.7}$$

makes the two line elements identical. Compatibility between the first two equalities of (2.7) requires that

$$c = \omega(\rho_0)^2/[1 - (\omega\rho_0)^2]^{1/2}. \tag{2.8}$$

Matching the exterior metric to the interior metric, as discussed above, on the surface $x^1 = x_0^1$ is then equivalent to matching it with the interior metric (2.4) on the surface $x^1 = [(\rho_0)^2 + c^2]^{1/2}$. The streamlines of the shell stress-energy tensor, in these coordinates, have $v^\mu = \delta_0^\mu$ as a tangent field. As mentioned above, these represent timelike worldlines rotating with the angular velocity ω with respect to the interior inertial frame. A nonzero value of the external parameter c reflects the fact that the field (2.2) is globally stationary[1]: The one-form or covector $V_\mu = \xi_\mu/\xi^2$, where $\xi^\mu = \delta^\mu{}_0$ is the globally timelike Killing vector of (2.2), has period $\oint V_\mu dx^\mu = 2\pi c$ around any closed curve encircling the cylindrical shell of matter. Equation (2.8) relates this parameter to ω, the rate of rotation of the cylinder as defined with respect to the interior inertial frame. For a fixed value of ρ_0, the radius of the cylinder with respect to this inertial frame, c is a monotonically increasing function of ω for $0 \leqslant \omega < \omega_0 = 1/\rho_0$ (the velocity of points on the shell must be less than the speed of light). As ω varies between zero and ω_0, c varies between zero and infinity.

In Ref. 1 there is a discussion of how a classical optical interference experiment in the exterior region could be used in principle to ascertain the value of **c**.

ACKNOWLEDGMENTS

The author would like to thank the Mathematics Departments of Southampton University and King's College, London, for their hospitality while this work was being completed; and Doctor R. A. D'Inverno and Doctor Felix Pirani for their helpful advice. The financial support of the British Scientific Research Council, which made these visits possible, is also gratefully acknowledged.

APPENDIX

The unit tangent to the surface $x^1 = x_0^1$ is given by $n_\mu = \exp(\gamma - \psi)\delta_\mu^1$. If ξ^a $(a = 0, 2, 3)$ are intrinsic coordinates on this surface, then the coefficients of the second fundamental forms of this surface in these coordinates are given by

$$K_{ab} = n_{\mu;\nu}\frac{\partial x^\mu}{\partial \xi^a}\frac{\partial x^\nu}{\partial \xi^b}. \qquad \text{(A1)}$$

Since we are using x^0, x^2, x^3 as intrinsic coordinates on this surface, and since all metric components depend only on x^1, it follows that

$$K_{ab}^{\pm} = \tfrac{1}{2}g_{ab,1}^{\pm}, \qquad \text{(A2)}$$

where the derivatives on the rhs are evaluated on $x^1 = x_0^1$. Using (1.1), (1.4), and (1.6), Eqs. (1.7) follow.

[1] J. Stachel, "Globally Stationary but Locally Static Space-Times: A Gravitational Analogue of the Aharonov–Bohm Effect," Phys. Rev. D **26**, 1281 (1982).

[2] W. Israel, Nuovo Cimento B **44**, 1 (1966). See also A. Papapetrou and A. Hamoui, Ann. Inst. H. Poincaré A **9**, 179 (1968).

[3] L. Marder, Proc. Roy. Soc. London Ser. A **246**, 133 (1958). Indeed, this may be done as long as $T_0^0 + T_1^1 = 0$, as Marder shows.

[4] See, e.g., J. Stachel, J. Math. Phys. **7**, 1321 (1966).

[5] See, e.g., S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, England, 1973), Sec. 4.3.

[6] So does $a = 1$, but this corresponds to an interchange of the roles of the x^2 and x^3 coordinates. Indeed, $a \leftrightarrow 1 - a$ maps (1.11) into itself with $x^2 \leftrightarrow x^3$. So one should require $a < 1$.

[7] L. Marder, "Locally Isometric Space-Times," in *Recent Developments in General Relativity* (PWN-Polish Scientific, Warsaw, 1962), p. 333.

[8] L. Marder, Proc. Roy. Soc. London Ser. A **244**, 524 (1958).

[9] See J. Weber and J. A. Wheeler, Rev. Mod. Phys. **29**, 509 (1957), p. 512.

[10] W. J. Sarill [Masters thesis, Boston University, 1976 (unpublished)] and J. M. Cohen, W. J. Sarill, and C. V. Vishveshwara ("An Example of Induced Centrifugal Force in General Relativity," preprint) have discussed the induced rotation of inertial frames within a rotating cylindrical shell of matter using this form of metric tensor.

Null strings and Hertz potentials

G. F. Torres del Castillo
Departamento de Física Matemática, Instituto de Ciencias de la Universidad Autónoma de Puebla, Puebla, Pue., Mexico and Departamento de Física, Centro de Investigación y Estudios Avanzados del IPN, Apartado Postal 14-740, 07000 México, D.F., Mexico

(Received 6 May 1983; accepted for publication 19 August 1983)

A covariant expression for the massless spinor fields in terms of Hertz potentials is given, assuming that the space-time admits a congruence of null strings defined by a multiple Debever–Penrose spinor. The metric of these spaces is also given in a covariant form.

PACS numbers: 04.20.Jb

1. INTRODUCTION

Searching for exact solutions of the Einstein field equations, the assumption of the existence of a congruence of null strings (two-dimensional totally null surfaces) allows one to reduce them, in the case of vacuum, to just one differential constraint.[1,2] When there are sources present a similar simplification is obtained when, additionally, one imposes certain conditions on the energy-momentum tensor of the matter.[3,4] As a consequence, the conformal curvature is algebraically degenerate as in the case of vacuum. In order to achieve this reduction, it becomes necessary to deal with complexified space-times, for in real ones (i.e., with Lorentzian signature) there are no such surfaces.

Moreover, the equations for the massless spinor fields (the anti-self-dual ones if the congruence is defined by a self-dual two-form or vice versa) are reduced to a single wavelike equation for a scalar potential in terms of which the solution is expressed[4] when the space-time admits a congruence of null strings and the algebraic degeneracy of the conformal curvature is assumed.

In the case of Einstein's equations, as well as for the massless spinor fields, the integration process did depend on the existence of coordinate systems adapted to the congruence, and the respective solutions were given in terms of null tetrads induced by these preferred coordinates. It is of interest to establish these results in a covariant way in order to find their underlying structure and to apply them to any space-time belonging to the class referred to above without making reference to a particular coordinate system. In this paper a method is given to obtain such covariant expressions. Previously, Plebański and Robinson[5] have obtained a covariant description for the structure of $\mathscr{H}$ spaces ("half-flat" complex space-times) by a different approach to that followed here. The present work complements, in a certain sense, the one of Plebański and Rózga[6] on null strings.

Even though one has to consider complex space-times in the derivation presented here, all the final results apply as well in the case of real space-times. In Sec. 2, some basic facts about congruences of null strings and their relation with null massless fields are established. Some similar results for the case of real space-times can be found in Refs. 7 and 8. In Sec. 3, a covariant expression for the solutions of the massless spinor field equations in terms of Hertz potentials is obtained. In Sec. 4, a covariant description for the metric of any space-time which admits a congruence of null strings defined by a multiple Debever–Penrose spinor is given. The formalism and notation used here follow Ref. 9. *All* the spinorial indices are raised and lowered according to the conventions $\psi^B = \psi_A \epsilon^{AB}$, $\psi_A = \epsilon_{AB}\psi^B$, and similarly for dotted indices.

2. NULL MASSLESS FIELDS

In this section it is shown that in a space-time which admits a congruence of null strings defined by a multiple Debever–Penrose spinor, one can construct null massless fields of arbitrary spin $s \geqslant \frac{1}{2}$ without any restriction on the Ricci tensor.

Let $\partial_{A\dot{B}}$ denote a null tetrad (with $\partial_{A\dot{B}}\partial_{C\dot{D}} = -2\epsilon_{AC}\epsilon_{\dot{B}\dot{D}}$). A nonvanishing locally defined spinor field l_A determines a two-dimensional distribution spanned by the vector fields

$$v_{\dot{B}} \equiv l^A \partial_{A\dot{B}}. \tag{2.1}$$

This distribution is involutive, and hence integrable, if and only if

$$l^A l^B \nabla_{A\dot{C}} l_B = 0. \tag{2.2}$$

As a consequence of (2.2), the covariant derivatives $\nabla_{v_{\dot{A}}} v_{\dot{B}}$ are linear combinations of $v_{\dot{C}}$. Therefore if l_A satisfies the condition (2.2) then the vector fields $v_{\dot{A}}$ are tangent to a congruence of two-dimensional geodesic surfaces. Since $v_{\dot{A}} v_{\dot{B}} = 0$, these surfaces are null and they are called null strings.[10] In a real space-time (assuming that $\overline{\partial_{A\dot{B}}} = \partial_{B\dot{A}}$, where the bar denotes complex conjugation) condition (2.2) means that the vector field $l^A \bar{l}^{\dot{B}} \partial_{A\dot{B}}$ (where $\bar{l}^{\dot{B}} = \overline{l^B}$) is tangent to a congruence of shear-free null geodesics.

Let ψ be a nonvanishing function. The spinor field ψl_A satisfies Eq. (2.2) if and only if l_A does; then both define the same congruence of null strings. Based on this ambiguity, one can impose on the spinor field l_A the further condition[6]

$$\nabla_{A\dot{C}}(l^A l^B) = 0. \tag{2.3}$$

Indeed, in a spinor frame such that $l_A = \delta_A^2$, Eq. (2.2) is equivalent to

$$\Gamma_{11}(\partial_{1\dot{C}}) = 0, \tag{2.4}$$

where Γ_{AB} and $\Gamma_{\dot{A}\dot{B}}$ denote the connection one-forms for the tetrad $\partial_{A\dot{B}}$, hence, $\nabla_{A\dot{C}}(\psi l^A \psi l^B) = \psi^2 \delta_1^B [\partial_{1\dot{C}} \ln \psi^2$

$+\Gamma_{11}(\partial_{2\dot C})-2\Gamma_{12}(\partial_{1\dot C})]$. Therefore, $\nabla_{A\dot C}(\psi l^A\psi l^B)=0$ if and only if ψ satisfies the conditions

$$\partial_{1\dot A}\ln\psi^2=2\Gamma_{12}(\partial_{1\dot A})-\Gamma_{11}(\partial_{2\dot A}). \tag{2.5}$$

From the relation $[X,Y]=\nabla_X Y-\nabla_Y X$, the second structural equations, and (2.4) one finds that

$$\partial_1{}^{\dot A}\partial_{1\dot A}=-[\Gamma_{12}(\partial_1{}^{\dot A})+\Gamma^{\dot A\dot B}(\partial_{1\dot B})]\partial_{1\dot A}, \tag{2.6}$$

and, denoting by C_{ABCD} and $C_{\dot A\dot B\dot C\dot D}$ the components of the Weyl spinor,

$$\begin{aligned}
&C_{1111}=0,\\
&-4C_{1112}=\partial_1{}^{\dot A}(\Gamma_{11}(\partial_{2\dot A}))+[\Gamma_{12}(\partial_1{}^{\dot A})\\
&\qquad+\Gamma^{\dot A\dot B}(\partial_{1\dot B})]\Gamma_{11}(\partial_{2\dot A}),\\
&-2C_{1112}=\partial_1{}^{\dot A}(\Gamma_{12}(\partial_{1\dot A}))+\Gamma^{\dot A\dot B}(\partial_{1\dot B})\Gamma_{12}(\partial_{1\dot A}).
\end{aligned} \tag{2.7}$$

Thus, applying $\partial_1{}^{\dot A}$ to both sides of (2.5) one obtains an identity, showing the integrability of those equations.

Conversely, contracting (2.3) with l_B one finds that the spinor field l_A satisfies Eq. (2.2). In other words, there exists a function f such that the spinor field $fl^A l^B$ satisfies the massless field equations of spin 1 ($\nabla_{A\dot C} fl^A l^B=0$) if and only if l_A defines a congruence of null strings. In a real space-time this result is the spinorial image of a well-known theorem of Robinson.[11]

Defining the two-form $\Sigma\equiv l_A\, g^{A\dot 1}\wedge l_B\, g^{B\dot 2}$, one finds that the integrability condition (2.2) is equivalent to the existence of a one-form α such that $d\Sigma=\alpha\wedge\Sigma$, while Eq. (2.3) is equivalent to $d\Sigma=0$. The tangent vectors v satisfying $v\,\lrcorner\,\Sigma=0$ are those which are tangent to the null strings defined by l_A.

On the other hand, if there exist a function f and a spinor field l_A such that $\nabla_{A_1\dot R}(fl^{A_1}l^{A_2}\cdots l^{A_{2s}})=0$, then, assuming $s\geqslant 1$ and taking $l_A=\delta_A^2$ as before, one finds that l_A satisfies Eq. (2.2) and that the function f must be a solution of

$$\partial_{1\dot A}\ln f=2s\Gamma_{12}(\partial_{1\dot A})-\Gamma_{11}(\partial_{2\dot A}). \tag{2.8}$$

Applying $\partial_1^{\dot A}$ to both sides of this equation and using (2.6) and (2.7) it follows that

$$(s-1)C_{1112}=0. \tag{2.9}$$

Thus, for $s>1$, a necessary condition for the existence of massless fields of the form $fl^{A_1}\cdots l^{A_{2s}}$ is that the spinor field l_A must be a solution of (2.2) and at the same time, a multiple Debever–Penrose (DP) spinor (i.e., $l^A l^B l^C C_{ABCD}=0$). An alternative derivation of this result can be found in the more general discussion given in Ref. 8.

The converse is also true. In fact, given a spinor field l_A which satisfies (2.2) and is a multiple DP spinor there exists a function ϕ such that,[4]

$$l^B\nabla_{A\dot C}l_B=l_A l^B\partial_{B\dot C}\ln\phi. \tag{2.10}$$

Then, assuming that (2.3) holds, it follows that

$$\begin{aligned}
\nabla_{B\dot C}l^B&=-\tfrac12 l^B\partial_{B\dot C}\ln\phi,\\
l^B\nabla_{B\dot C}l_A&=\tfrac12 l_A l^B\partial_{B\dot C}\ln\phi.
\end{aligned} \tag{2.11}$$

Therefore, applying repeatedly (2.11) one finds[12]

$$\nabla_{A_1\dot R}(\phi^{1-s}l^{A_1}l^{A_2}\cdots l^{A_{2s}})=0. \tag{2.12}$$

By multiplying $\phi^{1-s}l^{A_1}\cdots l^{A_{2s}}$ by a function f, one gets another solution of (2.12) if and only if f satisfies $l^A\partial_{A\dot B}f=0$. This last condition means that f is a constant on each null string or, equivalently, that f is an arbitrary function of $q^{\dot 1}$ and $q^{\dot 2}$ (see Sec. 3).[13]

3. MASSLESS FIELDS AND HERTZ POTENTIALS

In the rest of this paper it will be assumed that the spinor field l_A is a multiple DP spinor which defines a congruence of null strings. As is shown in Ref. 4 the integrability condition (2.2) implies the existence of a pair of functions $q^{\dot 1}$, $q^{\dot 2}$ such that

$$l_A\, g^{A\dot B}=\sqrt2\, l^{\dot B}{}_{\dot C}\,dq^{\dot C}, \tag{3.1}$$

where $(l^{\dot B}{}_{\dot C})$ is a nonsingular matrix. Defining the functions $p^{\dot A}$ by the conditions

$$l^C\partial_{C\dot A}\,p^{\dot B}=-\sqrt2\phi^2 l_{\dot A}{}^{\dot B}, \tag{3.2}$$

where ϕ is a solution of (2.10), it follows that the functions $q^{\dot A}$ and $p^{\dot A}$ can be used as local coordinates and the set of vector fields

$$\begin{aligned}
\partial'_{1\dot A}&=\sqrt2\frac{\partial}{\partial p^{\dot A}}\equiv\sqrt2\partial_{\dot A},\\
\partial'_{2\dot A}&=\sqrt2\phi^2\left(\frac{\partial}{\partial q^{\dot A}}-Q^{\dot B}{}_{\dot A}\frac{\partial}{\partial p^{\dot B}}\right)\equiv\sqrt2\phi^2 D_{\dot A},
\end{aligned} \tag{3.3}$$

constitutes a null tetrad with

$$Q^{\dot A\dot B}=(1/\sqrt2)\,l^{\dot D(\dot A}m^{|C|}\partial_{C\dot D}\,p^{\dot B)}, \tag{3.4}$$

where m^A is a spinor field such that $m^A l_A=1$.

With respect to the tetrad $\partial'_{A\dot B}$, induced by the coordinates $q^{\dot A}$, $p^{\dot A}$, the most general solution of the ("right") massless field equations, $\nabla_R{}^{\dot A_1}\Psi_{\dot A_1\ldots\dot A_{2s}}=0$, can be written in terms of a scalar potential H' in the form

$$\Psi'_{\dot A_1\dot A_2\ldots\dot A_{2s}}=\phi^{s+1}\partial'_{1\dot A_1}\partial'_{1\dot A_2}\cdots\partial'_{1\dot A_{2s}}H', \tag{3.5}$$

where H' satisfies a wavelike equation[14] (see Ref. 4). In order to express the components of the field (3.5) with respect to an arbitrary null tetrad $\partial_{A\dot B}$, it is necessary to replace the derivatives $\partial'_{1\dot A}$ by covariant derivatives. To this end, one begins by noticing that the null tetrads $\partial'_{A\dot B}$ and $\partial_{A\dot B}$ are related by

$$\partial'_{A\dot B}=M_A{}^C M_{\dot B}{}^{\dot D}\partial_{C\dot D}, \tag{3.6}$$

where

$$\begin{aligned}
&M_{\dot B}{}^{\dot D}=\Delta^{-1/2}l^{\dot D}{}_{\dot B}=-\Delta^{1/2}l^{-1\dot D}_{\ \dot B},\\
&M_1{}^C=\Delta^{-1/2}\phi^{-2}l^C,\\
&M_2{}^C=-\Delta^{1/2}\phi^2(m^C+(l^C/2\sqrt2)\phi^{-2}\\
&\qquad\times l^{-1\dot B}_{\ \dot A}m^D\partial_{D\dot B}\,p^{\dot A})\\
&\qquad\equiv-\Delta^{1/2}\phi^2\tilde m^C,\\
&\Delta\equiv\det l^{\dot A}{}_{\dot B}.
\end{aligned} \tag{3.7}$$

Denoting by $\Gamma_{AB},\Gamma_{\dot A\dot B}$ and $\Gamma'_{AB},\Gamma'_{\dot A\dot B}$ the connection one-forms for the tetrad $\partial_{A\dot B}$ and $\partial'_{A\dot B}$, respectively, from (3.7) one gets

$$\Gamma^{\dot A}{}_{\dot B}=l_{\dot B}{}^{\dot D}l^{-1\dot A}_{\ \dot C}\Gamma'^{\dot C}{}_{\dot D}-\Delta^{-1/2}l_{\dot B}{}^{\dot C}d(\Delta^{1/2}l^{-1\ \dot A}_{\ \dot C}), \tag{3.8}$$

therefore,

$$dl^{-1\dot{B}}_{\dot{C}} = -l^{-1\dot{D}}_{\dot{C}}\Gamma^{\dot{B}}{}_{\dot{D}} + l^{-1\dot{B}}_{\dot{D}}\Gamma'^{\dot{D}}{}_{\dot{C}} - \tfrac{1}{2}l^{-1\dot{D}}_{\dot{C}}d\ln\Delta. \tag{3.9}$$

According to (3.1), the two-form Σ defined in the previous section is given by

$$\Sigma = \Delta dq^{A}\wedge dq_{\dot{A}}, \tag{3.10}$$

hence, $d\Sigma = d\ln\Delta\wedge\Sigma$. Thus, if l_A satisfies Eq. (2.3), then Δ is a function of q^A only and conversely. In the forthcoming it will be assumed that Eq. (2.3) holds, hence $v_{\dot{S}}\Delta = 0$ [see Eq. (2.1)].

Using the explicit expression for the one-forms $\Gamma'_{\dot{A}\dot{B}}$ (see Refs. 2 and 4), Eq. (3.9) yields

$$v_{\dot{S}}l^{-1\dot{B}}_{\dot{C}} = -l^{-1\dot{D}}_{\dot{C}}\Gamma^{\dot{B}}{}_{\dot{D}}(v_{\dot{S}}) + \tfrac{1}{2}(\delta^{\dot{B}}_{\dot{D}}l^{-1}_{\dot{C}\dot{S}} + \delta^{\dot{B}}_{\dot{S}}l^{-1}_{\dot{C}\dot{D}})v^{\dot{D}}\ln\phi. \tag{3.11}$$

[This equation can also be obtained from the condition, $0=\sqrt{2}d\,dq^{\dot{C}} = d(l^{-1\dot{C}}{}_{\dot{B}}l_A g^{A\dot{B}})$, using Eq. (2.11).] Therefore, if $\Phi_{\dot{B}_1\ldots\dot{B}_n}$ denote the components of a spinorial field (not necessarily symmetric) referred to the tetrad $\partial_{A\dot{B}}$, then by Eq. (3.11) one gets

$$\begin{aligned} v_{\dot{S}}(l^{-1\dot{B}_1}_{\dot{A}_1}\cdots l^{-1\dot{B}_n}_{\dot{A}_n}\Phi_{\dot{B}_1\ldots\dot{B}_n}) &= l^{-1\dot{B}_1}_{\dot{A}_1}\cdots l^{-1\dot{B}_n}_{\dot{A}_n}\{\phi^{n/2}\nabla_{\dot{S}}(\phi^{-n/2}\Phi_{\dot{B}_1\ldots\dot{B}_n}) \\ &\quad + (\epsilon_{\dot{S}\dot{B}_1}\Phi_{\dot{D}\dot{B}_2\ldots\dot{B}_n} + \cdots + \epsilon_{\dot{S}\dot{B}_n}\Phi_{\dot{B}_1\ldots\dot{D}})v^{\dot{D}}\ln\phi\}, \end{aligned} \tag{3.12}$$

where

$$\nabla_{\dot{S}} \equiv l^{C}\nabla_{C\dot{S}}. \tag{3.13}$$

Returning to Eq. (3.5), from Eqs. (3.6) and (3.7), using repeatedly (3.12) one finds

$$\begin{aligned} \Psi'_{\dot{A}_1\cdots\dot{A}_{2s}} &= (-1)^{2s}\phi^{2s+1}l^{-1\dot{B}_1}_{(\dot{A}_1}\cdots l^{-1\dot{B}_{2s}}_{\dot{A}_{2s})} \\ &\quad\times(\phi^{-5/2}\nabla_{\dot{B}_1})\cdots(\phi^{-5/2}\nabla_{\dot{B}_{2s}})H'. \end{aligned}$$

Thus, since $\nabla_{\dot{A}}(\phi^{-1/2}l_B)=0$ [see Eq. (2.11)], from (3.7) it follows that the components of the field with respect to the tetrad $\partial_{A\dot{B}}$ are given by

$$\begin{aligned} \Psi_{\dot{A}_1\ldots\dot{A}_{2s}} &= \Delta^{-s}\phi^{s+1}l^{B_2}l^{B_2}\cdots l^{B_{2s}}\phi^{-2} \\ &\quad\times\nabla_{B_1(\dot{A}_1}\phi^{-2}\nabla_{|B_2|\dot{A}_2\ldots}\phi^{-2}\nabla_{|B_2|\dot{A}_{2s})}H' \\ &= \phi^{s+1}l^{B_1}\cdots l^{B_{2s}}\phi^{-2}\nabla_{B_1(\dot{A}_1\ldots}\phi^{-2}\nabla_{|B_{2s}|\dot{A}_{2s})}H, \end{aligned} \tag{3.14}$$

where

$$H \equiv \Delta^{-s}H'. \tag{3.15}$$

Due to (2.12), the expression (3.14) can be written in terms of a $D(s,0)$ null Hertz potential

$$\begin{aligned} \Psi_{\dot{A}_1\ldots\dot{A}_{2s}} &= \phi^{2(s-1)}\nabla_{B_1(\dot{A}_1}\phi^{-2}\nabla_{|B_2|\dot{A}_2} \\ &\quad\cdots\phi^{-2}\nabla_{|B_{2s}|\dot{A}_{2s})}H\phi^{1-s}l^{B_1}l^{B_2}\cdots l^{B_{2s}}. \end{aligned} \tag{3.16}$$

Evidently, by redefining H, the factor ϕ^{1-s} which appears in the Hertz potential can be absorbed. Moreover, since any nonvanishing factor can be absorbed into H, Eq. (3.16) holds even when the spinor l_A does not satisfy condition (2.3). For example, in a spinor frame such that $l_A = \delta^2_A$, using (2.10) one finds that Eq. (3.16) reduces to the expressions for the components of the massless spinor fields found by Cohen and Kegeles.[15] The relation between the Cohen and Kegeles potential ψ and the potential H defined above is given by $\psi = \phi^{1-3s}H$.

Substituting the expression (3.16) into the massless field equations, $\nabla^{R\dot{A}_1}\Psi_{\dot{A}_1\ldots\dot{A}_{2s}}=0$, one obtains a differential condition that the Hertz potential has to satisfy. However, the equations $l_R\nabla^{R\dot{A}_1}\Psi_{\dot{A}_1\ldots\dot{A}_{2s}}=0$ are automatically fulfilled by (3.16). Hence, the field equations are reduced to $m_R\nabla^{R\dot{A}_1}\Psi_{\dot{A}_1\ldots\dot{A}_{2s}}=0$, where m_R is any spinor field such that $m^R l_R$ does not vanish. Thus, in the case $s=\frac{1}{2}$, the Hertz potential has to satisfy a single linear partial differential equation of second order.[15] In the case $s=1$, by using the ambiguity in the definition of H, one can further reduce the field equations to one linear partial differential equation of second order.[3,15]

For $s>1$, there exists integrability conditions which restrict the solution of the field equations. These conditions involve the conformal curvature and are expressed by[16]

$$C_{\dot{A}_1\dot{A}_2\dot{A}_3(\dot{B}_1}\Psi^{\dot{A}_1\dot{A}_2\dot{A}_3}{}_{\dot{B}_4\ldots\dot{B}_{2s})} = 0. \tag{3.17}$$

A similar condition holds for the left massless fields. These algebraic restrictions imply that the existence of null massless fields require the algebraic degeneracy of the conformal curvature (cf., Sec. 2).

If the integrability conditions (3.17) are fulfilled, then the potential H is subject to a single second-order differential constraint. In the coordinates $q^A, p^{\dot{A}}$, this condition is obtained by commuting the derivatives $\partial_{\dot{A}_1},\ldots,\partial_{\dot{A}_{2s-1}}$ and $D^{\dot{B}}$ in Eq. (4.3a) of Ref. 4.

By using an expression analogous to (3.16), Penrose[17] showed that in the case of flat space-time, the massless fields satisfy the peeling theorem, assuming that the scalar potential (analogous to H) has an appropriate asymptotic behavior.

4. THE COVARIANT FORM OF THE METRIC

Now, the metric determined by (3.3) will be written in a covariant way using a similar procedure to the one of Sec. 3.

From Eqs. (3.6), (3.7), and (3.3) one gets

$$\begin{aligned} \partial_{A\dot{B}} &= M^{C}{}_{A}M^{\dot{D}}{}_{\dot{B}}\partial'_{C\dot{D}} \\ &= \sqrt{2}\Big[\phi^2 l_B{}^{\dot{D}}\tilde{m}_A\partial_{\dot{D}} - l^{-1\dot{D}}{}_{\dot{B}}l_A\frac{\partial}{\partial q^{\dot{D}}}\Big] \\ &\quad - \sqrt{2}l_A l^{-1\dot{D}}{}_{\dot{B}}Q_{\dot{D}\dot{C}}\partial^{\dot{C}}. \end{aligned} \tag{4.1}$$

The tetrad

$$^0\partial_{A\dot{B}} \equiv \sqrt{2}\Big[\phi^2 l_{\dot{B}}{}^{\dot{D}}\tilde{m}_A\partial_{\dot{D}} - l^{-1\dot{D}}{}_{\dot{B}}l_A\frac{\partial}{\partial q^{\dot{D}}}\Big] \tag{4.2}$$

corresponds to a conformally flat metric (specifically, $^0\partial_{A\dot{B}}$ is a null tetrad with respect to the conformally flat metric $^0g = 2\phi^{-2}dq^A\otimes dp_{\dot{A}}$), while the remaining term in (4.1) can be written as $\phi^{-2}l_A\Omega_{\dot{B}}{}^{\dot{R}}l^S\partial_{S\dot{R}}$, where

$$\Omega_{\dot{A}\dot{B}} \equiv l^{-1\dot{C}}{}_{\dot{A}}l^{-1\dot{D}}{}_{\dot{B}}Q_{\dot{C}\dot{D}}. \tag{4.3}$$

Therefore,

$$\partial_{A\dot{B}} = {}^0\partial_{A\dot{B}} + \phi^{-2}l_A\Omega_{\dot{B}\dot{R}}l_S\partial^{S\dot{R}}. \tag{4.4}$$

The connection one-forms for the tetrad $\partial'_{A\dot{B}}$ (see Refs. 2 and 4) can be decomposed into a conformally flat part,

${}^0\Gamma'_{AB}$ and ${}^0\Gamma'_{\dot A\dot B}$ (which is obtained by setting $Q_{\dot A\dot B}=0$) plus $Q_{\dot A\dot B}$- dependent part

$$
\begin{aligned}
\Gamma'_{11} &= {}^0\Gamma'_{11},\\
\Gamma'_{12} &= {}^0\Gamma'_{12} + \tfrac{1}{2}\phi^2\partial^{\dot A}(\phi^{-2}Q_{\dot A\dot B})dq^{\dot B},\\
\Gamma'_{22} &= {}^0\Gamma'_{22} + \phi^4 D^{\dot A}(\phi^{-2}Q_{\dot A\dot B})dq^{\dot B}\\
&\quad + \phi Q^{\dot A\dot B}(D_{\dot A}\phi\, dq_{\dot B} + \partial_{\dot A}\phi\, dp_{\dot B}),\\
\Gamma'_{\dot A\dot B} &= {}^0\Gamma'_{\dot A\dot B} + \phi^2\partial_{(\dot A}\phi^{-2}Q_{\dot B)\dot C}dq^{\dot C}\\
&\quad + \phi^{-1}Q_{\dot A\dot B}\partial_{\dot C}\phi\, dq^{\dot C}.
\end{aligned}
\tag{4.5}
$$

Hence, the connection one-forms for the tetrad $\partial_{A\dot B}$ are given by

$$
\begin{aligned}
\Gamma_{AB} &= M^C{}_A M^D{}_B \Gamma'_{CD} - M_{CA}\, dM^C{}_B\\
&= {}^0\Gamma_{AB} - M_{1(A}M_{|2|B)}\phi^2\partial^{\dot C}(\phi^{-2}Q_{\dot C\dot D})dq^{\dot D}\\
&\quad + M_{1A}M_{1B}[\phi^4 D^{\dot C}(\phi^{-2}Q_{\dot C\dot D})dq^{\dot D}\\
&\quad + \phi Q^{\dot C\dot D}(D_{\dot C}\phi\, dq_{\dot D} + \partial_{\dot C}\phi\, dp_{\dot D})],
\end{aligned}
\tag{4.6}
$$

where

$$
{}^\circ\Gamma_{AB} \equiv M^C{}_A M^D{}_B {}^0\Gamma'_{CD} - M_{CA}\, dM^C{}_B. \tag{4.7}
$$

And, similarly,

$$
\begin{aligned}
\Gamma_{\dot A\dot B} &= -l_{\dot B}{}^{\dot D} l^{-1\dot C}{}_{\dot A}\Gamma'_{\dot C\dot D} + l^{-1}{}_{\dot C(\dot A}dl_{\dot B)}{}^{\dot C}\\
&= {}^0\Gamma_{\dot A\dot B} - l_{\dot B}{}^{\dot D} l^{-1\dot C}{}_{\dot A}[\phi^2\partial_{(\dot C}\phi^{-2}Q_{\dot D)\dot R}\\
&\quad + Q_{\dot C\dot D}\partial_{\dot R}\ln\phi\,]dq^{\dot R},
\end{aligned}
\tag{4.8}
$$

where

$$
{}^0\Gamma_{\dot A\dot B} \equiv l_{\dot B}{}^{\dot D} l^{-1\dot C\circ}{}_{\dot A}\Gamma'_{\dot C\dot D} + l^{-1}{}_{\dot C(\dot A}dl_{\dot B)}{}^{\dot C}. \tag{4.9}
$$

The one-forms ${}^0\Gamma_{AB}$ and ${}^0\Gamma_{\dot A\dot B}$ are the connection forms for the tetrad ${}^0\partial_{A\dot B}$.

Since $l^{\dot A}{}_{\dot C} = -\Delta l^{-1\dot A}{}_{\dot C}$, by Eqs. (4.8), (4.1), (4.3), (3.12), (3.1), and (2.3) it follows that

$$
\Gamma_{\dot A\dot B} = {}^0\Gamma_{\dot A\dot B} + \tfrac{1}{2}\nabla_{D(\dot A}(\phi^{-2}l^D\Omega_{\dot B)\dot S}l_C)g^{C\dot S}. \tag{4.10}
$$

On the other side, in order to express the derivatives appearing in (4.6) in a covariant form, from Eqs. (3.8) and (3.12), using the explicit expression of $\Gamma'_{\dot A\dot B}$, one finds

$$
\begin{aligned}
&\partial_{A\dot S}l^{\dot B}{}_{\dot C}\\
&= -l^{\dot D}{}_{\dot C}\Gamma^{\dot B}{}_{\dot D}(\partial_{A\dot S}) + \tfrac{1}{2}l^{\dot B}{}_{\dot C}\partial_{A\dot S}\ln\Delta + \tfrac{1}{2}l_{\dot S\dot C}\partial_A{}^{\dot B}\ln\phi\\
&\quad - \tfrac{1}{2}\delta^{\dot B}_{\dot S}l^{\dot D}{}_{\dot C}\partial_{A\dot D}\ln\phi + \phi^{-2}l_A l_{\dot D\dot C}l^R\\
&\quad \times[\nabla_R{}^{(\dot D}\Omega^{\dot B)}{}_{\dot S} - \Omega^{\dot B\dot D}\partial_{R\dot S}\ln\phi\,].
\end{aligned}
\tag{4.11}
$$

However, in (4.11) besides $\phi,\Omega_{\dot A\dot B}$ and l_A there appear the derivatives of the function Δ, which depend on the choice of the coordinates $q^{\dot A}$ made at the start. One can eliminate these terms by noticing that from Eq. (4.6) one has

$$
dM_{CB} = -M_C{}^A\Gamma_{AB} - M^D{}_B\Gamma'_{CD};
$$

hence, substituting (3.7) it follows that

$$
\begin{aligned}
\nabla(\Delta^{-1/2}\phi^{-2}l_B) &= -M^D\Gamma_{B1D} = -\Delta^{1/2}\phi^2\tilde m_B\Gamma'_{11}\\
&\quad - \Delta^{-1/2}\phi^{-2}l_B\Gamma'_{12},\\
\nabla(\Delta^{1/2}\phi^2\tilde m_B) &= M^D{}_B\Gamma'_{2D} = \Delta^{1/2}\phi^2\tilde m_B\Gamma'_{12}\\
&\quad + \Delta^{-1/2}\phi^{-2}l_B\Gamma'_{22},
\end{aligned}
\tag{4.12}
$$

with the action of the covariant differential ∇ defined by $\nabla\psi_A = -\tfrac{1}{2}(\nabla_{B\dot C}\psi_A)g^{B\dot C}$. Then, using the explicit form of Γ'_{AB} and Eqs. (3.6), (3.7), (4.3), and (3.12) one gets

$$
\begin{aligned}
\nabla_{A\dot C}l_B &= \tfrac{1}{2}l_B[3\partial_{A\dot C}\ln\phi + \partial_{A\dot C}\ln\Delta\\
&\quad + \phi^{-4}l_A l^D\nabla_D{}^{\dot R}(\phi^2\Omega_{\dot C\dot R})]\\
&\quad + \epsilon_{AB}l^D\partial_{D\dot C}\ln\phi
\end{aligned}
\tag{4.13}
$$

from which it follows that

$$
\begin{aligned}
\nabla_{A\dot C}l_B l_S &= l_B l_S\partial_{A\dot C}\ln\Delta + 3l_{(B}l_S\partial_{A)\dot C}\ln\phi\\
&\quad + \phi^{-4}l_B l_S l_A l^D\nabla_D{}^{\dot R}(\phi^2\Omega_{\dot C\dot R}).
\end{aligned}
\tag{4.14}
$$

Finally, substituting into (4.6) after some work one finds that

$$
\begin{aligned}
\Gamma_{AB} &= {}^0\Gamma_{AB} - \tfrac{1}{2}\nabla_{(A}{}^{\dot C}(\phi^{-2}l_{B)}\Omega_{\dot C\dot D}l_S)g^{S\dot D}\\
&\quad + \tfrac{1}{2}\phi^{-2}l_A l_B\Omega^{\dot R\dot C}[\nabla_{S\dot R}(\phi^{-2}l^S\Omega_{\dot C\dot D}l_P)\\
&\quad - \tfrac{3}{2}\phi^{-2}l_P\Omega_{\dot R\dot C}l^S\partial_{S\dot D}\ln\phi\,]g^{P\dot D}.
\end{aligned}
\tag{4.15}
$$

The components of the right conformal curvature are easily obtained from their expression referred to the tetrad $\partial'_{A\dot B}$.[2,4] From Eqs. (3.6), (4.3), (3.12), and (2.3) one has

$$
C_{\dot A\dot B\dot C\dot D} = -\tfrac{1}{2}\phi\nabla_{R(\dot A}\phi^{-2}\nabla_{|S|\dot B}\phi^{-1}l^R l^S\Omega_{\dot C\dot D)}. \tag{4.16}
$$

The components of the left conformal curvature referred to as an arbitrary tetrad can be obtained by a similar computation. However, this computation is considerably more involved than the previous ones. It requires knowledge of the covariant derivatives of $\tilde m_A$, which are given by (4.12). The result, written in a form which explicitly shows the algebraic degeneracy of this spinor, is

$$
\begin{aligned}
C_{ABCD} &= -\tfrac{1}{2}\Delta^{-1}\phi^{-2}l_{(A}l_B\nabla_C{}^{\dot R}\phi^{-2}\nabla_{D)}{}^{\dot S}\Delta\phi^2\Omega_{\dot R\dot S}\\
&\quad + \tfrac{1}{4}\phi^{-6}l_A l_B l_C l_D\\
&\quad \times[\tfrac{1}{2}\Delta^{-2}\nabla^{\dot R\dot S}\phi^2\partial_{\dot R\dot S}(\Delta^2\Omega^{\dot P\dot Q}\Omega_{\dot P\dot Q})\\
&\quad + \phi^{-2}l^M\nabla_M{}^{\dot R}(\phi^2\Omega_{\dot R}{}^{\dot S})l^N\partial_{N\dot S}(\Omega^{\dot P\dot Q}\Omega_{\dot P\dot Q})].
\end{aligned}
$$

The derivatives of the function Δ can be eliminated by using Eq. (4.14).

If one assumes that the Einstein field equations are satisfied, then $Q_{\dot A\dot B}$ can be written in terms of a key function and some *constants of integration* provided that the energy-momentum tensor of the matter has a constant trace and satisfies $l^A l^B T_{AB\dot C\dot D}=0$, where $T_{AB\dot C\dot D}$ denotes the spinorial components of the traceless part of the energy-momentum tensor.[3,4] Clearly, this condition is satisfied in the case of vacuum.[1,2] Using the procedure presented here one can find $\Omega_{\dot A\dot B}$ and then apply the general relations derived above.

For example, in the case of vacuum with $l^B\nabla_{A\dot C}l_B=0$ (called Case I in Refs. 1 and 2) ϕ is a function of $q^{\dot R}$ only and the object $Q_{\dot A\dot B}$ is given by

$$
Q_{\dot A\dot B} = -\partial_{\dot A}\partial_{\dot B}\Theta' - \tfrac{2}{3}\phi^2 L_{(\dot A}p_{\dot B)},
$$

where Θ' is a function which has to fulfill a partial differential equation of second order involving only quadratic nonlinearities and $L_{\dot A}=L_{\dot A}(q^{\dot R})$. Then from Eqs. (4.3), (3.6), (3.7), (3.12), and (2.3) it follows that

$$
\Omega_{\dot A\dot B} = -\tfrac{1}{2}\phi^{-4}\nabla_{C\dot A}\nabla_{D\dot B}\Theta l^C l^D - \tfrac{2}{3}\phi^2\xi_{(\dot A}\pi_{\dot B)}, \tag{4.17}
$$

where

$$
\Theta \equiv \Delta^{-2}\Theta',
$$

[cf., Eq. (3.15)] and

$$
\xi_{\dot A} \equiv l^{-1\dot C}{}_{\dot A}L_{\dot C},\quad \pi_{\dot B} \equiv l^{-1\dot D}{}_{\dot B}\,p_{\dot D}.
$$

Using Eq. (3.12) one finds that, in this case, the fact that $L_{\dot A}$ is a function of $q^{\dot R}$ only amounts to $l^A\nabla_{A\dot B}\xi_{\dot C}=0$.

5. SOME FINAL REMARKS

According to the terminology of Ref. 18, Eq. (4.4) means that the metric of a space-time which admits a congruence of null strings defined by a multiple DP spinor is KS conjugated to a conformally flat metric ${}^0g$. Denoting by ${}^0\nabla$ the (conformally flat) connection corresponding to ${}^0g$ and taking ${}^0\nabla_{A\dot B}\equiv{}^0\nabla_{{}^0\partial_{A\dot B}}$, from Eqs. (4.4), (4.6), and (2.11), one finds

$$ {}^0\nabla_{A\dot B} l_C = \nabla_{A\dot B} l_C + \tfrac{1}{2}\phi^{-1} l_A l_C l_S \nabla^{S\dot R}(\phi^{-1}\Omega_{\dot B\dot R}). \tag{5.1} $$

Thus, the spinor l_A also defines a congruence of null strings in the structure with the metric ${}^0g$ and from Eqs. (5.1), (4.13), and (4.4) it follows that

$$ {}^0\nabla_{A\dot B} l_C = \tfrac{1}{2} l_C\,[3\,{}^0\partial_{A\dot B}\ln\phi + {}^0\partial_{A\dot B}\ln\Delta\,] + \epsilon_{AC} l^D\,{}^0\partial_{D\dot B}\ln\phi. \tag{5.2} $$

The normalization condition (2.3) is equivalent to the existence of a vector field $Z_{A\dot B}$, the Sommers vector,[8,6] such that

$$ \nabla_{A\dot B} l_C = 3 Z_{A\dot B} l_C + 2\epsilon_{AC} l^D Z_{D\dot B}. \tag{5.3} $$

However, if ψ is a nonvanishing function such that $l^A\partial_{A\dot B}\psi = 0$, then ψl_A also satisfies (2.3) and $Z_{A\dot B}$ is replaced by $Z_{A\dot B} + \tfrac{1}{3}\partial_{A\dot B}\ln\psi$. Thus, given a congruence of null strings, the Sommers vector associated to it is defined modulo these transformations. By comparing Eqs. (5.3) and (4.13) one sees that the Sommers vector associated to the congruence defined by l_A can be chosen as

$$ Z_{A\dot B} = \tfrac{1}{2}\partial_{A\dot B}\ln\phi + \tfrac{1}{6}\phi^{-4} l_A l^D \nabla_D{}^{\dot R}(\phi^2\Omega_{\dot B\dot R}). $$

While from (5.2) one finds that, with respect to the conformally flat structure

$$ {}^0Z_{A\dot B} = \tfrac{1}{2}{}^0\partial_{A\dot B}\ln\phi. $$

It should be remarked that the metric ${}^0g$ is not uniquely defined since it depends on the choice of coordinates q^A, $p^{\dot A}$, and the function ϕ. If $\tilde\phi$ is another solution of (2.10) and $\tilde q^A$, $\tilde p^{\dot A}$ is another set of coordinates constructed as in (3.1) and (3.2) then ${}^0\tilde g\equiv 2\tilde\phi^{-2} d\tilde q^A \otimes_s d\tilde p_{\dot A} = {}^0g + 2\rho\phi^{-2}(\partial\tilde q^A/\partial q^B) \times(\partial\tilde p_{\dot A}/\partial q^{\dot C})dq^B \otimes_s dq^{\dot C}$, where $\rho\equiv\phi^2/\tilde\phi^2$ is a function of $q^{\dot R}$ only and $\tilde p_{\dot A} = \rho^{-1}(\partial q^{\dot C}/\partial\tilde q^{\dot A})p_{\dot C} + \sigma_{\dot A}$, where $\sigma_{\dot A} = \sigma_{\dot A}(q^{\dot R})$. In general, the metric ${}^0g$ will be complex. If $\partial_{A\dot B}$ is Hermitian, then not necessarily each term in (4.4) will be separately Hermitian.

Finally, it is easy to see that the Eqs. (4.4), (4.10), (4.15), (4.16), and (4.17) apply even when l_A does not satisfy the normalization condition (2.3).

ACKNOWLEDGMENT

The author gratefully acknowledges his indebtedness to Professor J. F. Plebański for very valuable suggestions and discussions.

[1] J. F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976).
[2] J. D. Finley, III, and J. F. Plebański, J. Math. Phys. **17**, 2207 (1976).
[3] A. García, J. F. Plebański, and I. Robinson, Gen. Relativ. Gravit. **8**, 841 (1977). For a spinorial derivation see J. D. Finley, III, and J. F. Plebański, J. Math. Phys. **18**, 1662 (1977).
[4] G. F. Torres del Castillo, J. Math. Phys. **24**, 590 (1983).
[5] J. F. Plebański and I. Robinson, J. Math. Phys. **19**, 2350 (1978).
[6] J. F. Plebański and K. Rózga, "The optics of null strings," J. Math. Phys. (in press).
[7] F. A. E. Pirani, in *Lectures on General Relativity*, 1964 *Brandeis Summer Institute*, Vol. 1, edited by S. Deser and K. W. Ford (Prentice-Hall, Englewood Cliffs, NJ, 1965).
[8] P. Sommers, Proc. R. Soc. London Ser. A **349**, 309 (1976).
[9] J. F. Plebański, J. Math. Phys. **16**, 2395 (1975).
[10] J. F. Plebański and S. Hacyan, J. Math. Phys. **16**, 2403 (1975).
[11] I. Robinson, J. Math. Phys. **2**, 290 (1961).
[12] Compare Ref. 7, p. 363.
[13] See also E. J. Flaherty, Jr., in *General Relativity and Gravitation*, Vol. 2, edited by A. Held (Plenum, New York, 1980), p. 228.
[14] For some applications of these potentials in the theory of Einstein–Maxwell and Einstein–Weyl equations see Refs. 3 and 4, respectively.
[15] J. M. Cohen and L. S. Kegeles, Phys. Lett. A **47**, 261 (1974); Phys. Rev. D **10**, 1070 (1974); Phys. Lett. A **54**, 5 (1975); Phys. Rev. D **19**, 1641 (1979). See also R. M. Wald, Phys. Rev. Lett. **41**, 203 (1978).
[16] J. F. Plebański, Acta Phys. Pol. **27**, 361 (1965).
[17] R. Penrose, Proc. R. Soc. London Ser. A **284**, 159 (1965).
[18] J. F. Plebański and A. Schild, Nuovo Cimento B **35**, 35 (1976).

Higher-dimensional Riemannian geometry and quaternion and octonion spaces

J. W. Moffat
Department of Physics, University of Toronto, Toronto, Ontario, M5S 1A7, Canada

(Received 10 May 1983; accepted for publication 5 August 1983)

An eight-dimensional Riemannian geometry is shown to be the basis of a nonsymmetric theory of gravitation. A hyperbolic complex structure is imposed and the group structure is $GL(8,R) \mapsto GL(4,R) \otimes GL(4,R) \supset GL(4,R)$. Octonion and quaternion division algebras are used to represent geometrical quantities and spinors. A Lagrangian is constructed that is related to supersymmetry and supergravity theories. The group structure for a hyperbolic octonion scheme is $GL(8,q_H) \mapsto GL(4,\mathbb{O}_H) \simeq GL(4,q_H) \otimes GL(4,q_H) \supset GL(4,q_H)$, while a simpler scheme based on hyperbolic quaternions is $GL(8,\mathbb{C}) \to GL(4,q_H) \simeq GL(4,\mathbb{C}) \otimes GL(4,\mathbb{C}) \supset GL(4,\mathbb{C})$.

PACS numbers: 04.50. + h, 02.40.Ky

1. INTRODUCTION

The nonsymmetric theory of gravitation[1-4] in four real space-time dimensions, based on a nonsymmetric metric $g_{\mu\nu} = g_{(\mu\nu)} + g_{[\mu\nu]}$ has a natural geometrical formulation in an eight-dimensional space.[5] The metric $g_{\mu\nu}$ and its conjugate $\tilde{g}_{\mu\nu}$ ($\tilde{g}_{\mu\nu} = g_{(\mu\nu)} - g_{[\mu\nu]}$) can be associated with two tangent spaces T_x and T'_x, and the $D = 8$ space can be described by the product $T_x \times T'_x$. In Ref. 5, the conjugate metric $\tilde{g}_{\mu\nu}$ was related to the hypercomplex numbers ϵ with $\epsilon^2 = +1$, so that $g^c_{\mu\nu} = g_{(\mu\nu)} + \epsilon g_{[\mu\nu]}$ and $\tilde{g}^c_{\mu\nu} = g_{(\mu\nu)} - \epsilon g_{[\mu\nu]}$. Such hyperbolic complex numbers were used by Gödel[6] in terms of "split" quaternions or "hyperbolic" quaternions, which belong to a real subalgebra of the complexified quaternion algebra that is not equivalent to the ordinary real quaternion algebra.

In the $D = 8$ space the geometry is (pseudo-) Riemannian and the metric is symmetric,

$$g_{\Sigma\Lambda} = g_{\Lambda\Sigma} \quad (\Sigma,\Lambda = 1,2,...,8). \tag{1.1}$$

Indeed the $g_{\Sigma\Lambda}$ can be written as a matrix

$$g_{\Sigma\Lambda} = \begin{pmatrix} g_{\mu\nu} & g_{\bar{\mu}\nu} \\ g_{\mu\bar{\nu}} & g_{\bar{\mu}\bar{\nu}} \end{pmatrix}, \tag{1.2}$$

where in this notation[5] $g_{\mu\nu} = g_{\nu\mu}$ ($\mu,\nu = 1,2,3,4$) ($\bar{\mu},\bar{\nu} = 5,6,7,8$), $g_{\bar{\mu}\nu} = -g_{\mu\bar{\nu}}$, and $g_{\bar{\mu}\bar{\nu}} = g_{\bar{\nu}\bar{\mu}} = -g_{\mu\nu}$.

In the following we shall be mainly concerned with the *bein* and spinor structure of the theory. Spinors have some unique properties in $D = 8$ spaces which we shall discuss below, and they can be associated with quaternions and octonions. Such an association suggests an intimate relation to supersymmetry and supergravity.[7]

2. *ACHTBEINS* AND SPIN CONNECTIONS

Since the metric in $D = 8$ space is symmetric we can write it in terms of "*achtbeins*" e^A_Σ ($A,\Sigma = 1,2,...,8$) as[8]

$$g_{\Sigma\Lambda} = e^A_\Sigma e^B_\Lambda \eta_{AB}, \tag{2.1}$$

where $\eta_{AB} = \mathrm{diag}(-1,-1,-1,+1,+1,+1,+1,-1)$ and

$$e^A_\Sigma = \left(\frac{\partial \xi^A_X(x)}{\partial x^\Sigma} \right)_{x=X}. \tag{2.2}$$

Here ξ^A_X denotes a locally "inertial" frame at each point $x = X$ in our $D = 8$ space. We choose the signature of our M_8 manifold such that the number of spatial dimensions is $s = 4$ and the number of time dimensions is $t = 4$. The *achtbeins* satisfy

$$e^A_\Sigma e^\Lambda_A = \delta^\Lambda_\Sigma \tag{2.3}$$

and

$$e^A_\Sigma e^\Sigma_B = \delta^A_B. \tag{2.4}$$

Then,

$$g_{AB} = e^\Sigma_A e^\Lambda_B \eta_{\Sigma\Lambda} = \eta_{AB}. \tag{2.5}$$

We can introduce a tangent space covariant derivative in the form

$$D_A = e^\Sigma_A \left(\frac{\partial}{\partial x_\Sigma} + \Omega_\Sigma \right), \tag{2.6}$$

where Ω_Σ is the spin connection given by

$$\Omega_\Sigma = \frac{1}{2} \sigma^{[AB]} e^\Lambda_A(x) \left(\frac{\partial}{\partial x^\Sigma} e_{B\Lambda}(x) \right). \tag{2.7}$$

The $\sigma_{[AB]}$ are a set of constant matrices that are skew symmetric in A and B and satisfy the relations

$$[\sigma_{AB},\sigma_{CD}] = \eta_{CB}\sigma_{AD} - \eta_{CA}\sigma_{BD} + \eta_{DB}\sigma_{CA} - \eta_{DA}\sigma_{CB}. \tag{2.8}$$

The *achtbeins* e^A_Σ satisfy

$$\partial_\Sigma e^A_\Lambda + (\Omega_\Sigma)^A_B e^B_\Lambda - \Gamma^\Omega_{\Sigma\Lambda} e^A_\Omega = 0, \tag{2.9}$$

where $\Gamma^\Omega_{\Sigma\Lambda}$ is the connection on the principal bundle of linear coframes in the $D = 8$ space. We can express Γ in terms of e and Ω:

$$\Gamma_{\Sigma\Lambda\Omega} = \Gamma^\Delta_{\Sigma\Lambda} g_{\Delta\Omega} = \eta_{AB} [\partial_\Sigma e^A_\Lambda + (\Omega_\Sigma)^A_C e^C_\Lambda] e^B_\Omega. \tag{2.10}$$

If we perform the transformation

$$e^A_\Sigma = U^A_\Omega e^\Omega_\Sigma, \tag{2.11}$$

then the metric (2.1) remains invariant if U is an element of SO(4,4). Demanding that Γ also be invariant under this transformation, leads to the equation

$$(\Omega_\Sigma)^A_B = [U\Omega_\Sigma U^{-1} - (\partial_\Sigma U)U^{-1}]^A_B. \tag{2.12}$$

Since the Ω_Σ transform like the generators of SO(4,4), then

$$\partial_\Sigma g_{\Omega\Lambda} - g_{\Delta\Lambda}\Gamma^{\Delta}_{\Sigma\Omega} - g_{\Omega\Delta}\Gamma^{\Delta}_{\Sigma\Lambda} = 0 \tag{2.13}$$

and the Γ acts in the $D = 8$ space as the connection coefficient for the metric in the sense of Riemannian differential geometry.

A calculation gives

$$([D_\Sigma, D_\Lambda])^A_B = (R_{\Sigma\Lambda})^A_B, \tag{2.14}$$

where

$$(R_{\Sigma\Lambda})^A_B = \partial_\Sigma(\Omega_\Lambda)^A_B - \partial_\Lambda(\Omega_\Sigma)^A_B + ([\Omega_\Sigma, \Omega_\Lambda])^A_B. \tag{2.15}$$

We can now form a $D = 8$ scalar curvature

$$R = \eta^{AC} e^\Sigma_A e^\Lambda_B (R_{\Sigma\Lambda})^B_C. \tag{2.16}$$

By using R we can form our Lagrangian density

$$\mathscr{L} = eR, \tag{2.17}$$

where $e = \det(e^A_\Sigma)$.

3. REDUCTION TO $D = 4$ SPACE-TIME MANIFOLD

Our eight-dimensional vector space can be described in terms of the tangent space[5]

$$T'_x = T_x \times T_x, \tag{3.1}$$

so that elements of T'_x are ordered pairs of vectors (X, Y). The fiber bundle $L'(M_4)$ associated with a given $D = 4$ real manifold M_4 of T_x is

$$L'(M_4) = L(M_4) X_{\mathrm{GL}(4,R)} \mathrm{GL}(8,R), \tag{3.2}$$

where GL(4,R) has a natural (subgroup) right action on GL(8,R). We introduce a hyperbolic complex structure J on R^8 with $J^2 = +1$, that reduces GL(8,R) to GL(4,R) $\otimes$ GL(4,R) while preserving J.

We now choose one of our $D = 4$ space-time manifolds M_4 to be a constant hypersurface in our $D = 8$ manifold M_8. This fixes the suffixes Σ, Λ, etc. to the values μ, σ, etc. that take on the values 1,2,3,4, i.e., we "freeze out" $\bar{\mu}, \bar{\sigma} = 5,6,7,8$. Then the connection $\Gamma^A_{\mu B}$ denotes the 4×8^2 degrees of freedom of a GL(8,R) connection over M_4. We now require that[5]

$$\nabla J = 0, \tag{3.3}$$

where ∇ is the covariant derivative operator defined by

$$\Gamma^C_{\mu A} e_c = \nabla_{e_\mu} e_A, \tag{3.4}$$

with $e_C = e^\mu_C\, dx_\mu$ ($\mu = 1,2,3,4$). The condition (3.3) yields

$$\Gamma^\alpha_{\mu\beta} = \Gamma^\alpha_{\mu\bar{\beta}}, \tag{3.5a}$$

$$\Gamma^\alpha_{\mu\bar{\beta}} = -\Gamma^{\bar{\alpha}}_{\mu\beta}, \tag{3.5b}$$

and we are left with the 2×4^3 components of the connection over GL(4,R) $\otimes$ GL(4,R). We thus have a complex-valued connection in M_4:

$$\Gamma^{c\lambda}_{\mu\nu} = \Gamma^\lambda_{\mu\nu} + \epsilon\Gamma^{\bar{\lambda}}_{\mu\nu}. \tag{3.6}$$

In familiar notation

$$\Gamma^{c\lambda}_{\mu\nu} = \Gamma^\lambda_{(\mu\nu)} + \epsilon\Gamma^\lambda_{[\mu\nu]} \tag{3.7}$$

with $\Gamma^\lambda_{(\mu\nu)} \equiv \frac{1}{2}(\Gamma^\lambda_{\mu\nu} + \Gamma^\lambda_{\nu\mu}) = \Gamma^\lambda_{\mu\nu}$ and $\Gamma^\lambda_{[\mu\nu]} \equiv \frac{1}{2}(\Gamma^\lambda_{\mu\nu} - \Gamma^\lambda_{\nu\mu}) = \Gamma^{\bar{\lambda}}_{\mu\nu}$. Then the conjugate connection is $\tilde{\Gamma}^{c\lambda}_{\mu\nu} = \Gamma^\lambda_{(\mu\nu)} - \epsilon\Gamma^\lambda_{[\mu\nu]}$ and $\tilde{\Gamma}^{c\lambda}_{\mu\nu} = \Gamma^{c\lambda}_{\nu\mu}$ is a "hypercomplex" Hermitian connection. The metric also has a (hypercomplex) sesquilinear form

$$g^c_{\mu\nu} = g_{\mu\nu} + \epsilon g_{\bar{\mu}\nu} \tag{3.8}$$

and $g_{\mu\nu}$ is hypercomplex Hermitian,

$$\tilde{g}^c_{\mu\nu} = g^c_{\nu\mu}. \tag{3.9}$$

In terms of hypercomplex *vierbeins* $e^{ca}_\mu = e^a_\mu + \epsilon e^a_{\bar{\mu}}$, we have

$$g^c_{\mu\nu} = \eta_{ab} e^a_\mu \tilde{e}^b_\nu \quad (\mu,\nu,a,b = 1,2,3,4). \tag{3.10}$$

Since $\epsilon^2 = 1$ the $g_{\mu\nu}$ acts as a "real" metric and does *not have any ghost poles* in the $D = 4$ version of the nonsymmetric theory of gravitation.[9] This is why we prefer working with the hypercomplex number structure which forms a *ring*, in contrast to the Hermitian structure, based on ordinary complex numbers ($J^2 = -1$) that form a field. The Lagrangian density in terms of our projected complex-valued tensors on M_4 is

$$\mathscr{L} = (-g^c)^{1/2}(a g^{c\mu\nu} R^{\ ca}_{\mu\nu\alpha} + b g^{c\mu\nu} R^{\ ca}_{\mu\alpha\nu}), \tag{3.11}$$

where a and b are real parameters. It was proved in Ref. 10 that $\mathscr{L}$ is pure real and is the Lagrangian of the nonsymmetric theory of gravitation.[1-4] Note that the second term in (3.11) is the generalized Ricci scalar in the nonsymmetric theory, while the first term is the trace of the second contraction of the generalized Ricci tensor. The latter is zero in four-dimensional Riemannian geometry but nonzero in the present case due to the antisymmetric part of the hyperbolic complex metric $g^c_{\mu\nu}$.

4. EIGHT-DIMENSIONAL SPINORS

In $D = 8$ space a general spinor Ψ has sixteen components. We write Ψ as

$$\Psi = \begin{pmatrix} \phi_i \\ \psi_i \end{pmatrix} \quad (i = 1,2,\ldots,8). \tag{4.1}$$

A rotation transforms (4.1) as

$$\Psi' = M_{AB}\Psi \quad (A, B = 1,2,\ldots,8), \tag{4.2}$$

where M_{AB} are 8×8 matrices. The Dirac equation is

$$(iD_A\Gamma^A + m)\Psi = 0, \tag{4.3}$$

where D_A is the covariant differential operator in Eq. (2.6) and the eight matrices Γ_A satisfy

$$\{\Gamma_A, \Gamma_B\} = 2\eta_{AB} \quad (A,B = 1,2,\ldots,8). \tag{4.4}$$

The smallest nontrivial representation of the Γ_A has to be 16-dimensional. We choose an Hermitian representation for Γ_A

$$\Gamma_A = \begin{pmatrix} 0 & \alpha_A \\ \beta_A & 0 \end{pmatrix}, \tag{4.5}$$

where all elements in (4.5) are 8×8 matrices with $\alpha^\dagger_A = \alpha_A$ and $\beta^\dagger_A = \beta_A$.

A vector in M_8 can be represented by

$$\mathbf{A} = \sum_{A=1}^{8} a_A \Gamma^A. \tag{4.6}$$

An infinitesimal rotation by an angle θ in the A,B plane is determined by the operator

$$M_{AB} = 1 + \tfrac{1}{2}\theta\Gamma_A\Gamma_B \tag{4.7}$$

and

$$\mathbf{A}' = M_{AB}\mathbf{A}M_{AB}^{-1}. \tag{4.8}$$

The quantity

$$\mathbf{\Psi}^T\mathbf{B}\mathbf{\Psi} \tag{4.9}$$

is invariant under rotations, where $\mathbf{B}$ is a matrix satisfying $\mathbf{B}^2 = 1$.

In an eight-dimensional space, the eight component spinors ψ and ϕ and the vector $\mathbf{A}$ are equivalent, since they all three have the same number of components in a $D = 8$ space. Let the dimension of Ψ in spinor space be 2^ν. We then have $D = 2\nu$ and $D = 2\nu - 1$ for even and odd D, respectively. The dimensions of the subspinors ψ and ϕ for even D is $2^{\nu-1}$. Only for $D = 8$ is $D = 2^{\nu-1}$ satisfied, which makes $D = 8$ a special space in that the principle of "triality" is valid,[11-13] i.e., a rotation by an angle φ induces corresponding rotations by angles $\frac{1}{2}\theta$ in $\mathbf{A}$ and ψ and ψ, φ, and $\mathbf{A}$ are equivalent and cannot be distinguished.

5. QUATERNIONS AND OCTONIONS

A division algebra is a linear algebra with an identity and an inverse for every element except zero. If a norm N exists then

$$N(AB) = N(A)N(B), \quad A,B \in A,\ N \in R\ (N \geqslant 0). \tag{5.1}$$

Hurwitz's theorem[14] states that there are only four such algebras. Their elements are identified, respectively, with the real numbers, the complex numbers C, the quaternions q, and the octonions O (Cayley numbers). The existence and properties of division algebras have been related to those of supersymmetric field theories in various higher-dimensional spaces.[15] There also appears to be a relation of division algebras to supergravity.

Octonions are based on the eight units $e_1,e_2,...,e_8$ that satisfy

$$\begin{gathered} e_1^2 = e_1, \quad e_k^2 = -e_1, \quad \tilde{e}_k = -e_k \quad (k = 2,3,...,8), \\ \{e_k,e_l\} = \delta_{kl} \quad (k,l = 2,3,4,...,8). \end{gathered} \tag{5.2}$$

We shall use split octonions or hyperbolic octonions defined by

$$\begin{gathered} j_1 = e_1, \quad j_2 = e_2, \quad j_k = ie_k, \quad j_k^2 = e_1, \\ \tilde{j}_k = j_k \quad (k = 3,4,...,8), \\ [j_k,j_l] = -\delta_{kl} \quad (k,l = 3,4,...,8), \end{gathered} \tag{5.3}$$

where i is the ordinary pure imaginary number ($i^2 = -1$) which is assumed to commute with all e_A ($A = 1,2,...,8$). A hyperbolic octonion can be written as

$$\mathbf{C} = c_1 j_1 + c_2 j_2 + c_3 j_3 + \cdots + c_8 j_8, \tag{5.4}$$

and the conjugate hyperbolic octonion is

$$\widetilde{\mathbf{C}} = c_1 j_1 - c_2 j_2 + c_3 j_3 + \cdots + c_8 j_8. \tag{5.5}$$

Then the norm of C is

$$N(\mathbf{C}) \equiv \mathbf{C}\cdot\widetilde{\mathbf{C}} = c_1^2 + c_2^2 - c_3^2 - c_4^2 - \cdots - c_8^2. \tag{5.6}$$

Hyperbolic octonion multiplication is noncommutative and nonassociative.

We now represent our *achtbeins* in terms of the eight hyperbolic octonions in the form

$$\mathbf{E}_\Sigma = e_\Sigma^1 j_1 + e_\Sigma^2 j_2 + \cdots + e_\Sigma^8 j_8. \tag{5.7}$$

Our eight-component spinors ψ and ϕ are represented by

$$\psi = \psi_1 j_1 + \psi_2 j_2 + \cdots + \psi_8 j_8, \tag{5.8a}$$

$$\phi = \phi_1 j_1 + \phi_2 j_2 + \cdots + \phi_8 j_8. \tag{5.8b}$$

Thus our eight-dimensional tangent space has hyperbolic octonion *achtbeins* and spinors that are noncommutative and nonassociative.

The Lagrangian density $\mathscr{L}$ in (2.17) is now a scalar density invariant under the transformations of the group $GL(8,q_H)$ or $GL(32,R)$, where q_H denotes the hyperbolic quaternion algebra. The reduction to a four-dimensional manifold, as described in Sec. 3, is done through the group reduction

$$GL(8,q_H) \rightarrow GL(4,O_H) \simeq GL(4,q_H) \otimes GL(4,q_H). \tag{5.9}$$

The imposition of a metrically compatible connection, then yields the reduction

$$GL(4,q_H) \otimes GL(4,q_H) \rightarrow GL(4,q_H). \tag{5.10}$$

In (5.9) O_H denotes the hyperbolic octonion algebra.

We should note the isomorphism $SL(8,q_H) \cong SU^*(16)$ which is important for our octonion space O_x.

We shall use hyperbolic quaternions in the form

$$\begin{gathered} j_1 = e_1, \quad j_k = ie_k \quad (k = 2,3,4), \\ j_1^2 = e_1, \quad \tilde{j}_k = -j_k, \end{gathered} \tag{5.11}$$

and

$$\{j_k,j_l\} = -\delta_{kl} \quad (k,l = 2,3,4). \tag{5.12}$$

Then the *vierbeins* e_μ^a in the four-dimensional quaternion tangent space Q_x and the four-dimensional manifold M_4 are represented by

$$\mathbf{e}_\mu = e_\mu^1 j_1 + e_\mu^2 j_2 + e_\mu^3 j_3 + e_\mu^4 j_4. \tag{5.13}$$

A four-component spinor ψ is represented by

$$\psi = \psi_1 j_1 + \psi_2 j_2 + \psi_3 j_3 + \psi_4 j_4. \tag{5.14}$$

The eight-dimensional hyperbolic quaternion tangent space $Q'_x = Q_x \times Q_x$ is described by the transformations of the group $GL(8,C)$. The latter group preserves the metric (2.5). We can also consider the invariance of the quaternion space in a real $D = 16$ space under the transformations of $GL(16,R)$. The reduction to a $D = 4$ space-time manifold is achieved through

$$GL(8,\mathbb{C}) \rightarrow GL(4,q_H) \simeq GL(4,\mathbb{C}) \otimes GL(4,\mathbb{C}) \rightarrow GL(4,\mathbb{C}). \tag{5.15}$$

Thus the group that preserves the $D = 4$ metric (3.10) in the four-dimensional tangent space Q_x is $GL(4,C)$ which generalizes the corresponding (gauge) group $GL(4,R)$ in the tangent space T_x discussed in Ref. 5.

The Lagrangian $\mathscr{L}$ in our hyperbolic octonion tangent space O_x is an interesting candidate for a supergravity theory. But the $\mathscr{L}$ in the hyperbolic quaternion space $Q'_x = Q_x \times Q_x$ is a simpler theory, since the algebra is associative.

6. CONCLUSIONS

We have shown that the nonsymmetric theory of gravitation has a unique geometrical structure that is (pseudo-)

Riemannian in a real eight-dimensional space with the signature $(- - - - + + + +)$. A hypercomplex structure is imposed as in Ref. 5, and the group $GL(8,R)$ reduces to $GL(4,R) \otimes GL(4,R) \supset GL(4,R)$. The spinors in the $D = 8$ space can be constructed and they satisfy the principle of *triality*. A Dirac equation can be written in terms of a Clifford algebra, based on eight matrices Γ_A. The principle of triality, which says that e^A_Σ, ψ, and ϕ are indistinguishable, already suggests a supersymmetric structure in terms of a *discrete* symmetry that is naturally built into the scheme in an eight-dimensional space.

We generalized our R^8 space to a hyperbolic octonion space O_x of *achtbeins* and found that the group $GL(8,q_H) \to GL(4,\mathbb{O}_H) \simeq GL(4,q_H) \otimes GL(4,q_H) \supset GL(4,q_H)$ was the basic fiber structure. A simpler structure was then developed for the hyperbolic quaternion space $Q'_x = Q_x \times Q_x$ with the group structure $GL(8,\mathbb{C}) \to GL(4,q_H) \simeq GL(4,\mathbb{C}) \otimes GL(4,\mathbb{C}) \to GL(4,\mathbb{C})$. This scheme is closely related to a Grassman algebra structure of noncommutative operators and could form the basis of an elegant supergravity theory.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada. I am grateful to Dr. M. Green, Dr. G. Kunstatter, and Dr. M. Kalinowski for helpful discussions.

[1]J. W. Moffat, Phys. Rev. D **19**, 3554 (1979).
[2]J. W. Moffat, J. Math. Phys. **21**, 1798 (1980).
[3]J. W. Moffat, Phys. Rev. Lett. **50**, 709 (1983).
[4]For a review see J. W. Moffat, in *The Origin and Evolution of Galaxies*, edited by V. de Sabbata (World Scientific, Singapore, 1982), p. 127.
[5]G. Kunstatter, J. W. Moffat, and J. Malzan, J. Math. Phys. **24**, 886 (1983).
[6]K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[7]T. Kugo and P. Townsend, Nucl. Phys. B **221**, 357 (1983).
[8]J. W. Moffat, Ann. Inst. Henri Poincaré **34**, 85 (1981).
[9]R. B. Mann and J. W. Moffat, Phys. Rev. D **26**, 1858 (1982).
[10]G. Kunstatter and R. Yates, J. Phys. A **14**, 847 (1981).
[11]E. Cartan, *Lecons sur la theorie des spineurs* (Hermann & Cie, Paris, 1938), Vols. I and II.
[12]C. C. Chevalley, *The Algebraic Theory of Spinors* (Columbia University Press, New York, 1954).
[13]A. Gamba, J. Math. Phys. **8**, 775 (1967).
[14]See, for example, Ref. 11.
[15]See, for example, Ref. 7 and references contained therein.

Unification of generalized electromagnetic and gravitational fields

B. S. Rajput
Department of Physics, Kumaun University, Naini Tal-263 002, India

(Received 27 July 1982; accepted for publication 7 January 1983)

Structural symmetry between generalized gravitational field (gravito-Heavisidian field) and the generalized electromagnetic field associated with dyons has been demonstrated, and the field equations, equation of motion, and the quantization condition for angular momentum operators for both these fields have been unified.

PACS numbers: 04.50. + h, 03.50.De, 11.10.Cd

Recently, the question of existence of monopole[1,2] has become a challenging new frontier and the object of more interest in connection with quark confinement problem of quantum chromodynamics. The eighth decade of this century witnessed a rapid development of the group theory and gauge field theory to establish the theoretical existence of monopoles and to explain their group properties and symmetries. Keeping in mind t'Hooft's solutions[3] and the fact that despite the potential importance of monopoles, the formalism necessary to describe them has been clumsy and not manifestly covariant, we have recently developed[4,5] the self-consistent quantum field theory of generalized electromagnetic fields associated with dyons (particles carrying electric and magnetic charges) by using two 4-potentials and assuming the generalized charge, generalized 4-current, and generalized 4-potential associated with dyons as complex quantities with their eal and imaginary parts as electric and magnetic constituents. Postulating the Heavisidian monopole[6,7] and keeping in mind the recent interest in the linear equations for gravitational field,[6] the present paper demonstrates the structural symmetry between the generalized gravitational field (linear) and the generalized electromagnetic field (associated with dyons), and unifies the field equations, equations of motion, and the quantization conditions for angular momentum operators for both these fields in a consistent symmetrical manner.

Assuming the existence of magnetic monopole in order to explain the quantization of electric chage, Dirac[1] generalized the Maxwell's field equations into following form:

$$\begin{aligned}
&\nabla\cdot\mathbf{E} = j_0,\\
&\nabla\cdot\mathbf{H} = k_0,\\
&\nabla\times\mathbf{E} = -\frac{\partial\mathbf{H}}{\partial t} - \mathbf{k},\\
&\nabla\times\mathbf{H} = \frac{\partial\mathbf{E}}{\partial t} + \mathbf{j} \qquad (c=\hbar=1),
\end{aligned} \tag{1}$$

where j_0 and k_0 are electric and magnetic charge densities and $\mathbf{j}$ and $\mathbf{k}$ are the corresponding current densities. On the other hand, the linear equations for gravitational field with Heavisidian monopoles may be written as[7]

$$\begin{aligned}
&\nabla\cdot\mathbf{G} = -\rho_G,\\
&\nabla\cdot\mathscr{H} = -\rho_H,\\
&\nabla\times\mathbf{G} = -\frac{\partial\mathscr{H}}{\partial t} + \mathbf{j}_H,\\
&\nabla\times\mathscr{H} = \frac{\partial\mathbf{G}}{\partial t} - \mathbf{j}_G,
\end{aligned} \tag{2}$$

where $\mathbf{G}$ is gravitational field, $\mathscr{H}$ is Heavisidian field, ρ_G is gravitational charge (mass) density, π_H is Heavisidian charge density, $\mathbf{j}_G$ is gravitational current density, and $\mathbf{j}_H$ is Heavisidian current density. In order to write the sets of equations (1) and (2) in a unified form, let us introduce the generalized field function $\psi(x)$ such that

$$\psi(x) = \begin{cases} \mathbf{E} - i\mathbf{H} & \text{(for generalized electromagnetic fields)} \\ \mathbf{G} - i\mathscr{H} & \text{(for generalized gravitational fields, i.e., for gravito-Heavisidian field).} \end{cases}$$

Let us also introduce the generalized charge (mass) q and generalized 4-current source density J_μ such that

$$\begin{aligned}
q &= \begin{cases} e - ig & \text{(for dyons)} \\ m - ih & \text{(for gravito-dyons),} \end{cases}\\
J_\mu &= \begin{cases} j_\mu - ik_\mu & \text{(generalized 4-current)} \\ j_\mu^{(G)} - ik_\mu^{(H)} & \text{(generalized gravitational 4-current),} \end{cases}
\end{aligned} \tag{3}$$

where e and g are electric and magnetic charges on a dyon; m and h are gravitational and Heavisidian charges (masses) on a gravito-dyon; $j_\mu^{(G)}$ and $k_\mu^{(H)}$ are gravitational and Heavisidian 4-current densities given by

$$j_\mu^{(G)} = (\mathbf{j}_G, \rho_G) \qquad \text{and} \qquad k_\mu^{(H)} = (\mathbf{j}_H, \rho_H).$$

Here we have assumed that the generalized electromagnetic fields satisfying Eqs. (1) are produced by dyons carrying the generalized charges and the generalized gravitational fields (i.e., gravito-Heavisidian fields); the fields satisfying Eqs. (2) are produced by gravito-dyons carrying the generalized masses (charges). We may now unify the generalized field Eqs. (1) and (2) into the following form:

$$\begin{aligned}
&\nabla\cdot\psi = \alpha J_0,\\
&\nabla\times\psi = -i\frac{\partial\psi}{\partial t} - i\alpha\mathbf{J},
\end{aligned} \tag{4}$$

where J_0 and $\mathbf{J}$ are temporal and spatial parts of 4-vector

$\{J_\mu\}$ and α is a constant having the values $\alpha = +1$ for generalized electromagnetic fields and $\alpha = -1$ for generalized gravito-Heavisidian fields. This equation shows that while passing from generalized electromagnetic field to the generalized gravitational field, one must replace J_μ by $-J_\mu$ in field equations.

The Lorentz equation of motion of generalized charge q in the generalized field ψ may be written as

$$M\ddot{x} = \mathrm{Re}[q(\psi^* - i\mathbf{v}\times\psi^*)], \tag{5}$$

where Re denotes the real part and the effective mass M is given by

$$M = m - (\alpha - 1)\mathbf{h}/2. \tag{6}$$

For the dyon moving in the generalized electromagnetic field $M = m$ and Eq. (5) reduces to

$$m\ddot{x} = e[\mathbf{E} + \mathbf{v}\times\mathbf{H}] + g[\mathbf{H} - \mathbf{v}\times\mathbf{E}], \tag{7}$$

which is the usual Lorentz equation of motion for a dyon. For the gravito-dyon (the particle carrying gravitational charge m and Heavisidian charge h) moving in the generalized gravitational field, we have

$$M = m + h \qquad [\text{since } \alpha = -1 \text{ in Eq. (6)}],$$

and hence Eq. (5) reduces to

$$(m + h)\ddot{x} = m[\mathbf{G} + \mathbf{v}\times\mathscr{H}] + h[\mathscr{H} - \mathbf{v}\times\mathbf{G}] \tag{8}$$

which is similar to the result derived recently by Singh.[7] We may therefore treat Eq. (5) as the unified Lorentz equation of motion of generalized charge (mass) in the generalized fields. We have shown in our earlier papers[4,5] that in order to avoid the occurrence of unphysical string variables in the solution of generalized field equation and in the quantum field theoretical description of generalized fields, we have to introduce two 4-potentials $\{A_\mu\}$ and $\{B_\mu\}$. Introducing the generalized 4-potential $\{V_\mu\}$ such that

$$V_\mu = A_\mu - iB_\mu \equiv (\mathbf{V}, V_0), \tag{9}$$

the unified generalized field equations (4) may be written as follows:

$$\frac{\partial^2 V_\mu}{\partial x_\nu^2} - \frac{\partial^2 V_\nu}{\partial x_\nu \partial x_\mu} = \alpha J_\mu, \tag{10}$$

which is the unified generalized field equation in terms of the generalized 4-potential and the generalized 4-current. When the 4-potential $\{V_\mu\}$ satisfies the Lorentz condition, this equation reduces to

$$\Box V_\mu = \alpha J_\mu.$$

In other words, under the Lorentz condition we have

$$\Box V_\mu = \begin{cases} J_\mu & \text{(for generalized electromagnetic fields)} \quad (11a) \\ -J_\mu & \text{(for generalized gravitational fields),} \quad (11b) \end{cases}$$

which also shows the replacement of J_μ by $-J_\mu$ in the field equations while switching over to generalized gravitational fields from generalized electromagnetic fields. We have already shown in our earlier papers[8,9] that the usual subluminal electromagnetic fields when observed from a superluminal frame or a superluminal electromagnetic field when observed from a subluminal frame appear to satisfy the field equations (11b). We may therefore conclude that the field equations for generalized electromagnetic fields when transformed under complex superluminal Lorentz transformations (10) become the field equations for gravitational fields.

We find that the unified field equations (4) and the equation of motion (5) are dual-invariant under the following continuous dual transformations:

$$\begin{aligned} \mathrm{Re}\,\psi' &\to \mathrm{Re}\,\psi\cos\theta + \mathrm{Im}\,\psi\sin\theta, \\ \mathrm{Im}\,\psi' &\to -\mathrm{Re}\,\psi\sin\theta + \mathrm{Im}\,\psi\cos\theta, \\ \mathrm{Re}\,J_\mu &\to \mathrm{Re}\,J_\mu\cos\theta + \mathrm{Im}\,J_\mu\sin\theta, \\ \mathrm{Im}\,J_\mu &\to -\mathrm{Re}\,J_\mu\sin\theta + \mathrm{Im}\,J_\mu\cos\theta, \end{aligned}$$

where Re and Im denote real and imaginary parts. We also find that Eqs. (4) and (5) possess the symmetry under simultaneous space and time reflections combined with rotation in charge space,[5] and also under the combination of reflection in charge space with space and time reflections separately.

Using the equation of motion (5) of generalized charge (mass) in a generalized field, we may write the following expression for the angular momentum vector of jth generalized charge (mass) moving in the generalized field of kth generalized charge (mass), which is assumed at rest:

$$\mathbf{J} = \mathbf{r}\times\mathbf{p}\,1\,\mathrm{Im}(q_j q_k^*)\mathbf{r}/r, \tag{12}$$

where $\mathbf{p} = M_j(d\mathbf{r}/dt)$ and Im denotes the imaginary part. In deriving this result, we have substituted

$$\psi_k = q_j \mathbf{r}/r^3 \tag{13}$$

in Eq. (5). Equation (12) is similar to the recent result of A. Singh.[7] But this angular momentum is not acceptable in the presence of magnetic monopole (Heavisidian charge) because it is not gauge invariant. The angular momentum vector, which is both gauge invariant and rotationally symmetric, may be written in the following form:

$$\mathbf{J} = \mathbf{r}\times[\mathbf{p} - \mathrm{Im}(q_j q_k^*)\mathbf{V}] + \mathrm{Im}(q_j q_k^*)\mathbf{r}/r, \tag{14}$$

where $\mathbf{V}$ is the spatial part of generalized 4-potential $\{V_\mu\}$. We may also write the following expression for the gauge-invariant linear momentum of ith particle carrying the generalized charge (mass) q_j in the field of kth generalized charge (mass):

$$\mathbf{\Pi} = \mathbf{p} - \mathrm{Im}(q_j q_k^*)\mathbf{V}. \tag{15}$$

Using Eqs. (14) and (15), we get the following gauge-invariant and rotationally symmetric commutation relations for the linear momentum operator $\hat{\mathbf{\Pi}}$, the position operator $\hat{r}$, and the angular momentum operator $\hat{J}$ of the generalized charges (masses):

$$\begin{aligned} [\hat{J}_m, \hat{r}_n] &= i\epsilon_{mnp}\hat{r}_p, \\ [\hat{\Pi}_m, \hat{r}_n] &= i\delta_{mn}, \\ [\hat{\Pi}_m, \hat{\Pi}_n] &= -\mathrm{Im}\, q_j q_k^* \epsilon_{mnp}\psi_p^T, \\ [\hat{J}_m, \hat{\Pi}_n] &= i\epsilon_{mnp}\hat{\Pi}_p, \\ [\hat{J}_m, \hat{J}_n] &= i\epsilon_{mnp}\hat{J}_p \end{aligned} \tag{16}$$

where $\psi^T = -i\nabla\times\mathbf{V}$ is the transverse part of ψ. From Eq. (14) we get the scalar

$$(\mathbf{r}\cdot\mathbf{J})/r = \mathrm{Im}(q_j q_k^*), \tag{17}$$

which commutes with all the observables. Equation (14) also shows that there is a residual angular momentum

$$\hat{J}_{\rm res} = \mathrm{Im}(q_j q_k^*)\hat{r}/r$$

carried by generalized fields of generalized charges besides the orbital and spin angular momentum of each particle. Furthermore, if the generalized momentum given by Eq. (14) is quantized along the line joining generalized charges (masses) q_j and q_k, we obtained the following quantization condition for generalized charges (gravitational charges):

$$\mathrm{Im}(q_j q_k^*) = n, \tag{18}$$

where n is an integer. It reduces to the following chirality quantization condition for the generalized electromagnetic fields:

$$(e_j g_k - e_k g_j) = 0, \pm 1, \pm 2, \cdots, \tag{19}$$

which is identical to the result of Zwanziger[11] and to that derived in our earlier paper[4] by different approach. For the generalized gravitational field (i.e., gravito-Heavisidian field), Eq. (18) reduces to

$$m_j h_k - m_k h_j = 0, \pm 1, \pm 2, \cdots, \tag{20}$$

where m_j and h_j are gravitational and Heavisidian charges (masses) on jth gravito-dyon.

It has been shown in our earlier papers[12] that the condition (19) leads to a very large force between two opposite magnetic charges as compared to that between negative and positive unit charges. It could explain, to some extent at least, the negative results of the experimental search for monopole. In a similar manner, Eq. (20) leads to the enormous forces between Heavisidian charges and suggests that Heavisidian monopoles are the most strongly interacting form of matter.

[1]P. A. M. Dirac, Proc. R. Soc. London A **133**, 60 (1931); Phys. Rev. **74**, 817 (1948).

[2]J. Schwinger, Science **165**, 757 (1969).

[3]G. t'Hooft, Nucl. Phys. B **79**, 276 (1974); B **138**, 1 (1978).

[4]B. S. Rajput and D. C. Joshi, Hadronic J. **4**, 1805 (1981); Pramana **13**, 637 (1979).

[5]B. S. Rajput and Om Prakash, Indian J. Phys. A **53**, 274 (1979); Indian J. Pure Appl. Phys. **16**, 593 (1978).

[6]D. D. Cantani, Nuovo Cimento B **60**, 67 (1980).

[7]A. Singh, Lett. Nuovo Cimento **32**, 1, 232 (1981).

[8]B. S. Rajput and O. P. S. Negi, Lett. Nuovo Cimento **32**, 117 (1981).

[9]O. P. S. Negi, H. C. Chandola, K. D. Purohit, and B. S. Rajput, Phys. Lett. B **105**, 281 (1981).

[10]E. Recami and R. Mignani, Nuovo Cimento **4**, 209 (1974) and references therein.

[11]D. Zwanziger, Phys. Rev. **176**, 1489 (1968); Phys. Rev. D **3**, 880 (1971).

[12]B. S. Rajput and Om Prakash, Indian J. Phys. **50**, 929 (1976).

Stochastic processes driven by dichotomous Markov noise: Some exact dynamical results

J. M. Sancho
Departamento de Física Teórica, Facultad de Física, Diagonal, 647, Barcelona-28, Spain

(Received 29 September 1982; accepted for publication 13 January 1983)

Stochastic processes defined by a general Langevin equation of motion where the noise is the non-Gaussian dichotomous Markov noise are studied. A non-Fokker–Planck master differential equation is deduced for the probability density of these processes. Two different models are exactly solved. In the second one, a nonequilibrium bimodal distribution induced by the noise is observed for a critical value of its correlation time. Critical slowing down does not appear in this point but in another one.

PACS numbers: 05.40. + j, 02.50. + s, 02.30.Jr

1. INTRODUCTION

Stochastic differential equations have an important and successful role in the theory of nonequilibrium phenomena. Most of them them are Langevin type, which are first-order differential equations with stochastic terms. In some cases the stochastic forces enter additively and often it is assumed that they represent internal fluctuations. In other cases the noise enters externally by means of a parameter of the phenomenological equation of motion which fluctuates. This kind of external noise has received a great deal of attention because it can represent a fluctuating external environment or a controlled noise generated in the laboratory by specific devices and introduced in the system in order to study its influence. This external noise is independent of the system and it is characterized by its intensity and correlation time. Examples of the influence of external noise can be found in a variety of systems, such as electric circuits[1] or liquid crystals,[2] among others.

The mathematical study of these equations begins with the modeling of the noise. The simplest assumption is to take a Gaussian white noise which has zero correlation time. In this case, the process is Markovian and a Fokker–Planck equation for the probability density always exists.[3] Nevertheless, this noise cannot always substitute for a real noise, which has a finite (perhaps small, but not zero) correlation time. In this case, the hypothesis of white noise, although suitable for a general description of the process, does not explore all the possibilities of a real noise. If we want to take into account the color of the noise, we should choose a mathematically tractable colored noise. Although many possible noises exist[4] only two of them have been receiving enough consideration in the literature.

The first one is the Ornstein–Uhlenbeck process, which is Gaussian and obeys the same equation of motion as the velocity of a free Brownian particle. Stochastic processes driven by this noise have been studied in Refs. 5–8 and interesting features have been found which do not appear in the white noise assumption.

The second noise is the two-step Markov process or dichotomous noise.[9,10] This noise is not Gaussian but Markovian and its influence in the stochastic process has been studied in Refs. 10–12. Interesting results, some of them quite similar to the former case, have been obtained, but only for the stationary state. A few dynamical properties are known in this case.[12] This paper will be mainly devoted to the study of the dynamics of two stochastic processes which allow an exact analysis. These are linear models except for a change of variables, but they are not trivial. Moreover they present mainly the second example—characteristics belonging to the nonlinear cases, such as the possibility of having a nonequilibrium bimodal distribution.

The study is carried out by means of the evaluation of the first two moments of the stochastic variable and the solution of the differential equation that obeys the probability density.

Section 2 is devoted to the problem of finding the differential equation satisfied by the probability of the stochastic process. A short summary of the mathematical tools is presented and a differential equation of the non-Fokker–Planck type is obtained for the probability density. This new result is particularized to exact cases, whose probability density obeys a second-order partial differential equation of the hyperbolic type.

In Sec. 3 we study two exact examples: the first one is a pure diffusive model and the second one is a linear case with linear drift and additive noise. This last case is the most relevant and it will show interesting nonequilibrium characteristics. In Sec. 4 we summarize the main results of this paper.

2. GENERAL THEORY

A. Differential equation for $P(q,t)$

Here we summarize some known results. One can assume quite generally[10,11] the following equation of motion for the variable q:

$$\dot{q} = f(q) + g(q)\xi(t), \tag{2.1}$$

where $\xi(t)$ is a stochastic force that we identify with the dichotomous noise or two-step Markov process.[13] This noise will only have two possible values $\pm \Delta$ with equal probability and jumps with probability $\frac{1}{2}\lambda\, dt$ for dt.[9] It has zero mean and autocorrelation

$$\langle \xi(t)\xi(t')\rangle = \Delta^2 \exp\{-\lambda|t-t'|\}. \tag{2.2}$$

By means of the "formula of differentiation" of Shapiro

 0022-2488/84/020354-06$02.50

and Loginov,[14]

$$\frac{\partial}{\partial t}\langle \xi(t)\phi[\xi(t)]\rangle = -\lambda\langle \xi(t)\phi[\xi(t)]\rangle + \left\langle \xi(t)\frac{\partial}{\partial t}\phi[\xi(t)]\right\rangle, \tag{2.3}$$

where $\phi[\xi(t)]$ is a functional of $\xi(t)$ and the average is over the distribution of $\xi(t)$, we can obtain a closed set of equations for the probability density $P(q,t)$.[11,14] The method we are following[11] is an alternative to that employed by Kitahara *et al.*[10] We begin with the stochastic Liouville equation[15] for the density $\rho(q,t)$ of a set of realizations of (2.1),

$$\dot\rho(q,t) = -\frac{\partial}{\partial q}(f(q)+g(q)\xi(t))\rho(q,t). \tag{2.4}$$

Taking the average over $\xi(t)$ and using Van Kampen's lemma[9]

$$P(q,t) = \langle \rho(q,t)\rangle, \tag{2.5}$$

we arrive at

$$\dot P(q,t) = -\frac{\partial}{\partial q}f(q)P(q,t) - \frac{\partial}{\partial q}g(q)P_1(q,t), \tag{2.6}$$

where

$$P_1(q,t) = \langle \xi(t)\rho(q,t)\rangle. \tag{2.7}$$

Since $\rho(q,t)$ is a functional of $\xi(t)$, we will use the formula of differentiation (2.3) to obtain an equation of motion for $P_1(q,t)$:

$$\dot P_1(q,t) = -\lambda P_1(q,t) - \frac{\partial}{\partial q}f(q)P_1(q,t) - \frac{\partial}{\partial q}g(q)\Delta^2 P(q,t), \tag{2.8}$$

where we have used the fact that the square of the dichotomous noise is a constant $\xi^2(t) = \Delta^2$.

The set of equations (2.6) and (2.8) form a closed system of linear partial differential equations whose solution will give us $P(q,t)$, provided that we know the initial condition $P(q,0)$. We also need another initial condition because we have two linear equations. The second one is obtained assuming statistical independence between $\xi(t)$ and $\rho(q,t)$ in $t = 0$[14]:

$$\langle \xi(t)\rho(q,t)\rangle|_{t=0} = P_1(q,0) = 0, \tag{2.9}$$

which implies in (2.6)

$$\left.\frac{\partial P(q,t)}{\partial t} - \frac{\partial}{\partial q}f(q)P(q,t)\right|_{t=0} = 0, \tag{2.10}$$

which together with

$$P(q,t)|_{t=0} = \delta(q) \tag{2.11}$$

will be the initial conditions of the system (2.6), (2.8).

A closed equation for $P(q,t)$ cannot easily be obtained. Nevertheless, a formal expression can be given[11] in the following way.

Let us formally integrate the linear equation (2.8):

$$P_1(q,t) = -\Delta^2\int_0^t \exp\left\{-\left(\lambda + \frac{\partial}{\partial q}f(q)\right)(t-t')\right\} \frac{\partial}{\partial q}g(q)P(q,t') \equiv -\Delta^2 B(q,t), \tag{2.12}$$

where we have used (2.9). Substituting it in (2.6) we arrive at a formal differential equation for $P(q,t)$[11]:

$$\dot P(q,t) = -\frac{\partial}{\partial q}f(q)P(q,t) + \Delta^2\frac{\partial}{\partial q}g(q)B(q,t). \tag{2.13}$$

The main object of this paper will be the exact solution, for two particular cases, of this integrodifferential equation.

Although the time-dependent solution is not always known, the stationary solution is well known and it reads, in the case that it exists,[10,11]

$$P_{st}(q) = N\frac{g(q)}{\Delta^2 g^2(q) - f^2(q)} \times\exp\left\{\lambda\int^q \frac{dq' f(q')}{\Delta^2 g^2(q') - f^2(q')}\right\}. \tag{2.14}$$

If in (2.13) we take

$$\exp\{-\lambda|t-t'|\} \simeq (2/\lambda)\delta(t-t'), \tag{2.15}$$

we arrive at the white noise limit for $\xi(t)$. This limit holds for $\lambda\to\infty$, $\Delta\to\infty$, and Δ^2/λ finite and it give us an insight about a possible perturbation procedure, taking λ as a large parameter.

B. Expansion in $1/\lambda$

Let us review a perturbative approach to (2.13) considering λ large. The limit $\lambda\to\infty$ is a very crude approximation because we lose all the specific characteristics of the dichotomous noise. Let us see how to retain the properties of $\xi(t)$ by means of an expansion in $1/\lambda$.

We take a time derivative in (2.13)

$$\ddot P(q,t) = -\frac{\partial}{\partial q}f(q)\dot P(q,t) + \Delta^2\frac{\partial}{\partial q}g(q)\frac{\partial}{\partial q}g(q)P(q,t) - \Delta^2\frac{\partial}{\partial q}g(q)\lambda\left(1 + \frac{1}{\lambda}\frac{\partial}{\partial q}f(q)\right)B(q,t). \tag{2.16}$$

If we use now the approximation (2.15) in $B(q,t)$, we obtain

$$\ddot P(q,t) = -\frac{\partial}{\partial q}f(q)\dot P(q,t) + \Delta^2\frac{\partial}{\partial q}g(q)\frac{\partial}{\partial q}g(q)P(q,t) - \Delta^2\frac{\partial}{\partial q}g(q)\left(1 + \frac{1}{\lambda}\frac{\partial}{\partial q}f(q)\right)\frac{\partial}{\partial q}g(q)P(q,t), \tag{2.17}$$

which is valid to first order in $1/\lambda$.

So we have reduced (2.13) to a second-order partial differential equation. The procedure is extended to the desired order in $1/\lambda$, deriving (2.16) succesively.

At this moment, a question arises: does a process exist which obeys an equation exactly similar to (2.17)? The answer is affirmative and we are going to study it in the next subsection.

C. Exact cases

In (2.16) we can use (2.13) to substitute for the term $+\Delta^2\lambda(\partial g(q)/\partial q)B(q,t)$,

$$\Delta^2\frac{\partial}{\partial q}g(q)B(q,t) = \dot P(q,t) + \frac{\partial}{\partial q}f(q)P(q,t). \tag{2.18}$$

The other term $-\Delta^2(\partial g(q)/\partial q)(\partial f(q)/\partial q)B(q,t)$ needs a careful analysis. By means of the commutation of the q-derivatives, this term is expressed as (Note: from now on an upper point means a partial time-derivative and a comma means a q-derivative):

$$\begin{aligned}-\Delta^2 \frac{\partial}{\partial q} g(q) \frac{\partial}{\partial q} f(q) B(q,t) &= -\Delta^2 \frac{\partial}{\partial t}(g(q)f'(q) - g'(q)f(q))B(q,t) \\ &\quad -\Delta^2 \frac{\partial}{\partial q} f(q) \frac{\partial}{\partial q} g(q) B(q,t),\end{aligned} \tag{2.19}$$

where the last term can be written in terms of $P(q,t)$ by means of (2.13) so that

$$\begin{aligned}-\Delta^2 \frac{\partial}{\partial q} f(g) \frac{\partial}{\partial q} g(q) B(q,t) &= -\frac{\partial}{\partial q} f(q) \dot{P}(q,t) - \frac{\partial}{\partial q} f(q) \frac{\partial}{\partial q} f(g) P(q,t).\end{aligned} \tag{2.20}$$

Joining all these partial results, the general equation (2.16) can be written in the following form:

$$\begin{aligned}\ddot{P}(q,t) = &-\lambda \dot{P}(q,t) - 2\frac{\partial}{\partial q} f(q) \dot{P}(q,t) \\ &-\lambda \frac{\partial}{\partial q} f(q) P(q,t) \\ &+\Delta^2 \frac{\partial}{\partial q} g(q) \frac{\partial}{\partial q} g(q) P(q,t) \\ &-\frac{\partial}{\partial q} f(q) \frac{\partial}{\partial q} f(q) P(q,t) \\ &-\frac{\partial}{\partial q}(g(q)f'(q) - g'(q)f(q))\Delta^2 B(q,t).\end{aligned} \tag{2.21}$$

As far as solvability is concerned, this formal equation is equivalent to (2.16). In order to advance in this way, we need to eliminate $B(q,t)$ in the last term in (2.21). This can be done in the case that

$$g(q)f'(q) - g'(q)f(q) = g^2(q)(f(q)/g(q))' = Cg(q), \tag{2.22}$$

where C is a constant. If this condition holds, the last term in (2.21) is $-\partial(Cg(q)B(q,t))/\partial q$ and by means of (2.13), it is transformed into

$$\begin{aligned}-C\frac{\partial}{\partial q} g(q)\Delta^2 B(q,t) &= -CP(q,t) - C\frac{\partial}{\partial q} f(q) P(q,t),\end{aligned} \tag{2.23}$$

and hence (2.21):

$$\begin{aligned}\ddot{P}(q,t) = &-(\lambda + C)\dot{P}(q,t) - 2\frac{\partial}{\partial q}(q)\dot{P}(q,t) \\ &-(\lambda + C)\frac{\partial}{\partial q} f(q) P(q,t) \\ &+\Delta^2 \frac{\partial}{\partial q} g(q) \frac{\partial}{\partial q} g(q) P(q,t) \\ &-\frac{\partial}{\partial q} f(q) \frac{\partial}{\partial q} f(q) P(q,t),\end{aligned} \tag{2.24}$$

which is a second-order partial differential equation for the probability density of the process (2.1). This equation is one of the main results of this paper and it has a time-dependent exact solution with the initial conditions (2.10) and (2.11). For this reason, those processes (2.1) obeying (2.22) will be called exactly solvable models and not surprisingly the condition (2.22) is the same one that was discussed by Hänggi[16] and San Miguel[17] in the context of the white noise hypothesis for the stochastic force $\xi(t)$ in (2.1) and by Sancho and San Miguel[5] in the case of a Gaussian but a nonwhite assumption for $\xi(t)$.

In all these cases, the equation for the probability density was of first order in time (Fokker–Planck equation) with a linear drift and a constant diffusion whose solution is well known. In our case, there are higher time derivatives and the exact solution is, as of now, unknown. The exact solution of (2.24) will be another important result of this paper.

Before starting with the process of solving (2.24), we will write it in the standard form of second-order partial differential equations. This is done by commuting the q-derivatives

$$\begin{aligned}\Big[\frac{\partial}{\partial t^2} &+ (f^2(q) - \Delta^2 g(q))\frac{\partial^2}{\partial q^2} \\ &+ 2f(q)\frac{\partial^2}{\partial t \partial q} + (\lambda + C + 2f'(q))\frac{\partial}{\partial t} \\ &+ ((\lambda + C)f(q) - 3\Delta^2 g(q)g'(q) + 3f(q)f'(q))\frac{\partial}{\partial q} \\ &+ ((f(q)f'(q))' - \Delta^2(g(q)g'(q))' \\ &+ (\lambda + C)f'(q))\Big] P(q,t).\end{aligned} \tag{2.25}$$

In order to classify this partial differential equation, we need to evaluate the discriminant, which is

$$(\Delta^2 g^2(q))^{1/2} = \Delta g(q) \geqslant 0 \tag{2.26}$$

because $g(q)$ should always be positive. Equation(2.25) is classified as an hyperbolic second-order partial differential equation for whose solution we are going to follow the current studies on this mathematical topic.

3. EXAMPLES

If our process (2.1) obeys the necessary and sufficient condition (2.22) to be exactly solvable, we define a new variable $Q(q(t))$,[5,16,17]

$$Q(q) = \int^q \frac{dq}{g(q)}, \tag{3.1}$$

and we have

$$\dot{Q} = f(q)/g(q) + \xi(t) = CQ + A + \xi(t), \tag{3.2}$$

where we have used (2.22) in the integrated form and A is an irrelevant integration constant. We can assume A equal to zero. Our problem has been reduced to a linear one with additive noise. This is well known[5,16,17] in the context of soluble cases.

Equation (3.2) with $A = 0$ presents two possible and different versions: C equal to zero or not. These two cases will be called the pure diffusive case and the linear case, respectively.

A. Pure diffusive case

This case corresponds to the equation of motion (2.1) with $f(q) = 0$. The representative model can be written, after performing the changes (3.1) and relabeling the variable as

$$\dot{q}(t) = \xi(t). \tag{3.3}$$

Some exact results can be obtained without solving the corresponding equation of motion for $P(q,t)$. For example, the statistical averages are obtained as follows: The solution of (3.3) is

$$q(t) = q(0) + \int_0^t \xi(t')dt', \tag{3.4}$$

and hence the first two moments of the variable q are

$$\langle q(t)\rangle = \langle q(0)\rangle_{\rm IC} = 0, \tag{3.5a}$$

$$\langle q^2(t)\rangle - \langle q^2(0)\rangle_{\rm IC} = \int_0^t \xi(s)ds \int_0^t \xi(s')ds' = \frac{2\Delta^2}{\lambda}\left(t - \frac{(1-e^{-\lambda t})}{\lambda}\right). \tag{3.5b}$$

They are expressed in terms of the initial averages and we have used the statistical properties of $\xi(t)$ in (2.2). These results coincide with the well-known ones for the position of a free Brownian particle. We can obtain more interesting information about the system obeying (3.3). The knowledge of the time-dependent probability density $P(q,t)$ shows interesting behavior very different from that in the white noise limit.

From the general expression (2.25), particularized to the model (3.3), $f(q) = 0$, $g(q) = 1$, we obtain the equation of motion for $P(q,t)$:

$$\frac{\partial^2}{\partial t^2}P(q,t) - \Delta^2\frac{\partial^2}{\partial q^2}P(q,t) + \lambda\frac{\partial}{\partial t}P(q,t) = 0. \tag{3.6}$$

The initial conditions (2.10) and (2.11) become in this case

$$P(q,t)|_{t=0} = \delta(q), \tag{3.7}$$

$$\left.\frac{\partial P(q,t)}{\partial t}\right|_{t=0} = 0. \tag{3.8}$$

A similar equation to (3.6) with the initial conditions (3.7) and (3.8) appeared in the context of the generalized Smoluchowski diffusion equations[18,19] and the present example was solved by Hemmer.[20] Therefore, we will not reproduce the details. The probability density is

$$P(q,t) = \tfrac{1}{2}e^{-\lambda t/2}[\delta(\Delta t - q) + \delta(\Delta t + q)] + \frac{\lambda}{2\Delta}I_0\left(\frac{\lambda}{4\Delta}(\Delta^2t^2 - q^2)^{1/2}\right) + \frac{\lambda t}{2(\Delta^2t^2 - q^2)^{1/2}}I_1\left(\frac{\lambda}{2\Delta}(\Delta^2t^2 - q^2)^{1/2}\right), \tag{3.9}$$

where I_0, I_1 are the Bessel functions of the imaginary arguments of orders 0,1, respectively.

From (3.9) one can see that $P(q,t)$ is almost a flat distribution bounded by two delta functions moving to $q = \pm\infty$ with a velocity $\pm\Delta$, respectively. This shape is very different from that corresponding to the white noise case,[20] which presents a Gaussian distribution spreading out in time.

B. Linear case

In this case, Eq. (2.1) takes the general form (3.2), and after relabeling the variables, it is expressed as

$$\dot q = -\gamma q + \xi(t). \tag{3.10}$$

As in the former case, some interesting results can be obtained using the formal solution of (3.10). This is

$$q(t) = q(0)e^{-\gamma t} + \int_0^t e^{-\gamma(t-t')}\xi(t')dt', \tag{3.11}$$

so the mean value is

$$\langle q(t)\rangle = \langle q(0)\rangle_{\rm IC}e^{-\gamma t}, \tag{3.12}$$

which goes to zero for $t\to\infty$.

An interesting dynamical quantity is the correlation function

$$\langle q(t)q(t')\rangle. \tag{3.13}$$

Using the solution (3.11), the statistical properties of $\xi(t)$ in (2.2), and assuming statistical independence between the initial conditions and $\xi(t)$, with $q(0) = 0$, the quantity (3.13) is

$$\langle q(t)q(t')\rangle = \frac{2\Delta^2}{\lambda+\gamma}\left(\frac{1}{\gamma-\lambda} - \frac{1}{2\gamma}\right)e^{-\gamma(t+t')} - \frac{\Delta^2}{\gamma^2-\lambda^2}e^{-\gamma t-\lambda t'} - \frac{\Delta^2}{\gamma^2-\lambda^2}e^{-\gamma t'-\lambda t} + \left(\frac{\Delta^2}{(\lambda+\gamma)\gamma} - \frac{\Delta^2}{\gamma^2-\lambda^2}\right)e^{-\gamma(t-t')} + \frac{\Delta^2}{\gamma^2-\lambda^2}e^{-\lambda(t-t')}, \tag{3.14}$$

and in the stationary state $(t, t'\to\infty)$ but $t - t'$ finite,

$$\langle q(t)q(t')\rangle_{\rm st} = -\frac{\lambda\Delta^2}{\gamma(\gamma^2-\lambda^2)}e^{-\gamma(t-t')} + \frac{\Delta^2}{\gamma^2-\lambda^2}e^{-\lambda(t-t')}. \tag{3.15}$$

The equal-time correlation function in the stationary state is

$$\langle q^2\rangle_{\rm st} = \Delta^2/\gamma(\lambda+\gamma). \tag{3.16}$$

From the exact solution (3.15), we can obtain the linear relaxation time and see if critical slowing down exists at any point:

$$\tau = \int_0^\infty \frac{\langle q(t)q(t+t')\rangle_{\rm st}}{\langle q^2(t)\rangle_{\rm st}}dt' = \frac{\lambda+\gamma}{\lambda\gamma}. \tag{3.17}$$

We can see that only in the cases $\gamma\to0$ or $\lambda\to0$, the linear relaxation time diverges. In both cases, the formal stationary probability density is not normalizable, as we will see. The first case $(\gamma\to0)$ is a trivial one because the dissipative drift disappears and the problem reduces to the case 3A. The second case $(\lambda\to0)$ is new and more interesting because even with nonzero dissipative drift, the "color" λ and not the intensity Δ of the noise $\xi(t)$ precludes the existence of a stationary state. This corresponds to having a noise with infinite correlation time, and hence in the opposite limit of white noise.

The stationary distribution can be obtained from (2.14):

$$P_{\rm st}(q) = N(\Delta^2 - \gamma^2q^2)^{\lambda/2\gamma-1}, \tag{3.18}$$

defined between the boundaries $q = \pm\Delta/\gamma$. The normalization constant is

$$N = \frac{\gamma\Gamma(\frac{1}{2}+\lambda/2\gamma)}{\Delta^{\lambda/\gamma-1}\Gamma(\frac{1}{2})\Gamma(\lambda/2\gamma)}. \tag{3.19}$$

One can see that the correlation function in the stationary state evaluated with (3.18) and (3.19) agrees with (3.16).

The study of (3.18) manifests two different shapes for $P_{\rm st}(q)$, which we relate to a nonequilibrium phase transition.

The critical value of the parameters is $\lambda = 2\gamma$. In the case $\lambda > 2\gamma$, $P_{st}(q)$ has a maximum in $q = 0$ and goes to zero in the boundaries. For $\lambda < 2\gamma$, $P_{st}(q)$ has a minimum in $q = 0$ and goes to infinity at the boundaries. This gives rise to a bimodal distribution when we change the "color" of the noise λ from $\lambda > 2\gamma$ to $\lambda < 2\gamma$. Then we have found a nonequilibrium phase transition induced only by the "color" of the noise. The transition takes place at the value $\lambda = 2\gamma$. We have seen in (3.17) that no critical slowing down appears in this point but in $\lambda \to 0$. This is a very interesting example of Suzuki's theory,[8] which states that in nonequilibrium phase transitions, the phenomena of critical slowing down and the appearance of new maxima in $P_{st}(q)$ are separate processes which can take place for different values of the parameters.

We should mention that in the case $\lambda \to \infty$, we recover the white noise $P_{st}(q)$ corresponding to this problem. Let us now return to the problem of finding the solution of $P(q,t)$, which in this case is not known. As the details of the mathematical process are very cumbersome, we will state only the main steps and references used in the evaluation.

In the model (3.10), where $f(q) = -\gamma q$, $g(q) = 1$, the equation of motion (2.25) for the probability density is

$$\left[\frac{\partial^2}{\partial t^2} + (\gamma^2 q^2 - \Delta^2)\frac{\partial^2}{\partial q^2} - 2\gamma q \frac{\partial^2}{\partial t \partial q} - (3\gamma - \lambda)\frac{\partial}{\partial t} - \gamma(\lambda - 4\gamma) q \times \frac{\partial}{\partial q} - \gamma(\lambda - 2\gamma)\right] P(q,t) = 0 \tag{3.20}$$

with the corresponding initial conditions (2.10) and (2.11):

$$P(q,t)\big|_{t=0} = \delta(q), \tag{3.21}$$

$$\left(\frac{\partial}{\partial t} - \gamma q \frac{\partial}{\partial q} - \gamma\right) P(q,t)\bigg|_{t=0} = 0. \tag{3.22}$$

The standard approach in the solution of $(3.20)^{21}$ begins with its reduction to the canonical form. This is done by means of a change of variables

$$\xi = e^{\gamma t}(\gamma q - \Delta), \tag{3.23}$$

$$\eta = e^{\gamma t}(\gamma q + \Delta), \tag{3.24}$$

where the new variables ξ,η are called the characteristics. The partial differential equation for $P(\xi,\eta)$ takes the hyperbolic form

$$\left[(\eta - \xi)^2 \frac{\partial^2}{\partial \xi \partial \eta} - \alpha(\eta - \xi)\frac{\partial}{\partial \xi} + \alpha(\eta - \xi)\frac{\partial}{\partial \eta} - 2\alpha\right] P(\xi,\eta) = 0, \tag{3.25}$$

where

$$\alpha = 1 - \lambda/2\gamma. \tag{3.26}$$

The equation (3.25) has been studied by Koshlyakov *et al.*[21] in several cases. We follow their approach. The next step is to transform (3.25) into the Euler–Darbouse equation by the change

$$P(\xi,\eta) = (\eta - \xi)^\beta Q(\xi,\eta). \tag{3.27}$$

$Q(\xi,\eta)$ obeys the partial differential equation

$$\left[\frac{\partial^2}{\partial \xi \partial \eta} + \frac{\alpha}{\eta - \xi}\frac{\partial}{\partial \xi} - \frac{\alpha}{\eta - \xi}\frac{\partial}{\partial \eta}\right] Q(\xi,\eta) = 0 \tag{3.28}$$

with

$$\beta = 2\alpha. \tag{3.29}$$

The solution of the Cauchy problem associated with (3.28) with the corresponding boundary conditions given by (3.21) and (3.22), following the Rieman method, is indicated in Ref. 21. This gives, after transforming variables,

$$P(q,t) = P_1(q,t) + P_2(q,t) + P_3(q,t), \tag{3.30}$$

where

$$P_1(q,t) = \tfrac{1}{2}\gamma e^{(\alpha-1)\gamma t}\{\delta(\Delta(1 - e^{-\gamma t}) + \gamma q) + \delta(\gamma q - \Delta(q - e^{-\gamma t}))\}, \tag{3.31}$$

$$P_2(q,t) = \gamma(1 - 2\alpha)(2\Delta)^{2\alpha-1}(\Delta^2(1 + e^{-\gamma t})^2 - \gamma^2 q^2)^{-\alpha} F(\alpha,\alpha,1;\sigma), \tag{3.32}$$

$$\begin{aligned} P_3(q,t) = {} & -\tfrac{1}{2}\gamma(2\Delta)^{2\alpha+1} e^{-\gamma t}(\gamma q - \Delta(1 + e^{-\gamma t}))^{-1}(\Delta^2(1 + e^{-\gamma t})^2 - \gamma^2 q^2)^{-\alpha} F(\alpha,\alpha,1;\sigma) \\ & + \tfrac{1}{2}\gamma(2\Delta)^{2\alpha+1} e^{-\gamma t}(\Delta^2(1 + e^{-\gamma t})^2 - \gamma^2 q^2)^{-\alpha} F'(\alpha,\alpha,1;\sigma)(\gamma^2 g^2 - \Delta^2(1 + e^{-\gamma t})^2)^{-1} \\ & \times\{(\gamma q + \Delta(1 - e^{-\gamma t}))(\gamma q + \Delta(1 + e^{-\gamma t}))^{-1} + (\gamma q - \Delta(1 - e^{-\gamma t}))(\gamma q - \Delta(1 + e^{-\gamma t}))^{-1}\}, \end{aligned} \tag{3.33}$$

$$\sigma = (\Delta^2(1 - e^{-\gamma t})^2 - \gamma^2 q^2)(\Delta^2(1 + e^{-\gamma t})^2 - \gamma^2 q^2)^{-1}, \tag{3.34}$$

$$|q| \leqslant (\Delta/\gamma)(1 \pm e^{-\gamma t}). \tag{3.35}$$

$F(\alpha,\alpha,1;\sigma)$ and $F'(\alpha,\alpha,1;\sigma)$ are the hypergeometric function and its derivative with respect to σ.

$P_1(q,t)$ gives the behavior of $P(q,t)$ in the initial regime as one can see taking the limit $t \to 0$. $P_2(q,t)$ dominates in the limit $t \to \infty$, giving the stationary solution $P_{st}(q)$, which coincides with (3.18) and (3.19). $P_3(q,t)$ refers to the intermediate regime.

Although the solution(3.30)–(3.35) has its own importance because of its existence, only a few results can be obtained from it because of its extraordinary complexity. This shows that the non-Gaussianity of the process (3.10) manifests itself by the complexity of the solution, even in the case that we have a linear problem. As in the former case, we have that $P(q,t)$ is bounded by two delta functions moving to the stationary boundaries $\pm\Delta/\gamma$, and following a deterministic equation given by

$$\dot{q}_\pm = -\gamma q_\pm \pm \Delta \quad \text{or} \quad \dot{q}_\pm = f(q_\pm) \pm g(q_\pm)\Delta, \tag{3.36}$$

where $q_{\pm}$ is the peak position of the delta functions. This is easily understood if we think that the stochastic dichotomous process has only two possible values $\pm\Delta$. This is a characteristic of this kind of stochastic process modeled by this special noise, which has also been found in nonlinear equations by means of numerical simulation, and whose results will be presented elsewhere.

4. CONCLUSIONS

We have explored the possibility of obtaining dynamical properties for a Langevin-like equation of motion with dichotomous Markov noise. We have presented a general method to obtain the differential equation obeyed by the probability density of the process. This differential equation involves higher time derivatives and hence it is not of the Fokker–Planck type. The main point in this deduction is to consider the correlation time of the noise as an expansion parameter. In two particular cases, we have been able to write an exact differential equation which is a second-order partial differential equation of the hyperbolic type. In these two cases, we have found the exact solution of $P(q,t)$. In the second case, which has a nontrivial stationary state, we have studied the possibility of the appearance of critical slowing down by means of the explicit evaluation of the correlation time. Although the stationary analysis gives the existence of a phase transition for some value of the noise parameters, no critical slowing down appears in this point but in another one. This is an example of Suzuki's criteria of the appearance of slowing down in nonequilibrium stochastic processes.

ACKNOWLEDGMENT

I am indebted to Dr. M. San Miguel for his valuable comments.

[1]S. Kabashima, S. Kogure, S. Kawakubo, and T. Okada, J. Appl. Phys. **50**, 6296 (1979).
[2]S. Kai, T. Kai, M. Takata, and K. Hirakawa, J. Phys. Soc. Jpn. **47**, 1379 (1979).
[3]R. L. Stratonovich, *Topics in the Theory of Random Noise*, Vol. 1 (Gordon and Breach, New York, 1963).
[4]A. Brissaud and U. Frisch, J. Math. Phys. **15**, 524 (1974).
[5]J. M. Sancho and M. San Miguel, Z. Phys. B **36**, 357 (1980).
[6]M. San Miguel and J. M. Sancho, Phys. Lett. A **76**, 97 (1980).
[7]W. Horsthemke and R. Lefever, Z. Phys. B **40**, 241 (1980).
[8]M. Suzuki, K. Kaneko, and F. Sasagawa, Progr. Theoret. Phys. **65**, 828 (1981).
[9]N. G. Van Kampen, Phys. Rep. **24**, 171 (1976).
[10]K. Kitahara, W. Horsthemke, and R. Lefever, Phys. Lett. A **70**, 377 (1979); K.Kitahara, W. Horsthemke, R. Lefever, and Y. Inaba, Progr. Theoret. Phys. **64**, 1233 (1980).
[11]J. M. Sancho and M. San Miguel, Progr. Theoret. Phys. (1983) (to be published).
[12]M. Suzuki, in *Systems far from Equilibrium*, edited by L. Garrido, *Lecture Notes in Physics*, Vol. 132 (Springer-Verlag, Berlin, 1980); K. Kitahara and K. Ishii, "Relaxation of systems under the influence of two level Markovian noise," oral communication in *Statphys* 14 (Edmonton, Canada, 1981); C. Van den Broeck, Phys. Lett. A **91**, 399 (1982).
[13]R. C. Bourret, U. Frisch, and A. Pouquet, Physica **65**, 303 (1973).
[14]V. E. Shapiro and W. M. Loginov, Physica A **91**, 563 (1978).
[15]R. Kubo, J. Math. Phys. **4**, 174 (1963); M. Suzuki, Progr. Theoret. Phys. Suppl. **69**, 169 (1980).
[16]P. Hänggi, Z. Phys. B **30**, 85 (1978).
[17]M. San Miguel, Z. Phys. B **33**, 307 (1979).
[18]H. C. Brinkman, Physica **22**, 29 (1956).
[19]R. A. Sack, Physica **22**, 917 (1956).
[20]P. Chr. Hemmer, Physica **27**, 79 (1961).
[21]N. S. Koshlyakov, M. M. Smirnov, and R. R. Gliner, in *Differential Equations of Mathematical Physics* (North-Holland, Amsterdam, 1964).

A spin-$\frac{1}{2}$ particle formalism in curved space–time

Mario A. Castagnino
Instituto de Física de Rosario, Av. Pellegrini 250, 2000 Rosario, Argentina
Luis Chimento
Departamento de Física, Universidad de Buenos Aires, Ciudad Universitaria, Pab. I, 1428 Buenos Aires, Argentina
Diego D. Harari and Carmen A. Núñez
Instituto de Astronomía y Física del Espacio, Casilla de Correo 67, Sucursal 28, 1428 Buenos Aires, Argentina

(Received 27 August 1982; accepted for publication 28 January 1983)

A formalism is presented to solve the Dirac equation in curved space–time and the spinorial solution is obtained. A general method is introduced to compute the particle creation. The kernel $S_1(x,x')$ leading to the particle model is built out of considerations on the Minkowskian limit and high energy behavior.

PACS numbers: 11.10.Qr, 04.90. + e

I. INTRODUCTION

The introduction of a particle language in a curved space–time semiclassical quantum field theory could help to solve important problems and to get interesting cosmological consequences, as was already discussed in a previous paper.[1] The so-called quantum equivalence principle (QEP)[2] was applied there to quantize a scalar field in the presence of a classical gravitational field; a scalar particle model was explicitly obtained, and a finite number of particles created during the expansion of the universe was computed. A Robertson–Walker spatially flat metric was considered, and the calculations were performed up to second order in a power series development of the Hubble coefficient (H).

The same principle (QEP) was extended to the Dirac case by one of us (Castagnino[3]), and it was again proved to be implementable when developed up to first order in H. In fact, an analogy can be made between the ambiguity in the determination of the biscalar kernel $G_1^{(0)}(x,x')$ of the generalized Klein–Gordon operator and the one of the bispinorial distribution $S_1(x,x')$, which "contains" the spin-$\frac{1}{2}$ particle–antiparticle model as was shown in Ref. 3. The properties that are naturally expected to be satisfied by $S_1(x,x')$ (generalization of the flat space–time ones) are not enough to univocally determine it. However, if the QEP is used as the selective criterion, it leads to inconsistencies when higher orders of H are taken into account. Therefore, another additional condition has to be used as was already pointed out in Ref. 4, where the scalar case was considered. The selecting criterion introduced there is based on the argument that the high energy behavior of a field theory, which is governed by the singular structure of the kernels, should resemble the flat space–time one. In fact, the identification between flat and curved space–time kernels over all a Cauchy surface (as was assumed by the QEP) could be an excessively strong requirement. The additional condition introduced in Ref. 4 completely defines a scalar particle–antiparticle model when it is developed in a spatially flat expanding universe in a power series of the metric and its derivatives up to second order.

We now show that the method can be extended to the Dirac case and that the kernel $S_1(x,x')$ we obtain is the most natural generalization of the flat space–time one.

In Sec. II A a general formalism is presented to solve the Dirac equation in curved space–time, and the spinorial solution is given.

A general method to compute the particle production between two states of the universe (i.e., between two different times t_1 and t_2) is displayed in Sec. II B.

The kernel $S_1(x,x')$ built in Sec. III leads to no particle creation. However, it could contain terms that cannot be "caught" by a series development (such as those having the form e^{-k}, when the expansion is made in powers of k^{-1}) but which also satisfy the minimal hypothesis needed to integrate a good projector, and which could give rise to a nonzero number of created particles.

II. SPINORIAL PARTICLE MODEL

A. Dirac equation solutions in curved space–time

A spinor field in curved space–time should satisfy the generalization of Dirac equation, namely:

$$[\gamma^i(x)\nabla_i + m]\Psi(x) = 0, \tag{1}$$

where $i = 0,1,2,3$, $\gamma^i(x)$ are the generalized Dirac matrices verifying the anticommutation rule

$$\{\gamma^i,\gamma^j\} = -2g^{ij}I \tag{2}$$

(I = identity 4×4 matrix), $\nabla_i = \partial_i \pm \sigma_i$ denotes the covariant ($-$) (contravariant ($+$)) spinor derivative, being the spinorial connection partially defined by $\nabla_i\gamma_j = 0$, and m is the mass of the field. For more details see Refs. 3 and 5.

In a spatially flat Robertson–Walker universe characterized by the following space–time interval,

$$ds^2 = dt^2 - a(t)^2\delta_{\alpha\beta}\,dx^\alpha dx^\beta, \tag{3}$$

Eq. (1) takes the form

$$(\gamma^i\partial_i + \tfrac{3}{2}H\gamma^0 - m)\Psi(x) = 0. \tag{4}$$

The spinorial affine connection and Dirac matrices are explicitly listed in the Appendix.

Equation (4) can be solved by separation of variables. The following spinor can be given as a general solution:

$$\Psi_{\mathbf{k}}(\mathbf{r},t) = [1/(2\pi a)^{3/2}]\Phi_{\mathbf{k}}(t)e^{-i\mathbf{k}\cdot\mathbf{r}}. \tag{5}$$

If (5) is replaced in (4) and the commutation rules (2) are used, Eq. (4) reduces to

$$[\partial_0 + \gamma^0(ik_\alpha\gamma^\alpha + m)]\Phi_{\mathbf{k}} = 0. \tag{6}$$

Now the spinor $\Phi_{\mathbf{k}}$ could *a priori* be any function of time satisfying (6). However, it is well known that every solution of the Dirac equation is also a solution of the Klein–Gordon equation with a d'Alembert operator defined by $\Delta^{(1/2)} = -\nabla_i\nabla^i + \frac{1}{4}R$, where $R = g^{ij}R_{ij} = g^{ij}R^{h}{}_{ihj}$. Due to the particular representation of the Dirac matrices we have chosen when $\Delta^{(1/2)}$ is applied to $\Phi_{\mathbf{k}}$, two independent differential equations arise: One should be satisfied by the first two components of $\Phi_{\mathbf{k}}$ and the other by the remaining two. Then we suggest

$$\Phi_{\mathbf{k}} = \begin{pmatrix} A_{1\mathbf{k}}e^{i\int\Omega_k dt} \\ A_{2\mathbf{k}}e^{i\int\Omega_k dt} \\ A_{3\mathbf{k}}e^{i\int\Lambda_k dt} \\ A_{4\mathbf{k}}e^{i\int\Lambda_k dt} \end{pmatrix}, \tag{7}$$

where $A_{1\mathbf{k}},\ldots,A_{4\mathbf{k}}$ are coefficients depending only on the momentum $\mathbf{k}$ and Ω_k and Λ_k are arbitrary complex functions of time depending on the momentum modulus. It can also be pointed out that the coefficients $A_{1\mathbf{k}},\ldots,A_{4\mathbf{k}}$ are those appearing in the spinorial solution of the flat space–time Dirac equation.

Replacing (7) in (6), we obtain a homogeneous system which has a nontrivial solution if

$$(\Omega_k + m)(\Lambda_k - m) = k^2/a^2. \tag{8}$$

When this condition is satisfied, two independent spinors $\Phi_{\mathbf{k}}$ are obtained. However, as a complete base of solutions of the Dirac equation must consist of four spinors, two more spinors have to be found. In fact, note that the determinant (8) is not modified under the following change:

$$\Omega_k \to -\Lambda_k^*, \quad \Lambda_k \to -\Omega_k^*. \tag{9}$$

Therefore, when this change is introduced in (7) and the resultant expression is replaced in (6), the other two spinors are obtained.

The base can then be written as

$$\begin{aligned}\Psi_{\mathbf{k}}^{(1),(2)} &= \left(\frac{\Omega_k^* + m}{\Omega_k^* + \Lambda_k}\right)^{1/2} \times \begin{pmatrix} I \\ \dfrac{k_\alpha\sigma^\alpha}{a(\Omega_k^* + m)} \end{pmatrix} \frac{e^{-i\int\lambda_k dt}}{(2\pi a)^{3/2}} e^{-i\mathbf{k}\cdot\mathbf{r}}, \\ \Psi_{\mathbf{k}}^{(3),(4)} &= \left(\frac{\Omega_k^* + m}{\Omega_k^* + \Lambda_k}\right)^{1/2} \times \begin{pmatrix} \dfrac{-k_\alpha\sigma^\alpha}{a(\Omega_k^* + m)} \\ I \end{pmatrix} \frac{e^{i\int\lambda_k dt}}{(2\pi a)^{3/2}} e^{-i\mathbf{k}\cdot\mathbf{r}},\end{aligned} \tag{10}$$

where $\lambda_k = \mathrm{Re}(\Lambda_k)$.

These spinors have been normalized according to the following internal product (see Ref. 3):

$$\langle\Psi_{\mathbf{k}}^{(i)},\Psi_{\mathbf{k}'}^{(l)}\rangle = -i\int_\Sigma \bar{\Psi}_{\mathbf{k}}^{(i)}\gamma_k\Psi_{\mathbf{k}'}^{(l)}\,d\sigma^k = \delta^{il}\delta(\mathbf{k}-\mathbf{k}'), \tag{11}$$

where $\bar{\Psi}_{\mathbf{k}}$ means Dirac adjoint ($\bar{\Psi}_{\mathbf{k}} = \Psi_{\mathbf{k}}^\dagger\beta; \beta = -i\gamma_0$). It is clearly seen that the integral (11) does not depend on the surface Σ, where it is performed.

However, spinors (10) represent a formal solution. They will not be a solution of the Dirac equation unless Ω_k and Λ_k satisfy the following differential equations obtained replacing the spinors $\Phi_{\mathbf{k}}$ in (6):

$$i\dot{\Omega}_k - \Omega_k^2 + iH(\Omega_k + m) + \omega_{\mathbf{k}}^2 = 0, \tag{12}$$

$$i\dot{\Lambda}_k - \Lambda_k^2 + iH(\Lambda_k - m) + \omega_{\mathbf{k}}^2 = 0, \tag{13}$$

being

$$\omega_{\mathbf{k}}^2 = k^2/a^2 + m^2. \tag{14}$$

It can easily be seen that Eqs. (12)–(13) admit a simple solution when the massless case is considered. In fact, it is

$$\Omega_k = \pm k/a, \quad \Lambda_k = \pm k/a. \tag{15}$$

However, there is in general no simple solution for massive particles. Nevertheless, it is possible to develop a solution of (12)–(13) in a power series of the Hubble coefficient H (see the Appendix).

There are therefore two positive and two negative frequency solutions which will be identified with particles and antiparticles with positive or negative helicity when the corresponding kernels were found. Indeed a base $\Phi_{\mathbf{k}}^{(a)}$ ($a = 1,2,3,4$) there also exists in flat space–time where these four functions have a precise physical meaning: Two of them are the particle model with positive or negative helicity, and the other two constitute the antiparticle model with positive or negative helicity. Instead we have a set of completely equivalent base spinors $\{\Phi_{\mathbf{k}}^{(a)}\}$, and we do not have a criterion to decide, within that base, which vectors correspond to particles and which to antiparticles. In fact, the particle–antiparticle model is associated with the bispinorial kernels $S(x,x')$ and $S_1(x,x')$ (generalizations of their flat space–time analogs). While the kernel $S(x,x')$ is well defined in an expanding universe and in general in any curved space–time, the kernel $S_1(x,x')$ is not. It was in fact shown in Ref. 3 that $S_1(x,x')$ is not invariant under a change of the base of Dirac equation solutions. The fact that a different kernel $S_1(x,x')$ and thus a different particle–antiparticle model could be selected over different Cauchy surfaces can be interpreted as a phenomenon of particle creation. We therefore introduce in the next paragraph a formalism to compute the number of particles created during the expansion of the universe.

B. Particle creation

It will be useful to define a matrix $F_{\mathbf{k}}$ the columns of which are the spinors $\Psi_{\mathbf{k}}^{(a)}$ without the factor $e^{-i\mathbf{k}\cdot\mathbf{r}}/(2\pi a)^{3/2}$, which is not necessary to evaluate the creation of particles. It can be seen that $F_{\mathbf{k}}$ is unitary, i.e.,

$$F_{\mathbf{k}}^{-1} = F_{\mathbf{k}}^\dagger. \tag{16}$$

Any base defining the particle–antiparticle model over a Cauchy surface $\Sigma(\tau) = \{x/t(x) = \tau\}$ can be expressed as a linear combination of the spinors $\Psi_{\mathbf{k}}^{(a)}$ as

$$P_{\mathbf{k}}^{(\tau)}(t) = F_{\mathbf{k}}(t) A_{\mathbf{k}}^{(\tau)}, \tag{17}$$

where $A_{\mathbf{k}}$ is a 4×4 matrix. These $P_{\mathbf{k}}^{(\tau)}(t)$, the columns of which will be denoted by $\Pi_{\mathbf{k}}^{(\tau)(a)}(t)$, would be fixed with a, by now unknown, physical principle providing the Cauchy data $\Pi_{\mathbf{k}}^{(\tau)(a)}(\tau)$ (e.g., in Ref. 3 the QEP was used).

The coefficients $A_{\mathbf{k}}^{(\tau)}$ can be obtained as a function of the initial conditions $P_{\mathbf{k}}^{(\tau)}(\tau)$ on the Cauchy surface using Eq. (16)

$$A_{\mathbf{k}}^{(\tau)} = F_{\mathbf{k}}^{\dagger}(\tau) P_{\mathbf{k}}^{(\tau)}(\tau). \tag{18}$$

The particle–antiparticle models at two different times τ_1 and τ_2 can also be related by a linear transformation (Bogoliubov transformation)

$$P_{\mathbf{k}}^{(\tau_2)}(t) = P_{\mathbf{k}}^{(\tau_1)}(t)\alpha_{\mathbf{k}}(\tau_1,\tau_2), \tag{19}$$

$\alpha_{\mathbf{k}}$ being the 4×4 matrix of the transformation, namely from (19) and, using (16) and (18),

$$\alpha_{\mathbf{k}}(\tau_1,\tau_2) = P_{\mathbf{k}}^{(\tau_1)\dagger}(\tau_1) F_{\mathbf{k}}(\tau_1) F_{\mathbf{k}}^{\dagger}(\tau_2) P_{\mathbf{k}}^{(\tau_2)}(\tau_2) \tag{20}$$

as a function of the Cauchy data $P_{\mathbf{k}}^{(\tau_1)}(\tau_1)$ and $P_{\mathbf{k}}^{(\tau_2)}(\tau_2)$.

The matrix $\alpha_{\mathbf{k}}$ has two very important properties:

(i) It is unitary

$$\alpha_{\mathbf{k}}^{\dagger} = \alpha_{\mathbf{k}} \tag{21}$$

because it transforms the orthonormal base $P_{\mathbf{k}}^{(\tau_1)}$ in another base $P_{\mathbf{k}}^{(\tau_2)}$, which is also orthonormal. This property is directly obtained from (20).

(ii) If the Dirac equation (6) is charge conjugated and some properties of Dirac matrices γ_i are taken into account, it can be easily seen that $\gamma_2 \Phi^*_{-\mathbf{k}}$ is a solution of the Dirac equation if $\Phi_{\mathbf{k}}$ also is a solution. The charge conjugation operator C can then be defined as $C=\gamma_2$ plus complex conjugation and $\mathbf{k}\to -\mathbf{k}$, and after some calculation the following can be obtained:

$$\begin{aligned} &C F_{\mathbf{k}} C = F^*_{-\mathbf{k}},\\ &C P_{\mathbf{k}}^{(\tau)} C = P_{-\mathbf{k}}^{(\tau)*},\\ &C A_{\mathbf{k}}^{(\tau)} C = A_{-\mathbf{k}}^{(\tau)*},\\ &C \alpha_{\mathbf{k}} C = \alpha^*_{-\mathbf{k}}. \end{aligned} \tag{22}$$

On the other hand multiplying by $e^{-i\mathbf{k}\cdot\mathbf{r}}/(2\pi a)^{3/2}$ the first of Eq. (22) and using spinors (10), it can be seen that

$$C\Psi_{\mathbf{k}}^{(1)} = \Psi_{-\mathbf{k}}^{(4)*}, \quad C\Psi_{\mathbf{k}}^{(2)} = \Psi_{-\mathbf{k}}^{(3)*}. \tag{23}$$

Moreover, for any column of the matrices $P_{\mathbf{k}}, A_{\mathbf{k}}$, and $\alpha_{\mathbf{k}}$ an analogous relation holds. If it is developed for $\alpha_{\mathbf{k}}$, some relations are found between its elements which reduce the number of independent elements, namely

$$\begin{aligned} &(\alpha_{\mathbf{k}})_{41} = (\alpha^*_{-\mathbf{k}})_{14}, \quad (\alpha_{\mathbf{k}})_{42} = -(\alpha^*_{-\mathbf{k}})_{13},\\ &(\alpha_{\mathbf{k}})_{31} = -(\alpha^*_{-\mathbf{k}})_{24}, \quad (\alpha_{\mathbf{k}})_{32} = (\alpha^*_{-\mathbf{k}})_{23},\\ &(\alpha_{\mathbf{k}})_{21} = -(\alpha^*_{-\mathbf{k}})_{34}, \quad (\alpha_{\mathbf{k}})_{22} = (\alpha^*_{-\mathbf{k}})_{33},\\ &(\alpha_{\mathbf{k}})_{11} = (\alpha^*_{-\mathbf{k}})_{44}, \quad (\alpha_{\mathbf{k}})_{12} = -(\alpha^*_{-\mathbf{k}})_{43}. \end{aligned} \tag{24}$$

Now, the Dirac field must be quantized in order to compute the number of particles created between τ_1 and τ_2. As was already discussed, $P_{\mathbf{k}}^{(\tau_1)}$ and $P_{\mathbf{k}}^{(\tau_2)}$ are the bases which will represent the particle and antiparticle models at τ_1 and τ_2. The field can then be expressed in terms of any of these bases with the corresponding creation and annihilation operators over each Cauchy surface, i.e.,

$$\begin{aligned} \Psi_{\mathbf{k}}(t) &= \Pi_{\mathbf{k}}^{(\tau_1)(1)}(t)(a_{\mathbf{k}}^{(\tau_1)})_1 + \Pi_{\mathbf{k}}^{(\tau_1)(2)}(t)(a_{\mathbf{k}}^{(\tau_1)})_2\\ &\quad + \Pi_{\mathbf{k}}^{(\tau_1)(3)}(t)(a_{-\mathbf{k}}^{(\tau_1)})_3^{\dagger} + \Pi_{\mathbf{k}}^{(\tau_1)(4)}(t)(a_{-\mathbf{k}}^{(\tau_1)})_4^{\dagger},\\ \Psi_{\mathbf{k}}(t) &= \Pi_{\mathbf{k}}^{(\tau_2)(1)}(t)(a_{\mathbf{k}}^{(\tau_2)})_1 + \Pi_{\mathbf{k}}^{(\tau_2)(2)}(t)(a_{\mathbf{k}}^{(\tau_2)})_2\\ &\quad + \Pi_{\mathbf{k}}^{(\tau_2)(3)}(t)(a_{-\mathbf{k}}^{(\tau_2)})_3^{\dagger} + \Pi_{\mathbf{k}}^{(\tau_2)(4)}(t)(a_{-\mathbf{k}}^{(\tau_2)})_4^{\dagger}, \end{aligned} \tag{25}$$

where again the common factor $e^{-i\mathbf{k}\cdot\mathbf{r}}/(2\pi a)^{3/2}$ has been omitted and $(a_{\mathbf{k}})_{1,2}$ are the particle creation operators and $(a_{\mathbf{k}})_{3,4}$ are the antiparticle annihilation operators. To simplify the notation, the following column vector operator is introduced:

$$a_{\mathbf{k}}^{(\tau)} = \begin{pmatrix} (a_{\mathbf{k}}^{(\tau)})_1 \\ (a_{\mathbf{k}}^{(\tau)})_2 \\ (a_{-\mathbf{k}}^{(\tau)})_3^{\dagger} \\ (a_{-\mathbf{k}}^{(\tau)})_4^{\dagger} \end{pmatrix}. \tag{26}$$

Then Eq. (25) can be written in the following simpler form:

$$\begin{aligned} \Psi_{\mathbf{k}}(t) &= P_{\mathbf{k}}^{(\tau_1)}(t) a_{\mathbf{k}}^{(\tau_1)},\\ \Psi_{\mathbf{k}}(t) &= P_{\mathbf{k}}^{(\tau_2)}(t) a_{\mathbf{k}}^{(\tau_2)}. \end{aligned} \tag{27}$$

But as $P_{\mathbf{k}}^{(\tau_1)}$ and $P_{\mathbf{k}}^{(\tau_2)}$ are related by a Bogoliubov transformation replacing one base as a function of the other in (27), $a_{\mathbf{k}}$ transforms according to the following law:

$$a_{\mathbf{k}}^{(\tau_2)} = \alpha_{\mathbf{k}}^{\dagger}(\tau_1,\tau_2) a_{\mathbf{k}}^{(\tau_1)}. \tag{28}$$

Everything is now ready to compute the number of particles and antiparticles created by an expanding universe as it goes from τ_1 to τ_2. If the state of the universe at τ_1 is taken as the vacuum state, then

$$\begin{aligned} {}_1\langle 0|N_{1\mathbf{k}}^{(\tau_2)}|0\rangle_1 &= {}_1\langle 0|(a_{\mathbf{k}}^{(\tau_2)})_1 (a_{\mathbf{k}}^{(\tau_2)})_1|0\rangle_1\\ &= |(\alpha_{\mathbf{k}})_{31}|^2 + |(\alpha_{\mathbf{k}})_{41}|^2,\\ {}_1\langle 0|(N_{\mathbf{k}}^{(\tau_2)})_2|0\rangle_1 &= |(\alpha_{\mathbf{k}})_{42}|^2 + |(\alpha_{\mathbf{k}})_{32}|^2,\\ {}_1\langle 0|(N_{\mathbf{k}}^{(\tau_2)})_3|0\rangle_1 &= |(\alpha_{\mathbf{k}})_{42}|^2 + |(\alpha_{\mathbf{k}})_{32}|^2,\\ {}_1\langle 0|(N_{\mathbf{k}}^{(\tau_2)})_4|0\rangle_1 &= |(\alpha_{\mathbf{k}})_{41}|^2 + |(\alpha_{\mathbf{k}})_{31}|^2, \end{aligned} \tag{29}$$

where relations (24) have been used for the last two. It must be noticed that the creation of particles arises from the nondiagonal terms of $\alpha_{\mathbf{k}}$, i.e., the elements mixing the particle and antiparticle terms.

The formalism we have introduced is a generalization of that applied in Ref. 3 to compute the particle production when only the first order of H was considered. However, as the QEP is not implementable when the second order of H is taken into account, another criterion has to be given to select the particle–antiparticle base. We present, in the next section, an alternative method leading to the adiabatic particle model.

III. MINIMAL HYPOTHESES AND CONSTRUCTION OF THE MODEL

A. Minimal hypotheses

It is now necessary to explicitly find the base $P_{\mathbf{k}}^{(\tau)}$ giving rise to the particle–antiparticle model. As was seen in Ref. 3 this base is associated with the bispinorial kernel $S_1(x,x')$. So we now introduce a criterion to select one of the possible candidates to play that role.

We shall work in a generic globally hyperbolic manifold at the beginning of this section, though the formalism will be developed for a Robertson–Walker spatially flat expanding universe (3).

The generalization of the biscalar kernel $G^{(0)}(x,x')$ of the operator $(\Delta^{(0)} - m^2 - \xi R)$ to the spinorial case was done by Lichnerowicz,[6] who introduced the bispinorial distribution contravariant in x and covariant in x', $G^{(1/2)a}{}_{b'}(x,x')$, i.e., the elemental kernel of the operator $(\Delta^{(1/2)} - m^2)$.

The following notation has been used: $\Delta^{(0)} = -g^{ij}\nabla_i^{(0)}\partial_j$, where $\nabla_i^{(0)}$ denotes the ordinary convariant derivative of tensors; ξ defines the kind of coupling to the gravitational field ($\xi = 0$ minimal or $\xi = 1/6$ conformal); and $a,b' = 1,2,3,4$ are spinorial indices, which will be omitted from now on as it will be understood that the bispinors considered are contravariant in x and covariant in x'.

Operating on the bispinorial kernel $G^{(1/2)a}{}_{b'}(x,x')$ with the generalization of the flat space–time Dirac operator a new kernel is introduced, namely,

$$S(x,x') = (\gamma^i\nabla_i + m)G^{(1/2)}(x,x'), \tag{30}$$

which can also be expanded in terms of an orthonormal base of Dirac equation solutions $\{\chi_{\mathbf{k}}^{(s,h)}\}$ as

$$S(x,x') = i\sum_{s,h}\chi_{\mathbf{k}}^{(s,h)}(x)\bar{\chi}_{\mathbf{k}}^{(s,h)}(x'), \tag{31}$$

where

$$\chi_{\mathbf{k}}^{(s,h)} = \frac{1}{(2\pi a)^{3/2}}\Pi_{\mathbf{k}}^{(\tau)(s,h)}e^{-i\mathbf{k}\cdot\mathbf{r}} \tag{32}$$

(see Ref. 3), and we have now changed the four-valued indices (a) for a pair of two-valued indices (s,h) where $s = +(-)$, $h = +(-)$ correspond to particles (antiparticles) with positive (negative) helicity. $S(x,x')$ has all the properties to be considered the anticommutator of the field $[\{\Psi(x),\Psi(x')\} = -iS(x,x')]$.

Analogously the kernel $S_1(x,x')$ can be introduced as

$$S_1(x,x') = (\gamma^i\nabla_i + m)G_1^{(1/2)}(x,x'). \tag{33}$$

However, $S_1(x,x')$ is not invariant under an orthonormal base transformation as can be seen when it is expanded in terms of the base $\{\chi_{\mathbf{k}}^{(s,h)}\}$,

$$S_1(x,x') = \sum_{s,h} s\chi_{\mathbf{k}}^{(s,h)}(x)\bar{\chi}_{\mathbf{k}}^{(s,h)}(x') \tag{34}$$

(see Ref. 3). Therefore, the formalism stated in Ref. 4 to single out one of these possible biscalar kernels satisfying Lichnerowicz's conditions is now sketched and extended to the spinorial case. It will be used to select one of these $S_1(x,x')$ and, thus, the particle–antiparticle model.

The new condition introduced assumes that a different kernel could be associated with each point x_0 of space–time, which in this case will be called $S_1^{(x_0)}(x,x')$, due to the nonexistence of a global momentum space. However, if some symmetries hold all $S_1^{(x_0)}(x,x')$, for x_0 belonging to certain kind of surfaces (for example, over $\{t = \text{const}\}$ hypersurfaces in a Robertson–Walker spatially flat universe) will coincide (they will be called $S_1^{\Sigma(x_0)}$), and a particle definition will be possible. Two natural properties should be satisfied by these $S_1^{x_0}(x,x')$. They are: (i) Minkowskian limit; (ii) their singular structure must be the one of the flat space–time $S_1^{(\text{fl})}[s(x,x')]$ (s is being the geodetic interval between x and x'). This last property displays the fact that the particle model should reproduce the flat-space–time high-energy behavior.

We now use Eq. (33), and imposing $S_1(x,x')$ to be a solution of the Dirac equation shows that the analogy with the scalar case can go even further. In fact,

$$(\gamma^i\nabla_i - m)S_1(x,x') = (\Delta^{(1/2)} - m^2)G_1^{(1/2)}(x,x') = 0 \tag{35}$$

shows that $G_1^{(1/2)}(x,x')$ is obviously the generalization of the flat space–time $I\Delta_1[s(x,x')]$. Therefore, condition (ii) can be restated in these terms: $G_1^{(1/2)}(x,x')$ should reproduce the different dependences of $\Delta_1(s)$ on s. In fact, $\Delta_1[s(x,x')]$ as a function of the geodetic interval $s(x,x')$ is

$$\Delta_1(s) = (m^2/4\pi)\text{Im}\,[H_1^{(1)}(ms)/ms], \tag{36}$$

where $H_1^{(1)}$ is the first order and first-type Hankel function which when developed for small s has a quadratic divergence independent of mass, a logarithmic one, a constant term, and terms vanishing as s does. The different behaviors with respect to s can be identified with successive derivatives of Δ_1 with respect to m^2 (see DeWitt[7]), i.e., $\Delta_1(s)$ essentially has a quadratic divergence, $\partial\Delta_1/\partial m^2$ starts with a logarithmic one, $\partial^2\Delta_1/(\partial m^2)^2$ is regular, and so on. The generalized kernel $G_1^{(1/2)(x_0)}(x,x')$ should reproduce the flat space–time dependence of Δ_1 on s, but in principle any bispinorial regular function may appear as a multiplicative factor of each term provided it satisfies its Minkowskian limit.

(i) and (ii) can now be joined into

$$\lim_{\substack{x\to x_0\\ x'\to x_0}} G_1^{(1/2)}(x,x') = \lim_{\substack{x\to x_0\\ x'\to x_0}}\sum_{n=0}^{\infty}F_n^{(x_0)}(x,x')\frac{\partial^n\Delta_1(s)}{(\partial m^2)^n} \tag{37}$$

with

$$\left.\begin{aligned}F_0^{(x_0)}(x,x') &= 1\\ F_\alpha^{(x_0)}(x,x') &= 0\end{aligned}\right\}\quad \text{if } R_{ijkl} = 0 \quad \text{(flat space–time)},$$

$n = 0,1,2,\ldots$; $\alpha = 1,2,3,\ldots$, and it constitutes together with (34) the minimal hypotheses to define a "good" $S_1(x,x')$.

B. Construction of the model

The functions $F_n^{(x_0)}(x,x')$ are required to be real symmetric bispinors (see Ref. 3). As they and their derivatives will be evaluated in a power series development of the metric derivatives, general expressions for the coincidence limits have to be given. The curvature, the metric tensor, Dirac matrices, and the mass of the particles are the only candidates available for a covariant expression. The selection of the terms involved in the coincidence limits is based on dimensional and Minkowskian limit considerations. In atomic units ($\hbar = c = 1$) the mass has frequency dimensions, the scalar curvature R squared frequency dimensions, and expressions (A1) explicitly indicate the dimensions of Dirac matrices. There will be terms containing the curvature, but neither derivatives nor higher powers will be considered, as it will be explained below. It can then be written

$$\begin{aligned}
&\lim_{\substack{x\to x_0\\ x'\to x_0}} F_0^{(x_0)}(x,x') = I,\\
&\lim_{\substack{x\to x_0\\ x'\to x_0}} \nabla_i F_0^{(x_0)}(x,x') = 0,\\
&\lim_{\substack{x\to x_0\\ x'\to x_0}} F_\alpha^{(x_0)}(x,x') = (m^2)^{\alpha-1} A_\alpha R, \qquad (38)\\
&\lim_{\substack{x\to x_0\\ x'\to x_0}} \nabla_i F_\alpha^{(x_0)}(x,x') = (m^2)^{2\alpha-1}[B_\alpha R\gamma_i + C_\alpha R_{ij}\gamma^j],\\
&\lim_{\substack{x\to x_0\\ x'\to x_0}} \nabla_i\nabla_j F_n^{(x_0)}(x,x') = (m^2)^n[V_n R_{ij} + W_n R g_{ij} + X_n R\gamma_{[i}\gamma_{j]}\\
&\qquad + Y_n R_{ijkl}\gamma^k\gamma^l + Z_n R_{ijkl}\gamma^j\gamma^k]
\end{aligned}$$

as the most general covariant expression. The symmetry properties of the curvature tensor and of the products were used to eliminate other possible terms (e.g., $R_{ikjl}\gamma^k\gamma^l$; $R\gamma_{(i}\gamma_{j)}$, etc.).

It must be emphasized that dimensional considerations prevent including higher orders or higher derivatives of the curvature up to the second order of the Taylor development of any $F_n^{(x_0)}(x,x')$. Moreover, higher orders of the Taylor development will vanish at the order considered in the curvature. In fact, if the singular structure of the flat space–time derivatives of $I\Delta_1(s)$ with respect to m^2 is also expected to be reproduced by the derivatives of $G_1^{(1/2)}(x,x')$, i.e.,

$$\lim_{\substack{x\to x_0\\ x'\to x_0}} \frac{\partial^i G_1^{(1/2)}(x,x')}{(\partial m^2)^i} = \lim_{\substack{x\to x_0\\ x'\to x_0}} \sum_{n=i}^{\infty} F_n^{(x_0)}(x,x')\frac{\partial^n \Delta_1(s)}{(\partial m^2)^n} \qquad (39)$$

it can be proved that $F_0^{(x_0)}$ cannot depend on m^2, $F_1^{(x_0)}$ can only linearly depend on m^2, $F_2^{(x_0)}$ quadratically, etc. Therefore, the terms considered are all the terms that, up to the order of the curvature, will appear in the complete series.

The formalism is now developed in a spatially flat expanding Robertson–Walker universe (3). In such a metric the bispinorial distribution $G_1^{(1/2)(x_0)}(x,x')$ is expected to be the same function for all x_0 belonging to the surface $\Sigma(x_0) = \{x/t(x) = t(x_0)\}$ and therefore to define a good particle model if it verifies conditions (34) and (37).

The following expressions are obtained from (38) for the functions needed to develop $S_1^{(x_0)}(x,x')$:

$$\begin{aligned}
F_0^{(x_0)}(x,x')|_{\Sigma(x_0)} &= I - \tfrac{1}{2}H\gamma^0\gamma_\alpha r^\alpha + \tfrac{1}{2}a^2r^2\\
&\quad\times[(V_0+\tfrac{1}{4})H^2 - R(\tfrac{1}{6}V_0 + W_0)] + \cdots,\\
\partial_0 F_0^{(x_0)}(x,x')|_{\Sigma(x_0)}&\\
&= -[(X_0+\tfrac{1}{3}Y_0)R + 2Y_0H^2]\gamma^0\gamma_\alpha r^\alpha + \cdots,\\
F_\alpha^{(x_0)}(x,x')|_{\Sigma(x_0)} &= (m^2)^{\alpha-1}A_\alpha R + (m^2)^{\alpha-1}\\
&\quad\times[(B_\alpha + \tfrac{1}{6}C_\alpha) - C_\alpha H^2]\gamma_\alpha r^\alpha + \cdots, \qquad (40)\\
\partial_0 F_\alpha^{(x_0)}(x,x')|_{\Sigma(x_0)}&\\
&= (m^2)^{2\alpha-1}[(B_\alpha + \tfrac{1}{2}C_\alpha)R + 3C_\alpha H^2]\gamma^0 - (m^2)^\alpha\\
&\quad\times[(X_n + \tfrac{1}{3}Y_n)R + 2Y_nH^2]\gamma^0\gamma_\alpha r^\alpha + \cdots
\end{aligned}$$

[R and H should, of course, be evaluated at $t(x_0)$]; we have used expressions (A2) to eliminate some of the terms appearing in (38) and $\nabla_i\nabla_j F = \partial_i\partial_j F + (\partial_i\sigma_j)F + \sigma_j\partial_i F + \sigma_i\nabla_j F - \Gamma^h_{ij}\nabla_h F$.

We also need $\Delta_1(s)$ and its derivatives. The results obtained in Ref. 4 are only listed here:

$$\begin{aligned}
&s^2(x,x')|_\Sigma = a^2r^2(1 + \tfrac{1}{12}a^2r^2H^2),\\
&\partial_0 s^2(x,x')|_\Sigma = -a^2r^2H + \cdots,\\
&\partial_{00} s^2(x,x')|_\Sigma = 2 + \tfrac{2}{3}a^2r^2(H^2 + \tfrac{1}{16}R) + \cdots,\\
&\Delta_1(s)|_\Sigma = \frac{1}{(2\pi a)^3}\int d^3\mathbf{k}\,\frac{e^{-i\mathbf{k}\cdot\mathbf{r}}}{\omega_\mathbf{k}}\left(1 - \frac{5}{8}\frac{m^4}{\omega_\mathbf{k}^4}\frac{H^2}{\omega_\mathbf{k}^2} + \cdots\right), \qquad (41a)
\end{aligned}$$

$$\partial_0\Delta_1(s)|_\Sigma = \frac{-H}{(2\pi a)^3}\int d^3\mathbf{k}\,\frac{e^{-i\mathbf{k}\cdot\mathbf{r}}}{\omega_\mathbf{k}}\left(1 + \frac{m^2}{2\omega_\mathbf{k}^2} + \cdots\right), \qquad (41b)$$

$$\begin{aligned}
\partial_{00}\Delta_1(s)|_\Sigma &= \frac{1}{(2\pi a)^3}\int d^3\mathbf{k}\,\frac{e^{-i\mathbf{k}\cdot\mathbf{r}}}{\omega_\mathbf{k}}\\
&\quad\times\left[-\omega_\mathbf{k} + \left(\frac{7}{3}H^2 + \frac{R}{9}\right)\left(1 + \frac{m^2}{2\omega_\mathbf{k}^2}\right)\right.\\
&\quad\left. + \frac{5}{8}\frac{m^4}{\omega_\mathbf{k}^4}H^2 + \cdots\right], \qquad (41c)
\end{aligned}$$

where all the terms have been completely Fourier-analized.

We are now able to construct $S_1^{(x_0)}(x,x')$ and its derivatives and replacing them in (37) to determine the coefficients. It is easy to see that $F_n^{(x_0)}$ having $n > 2$ need not be included as they will contain higher powers than the second in m^2 and their coefficients will vanish or appear in a combination which will vanish when the different terms of the polynomial in $(m^2/\omega_\mathbf{k}^2)$ are equated to zero.

When (40) and (41) are replaced in (37) and the function $G_1^{(1/2)}(x,x')$ obtained in this way is replaced in (35), most of the coefficients are determined. In fact, if the particle model is to be defined for any possible evolution of the universe, terms involving R and H^2 must vanish independently, and it is therefore obtained that $V_0 = 1/6$, $A_1 = -1/12$ and $W_0 = C_0 = B_0 = C_1 = V_1 = W_1 = V_2 = W_2 = 0$.

We now ask Eq. (34) to be satisfied to determine the remaining coefficients.

Using the orthogonality conditions (11) and Eq. (33) (see Ref. 3), Eq. (34) can be written as

$$S_1(x,x') = \frac{1}{(2\pi a)^3}\sum_h \int d^3\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{r}}\left[-\Pi_{\mathbf{k}}^{(r)(+,h)}(t)\bar{\Pi}_{\mathbf{k}}^{(r)(+,h)}(t') + \Pi_{\mathbf{k}}^{(r)(-,h)} + \bar{\Pi}_{-\mathbf{k}}^{(r)(-,h)}(t)\right]. \tag{42}$$

On the other hand, if we replace in (33) the function $G_1^{(1/2)}(x,x')$ with the coefficients already determined, we get

$$\begin{aligned} S_1(x,x') = \frac{1}{(2\pi a)^3}\int d^3\mathbf{k}\frac{e^{-i\mathbf{k}\cdot\mathbf{r}}}{\omega_{\mathbf{k}}}\Bigg(&-ik_\alpha\gamma^\alpha\Bigg\{1+\frac{H^2}{\omega_{\mathbf{k}}^2}\left[\left(2Y_0+\frac{1}{4}\right)-\frac{m^2}{\omega_{\mathbf{k}}^2}\left(\frac{1}{8}+3Y_1\right)+\frac{m^4}{\omega_{\mathbf{k}}^4}\left(-\frac{5}{8}+\frac{15}{2}Y^2\right)\right] \\ &+\frac{R}{\omega_{\mathbf{k}}^2}\left[X_0+\frac{Y_0}{3}+\frac{1}{24}-\frac{m^2}{2\omega_{\mathbf{k}}^2}\left(\frac{1}{12}+3W_0+X_1+Y_1\right)+\frac{15}{4}\frac{m^4}{\omega_{\mathbf{k}}^4}\left(X_2+\frac{Y_2}{3}\right)\right]+\cdots \\ &+m\left\{1+\frac{5}{8}\frac{m^2}{\omega_{\mathbf{k}}^2}\frac{H^2}{\omega_{\mathbf{k}}^2}\left(1+\frac{m^2}{\omega_{\mathbf{k}}^2}\right)+\frac{1}{4}\frac{R}{\omega_{\mathbf{k}}^2}\left[\frac{1}{6}+\frac{m^2}{\omega_{\mathbf{k}}^2}\left(\frac{3}{4}A_2-\frac{1}{6}\right)\right]+\cdots\right\}\Bigg), \end{aligned} \tag{43}$$

which can be expressed as

$$S_1(x,x') = \frac{1}{(2\pi a)^3}\int d^3\mathbf{k}\frac{e^{-i\mathbf{k}\cdot\mathbf{r}}}{\omega_{\mathbf{k}}}M(\mathbf{k}), \tag{44}$$

where $M(\mathbf{k})$ is a matrix containing only the spinors I, $k_\alpha\gamma^\alpha$ and $\gamma^0 k_\alpha\gamma^\alpha$, and which under comparison with (4) has the explicit form

$$M(k) = \theta(\mathbf{k}) - i\beta(\mathbf{k})k_\alpha\gamma^\alpha + i\delta(\mathbf{k})\gamma^0 k_\alpha\gamma^\alpha.$$

Comparing (42) and (44) and using the orthogonality conditions (11), we get

$$(i\gamma^0\omega_{\mathbf{k}} + \theta + i\beta k_\alpha\gamma^\alpha - i\delta\gamma^0 k_\alpha\gamma^\alpha)\Pi_{\mathbf{k}} = 0. \tag{45}$$

Now, in order that this system of equations have a solution different from the trivial one, the determinant of the coefficients must be zero. This condition implies

$$(\theta^2 - \omega_{\mathbf{k}}^2)^2 - (k^4/a^4)(\beta^2+\delta^2) = 0,$$

i.e.,

$$[\theta^2 - \omega_{\mathbf{k}}^2 + (k^2/a^2)(\beta^2+\delta^2)][\theta^2 - \omega_{\mathbf{k}}^2 - (k^2/a^2)(\beta^2+\delta^2)] = 0.$$

One of the two members of the lhs of this expression must vanish, and the zeroth order (flat space-time limit) indicates that it is the first one. Therefore,

$$\theta^2 = \omega_{\mathbf{k}}^2 - (k^2/a^2)(\beta^2 - \delta^2).$$

Replacing θ, β, and δ from (43) and equating the corresponding orders in $(m^2/\omega_{\mathbf{k}}^2)$ and H and R, all the coefficients are determined. Indeed it turns out that $Y_0 = -1/8$, and the remaining coefficients vanish.

We can therefore write

$$\begin{aligned} S_1(x,x') = \frac{1}{(2\pi a)^3}\int d^3\mathbf{k}\frac{e^{-i\mathbf{k}\cdot\mathbf{r}}}{\omega_{\mathbf{k}}}\Bigg\{&-ik_\alpha\gamma^\alpha\left[1-\frac{m^2}{\omega_{\mathbf{k}}^2}\frac{H^2}{\omega_{\mathbf{k}}^2}\left(\frac{1}{8}+\frac{5}{8}\frac{m^2}{\omega_{\mathbf{k}}^2}\right)\right. \\ &\left.-\frac{R}{24\omega_{\mathbf{k}}^2}+\cdots\right]+m\left(1+\frac{5}{8}\frac{m^2}{\omega_{\mathbf{k}}^2}\frac{H^2}{\omega_{\mathbf{k}}^2}+\frac{1}{24}\frac{R}{\omega_{\mathbf{k}}^2}+\cdots\right)\frac{imH\gamma^0 k_\alpha\gamma^\alpha}{2\omega_{\mathbf{k}}^2}+\cdots\Bigg\}. \end{aligned} \tag{46a}$$

It can be seen that the same expression is obtained if we write, as it would have been naturally expected,

$$S_1(x,x') = (\gamma^i\nabla_i + m)T(x,x')G_1^{(0)}(x,x'), \tag{46b}$$

where T is the displacement bispinor, satisfying

$$\frac{1}{2}\frac{\partial s^2}{\partial x_i}T_{;i} = 0,\quad \frac{1}{2}\frac{\partial s^2}{\partial x_{i'}}T_{;i'} = 0,\ \lim_{x\to x'}T(x,x') = I$$

(see Ref. 7) and which, evaluated over a spatial surface $\Sigma = \{t = \text{const}\}$ reads

$$T(x,x') = 1 - \tfrac{1}{2}H\gamma^0\gamma_\alpha r^\alpha + \tfrac{1}{8}a^2r^2H^2 + \cdots$$

and its temporal derivative

$$\partial_0 T(x,x') = -(R/24) + (H/4)\gamma^0\gamma_\alpha r^\alpha + \cdots.$$

$G_1^{(0)}(x,x')$ is the scalar kernel obtained in Ref. 4, namely,

$$\begin{aligned} G_1^{(0)}(x,x') &= \Delta^{1/2}(x,x')\left[\Delta_1(s) - \left(\xi+\frac{1}{6}\right)R\frac{\partial\Delta_1}{\partial m^2}\right]+\cdots \\ &= \frac{1}{(2\pi a)^3}\int d^3\mathbf{k}\frac{e^{-i\mathbf{k}\cdot\mathbf{r}}}{\omega_{\mathbf{k}}} \\ &\quad\times\left\{1-\frac{5}{8}\frac{m^4}{\omega_{\mathbf{k}}^2}\frac{H^2}{\omega_{\mathbf{k}}^2}+\frac{1}{4}\frac{m^2}{\omega_{\mathbf{k}}^2}\frac{H^2}{\omega_{\mathbf{k}}^2}-\frac{R}{2\omega_{\mathbf{k}}^2}\right. \\ &\quad\times\left.\left[\left(\xi+\frac{1}{6}\right)+\frac{1}{12}\frac{m^2}{\omega_{\mathbf{k}}^2}\right]+\cdots\right\} \end{aligned} \tag{47}$$

with $\xi = -1/4$, a natural condition arising from Eq. (35), i.e., every Dirac equation solution is also a solution of the Klein–Gordon equation with a d'Alembert operator defined as $\Delta^{(1/2)} = -\nabla_i\nabla^i + 1/4R$, and

$$\Delta(x,x') = -g^{1/2}(x)\det\left[-\frac{1}{2}\partial_\mu\partial_\nu s^2(x,x')\right]g^{-1/2}(x')$$

is the Van Vleck determinant.

It must be pointed out that an equivalent expression has been obtained by DeWitt[7] and by Bunch and Parker[8] with different methods and based on different physical principles and that it can also be obtained if Eq. (A3) is used, i.e., we obtained the adiabatic particle model. In fact, it was shown in Ref. 8 that both the proper-time and the momentum-space representations lead to a Feynman propagator $G_F^{(1/2)}(x,x')$ equivalent to that explicitly indicated in expressions (46), namely $T(x,x')G_1^{(0)}(x,x')$. However, those $G_F^{(1/2)}(x,x')$ are constructed to solve the renormalization of a $\lambda\Phi^4$ interacting theory and neither is required to lead to a particle model.

Moreover, it can easily be seen that the kernel $S_1(x,x')$ found leads to the particle–antiparticle model expressed by spinors (10). In fact, if the following identification is made, $\Psi_{\mathbf{k}}^{(1)}(\Psi_{\mathbf{k}}^{(2)})$ is a particle with positive (negative) helicity, $\Pi_{\mathbf{k}}^{(\tau)(+,+)}(\Pi_{\mathbf{k}}^{(\tau)(+,-)})$, and $\Psi_{\mathbf{k}}^{(3)}(\Psi_{\mathbf{k}}^{(4)})$ is an antiparticle with positive (negative) helicity, $\Pi_{\mathbf{k}}^{(\tau)(-,+)}(\Pi_{\mathbf{k}}^{(\tau)(-,-)})$, and they are replaced in (42), expression (46) is obtained.

The formalism we have introduced can be extended to the massless case and the corresponding kernel can be expressed as

$$S_1(x,x') = \gamma^i\nabla_i[T(x,x')\mathscr{D}_1(x,x')], \tag{48}$$

where $\mathscr{D}_1(x,x')$ is the massless scalar kernel of the operator $(\Delta^{(0)} + 1/4R)$ namely

$$\mathscr{D}_1(x,x')|_\Sigma = -1/a^2r^2 + (R/48)\ln(a^2r^2) + \cdots \tag{49}$$

up to second order. In fact, it has already been pointed out by DeWitt[7] for the scalar case that an expansion corresponding to (37) involves inverse powers of m and can therefore not be used when $m = 0$. The corresponding expression is then obtained transcribing the singular structure of $D_1^{\mathrm{fl}}[s(x,x')]$.

Although this formalism can only be implemented in a power series development and the massless case has a simple exact solution, it can easily be seen that (48) leads to (15).

IV. DISCUSSION

The formalism stated above helps to select a function $S_1(x,x')$ which leads to no particle creation. Indeed as it is not able to select a different bispinorial kernel $S_1(x,x')$ on different Cauchy surfaces, it therefore defines the same particle–antiparticle model at different times [i.e., $S_1(x,x')$ turns out to be independent of the surface Σ]. However, the mechanism of particle creation has been studied in asymptotically static universes ("in–out" theories[9]) where the particle model is perfectly defined in the far future and past (plane waves) and a nonnull creation has been found. Therefore, there must be some terms which, although they satisfy the minimal hypotheses we have mentioned in the preceeding paragraph, are nonanalytical, i.e., they will never appear in a Taylor development (e.g., terms of the form e^{-k} will not appear in a power series of k^{-1}). This problem has been partially solved by Chitre and Hartle[10] for the scalar case. In fact, a kernel $G_F^{(0)}(x,x')$ was there found using path integral methods which leads to a reasonable particle creation (blackbody spectrum) when a linearly expanding universe is considered. In Ref. 11 some physical reasons supporting that choice are displayed, which will be implemented for the Dirac case in a forthcoming paper. Indeed different boundary conditions should be satisfied by the kernel $G_1^{(0)}(x,x')(G_1^{(1/2)}(x,x'))$ on different Cauchy surfaces, in particular near the singularity or on an adiabatic surface, allowing the choice of a different kernel in each case and thus leading to the creation of particles. The formalism which will be necessary to compute this creation has therefore already been introduced.

APPENDIX

The following identities evaluated in metric (3) were used throughout the paper.

The spinorial affine connection reads

$$\sigma_0 = 0,\quad \sigma_\alpha = \tfrac{1}{2}H\gamma^0\gamma_\alpha.$$

Dirac matrices γ_i can be written in terms of Pauli matrices as

$$\{\hat\gamma^i,\hat\gamma^j\} = 2\eta^{ij},$$

$$\hat\gamma^0 = i\begin{pmatrix} I & 0\\ 0 & -I\end{pmatrix},\quad \hat\gamma^\alpha = i\begin{pmatrix} 0 & -\sigma^\alpha\\ \sigma^\alpha & 0\end{pmatrix},$$

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

and they can be defined in this metric in terms of an arbitrary representation of the flat space–time constant ones $\hat\gamma_i$ as

$$\begin{aligned}&\gamma^0 = \hat\gamma^0,\quad \gamma^\alpha = a(t)^{-1}\hat\gamma^\alpha,\\ &\gamma_0 = \hat\gamma_0,\quad \gamma_\alpha = a(t)\hat\gamma_\alpha.\end{aligned} \tag{A1}$$

The nonvanishing coefficients of the Riemannian connection are

$$\Gamma^0_{\alpha\alpha} = a\dot a,\quad \Gamma^\alpha_{0\alpha} = \Gamma^\alpha_{\alpha 0} = H = \dot a/a.$$

The scalar curvature

$$R = -6(\dot H + 2H^2).$$

The nonvanishing components of the Ricci tensor are

$$R_{\alpha\alpha} = -a^2(R/6 - H^2),\quad R_{00} = 3H^2 + R/2,$$

and the independent ones of the curvature tensor

$$\begin{aligned}R_{\alpha 0\beta 0} &= -\delta_{\alpha\beta}a^2(R/6 + H^2),\\ R_{\alpha\beta\mu\nu} &= -a^2H^2(\delta_{\alpha\mu}\delta_{\beta\nu} - \delta_{\alpha\nu}\delta_{\beta\mu}).\end{aligned} \tag{A2}$$

The solution to Eqs. (12)–(13) up to second order in a power series of the Hubble coefficient H are

$$\begin{aligned}\Omega_k &= \omega_{\mathbf{k}}\bigg[1 + \frac{im}{2\omega_{\mathbf{k}}}\Big(1 + \frac{m}{\omega_{\mathbf{k}}}\Big)\frac{H}{\omega_{\mathbf{k}}}\\ &\quad + \Big(\pm\frac{5}{8}\frac{m^4}{\omega_{\mathbf{k}}^4} \mp \frac{m^2}{8\omega_{\mathbf{k}}^2} + \frac{m^3}{2\omega_{\mathbf{k}}^3}\Big)\frac{H^2}{\omega_{\mathbf{k}}^2}\\ &\quad \mp \frac{m}{24\omega_{\mathbf{k}}}\Big(1 + \frac{m}{\omega_{\mathbf{k}}}\Big)\frac{R}{\omega_{\mathbf{k}}^2} + \cdots\bigg],\\ \Lambda_k &= \omega_{\mathbf{k}}\bigg[1 - \frac{im}{2\omega_{\mathbf{k}}}\Big(1 - \frac{m}{\omega_{\mathbf{k}}}\Big)\frac{H}{\omega_{\mathbf{k}}}\\ &\quad + \Big(\pm\frac{5}{8}\frac{m^4}{\omega_{\mathbf{k}}^4} \mp \frac{m^2}{8\omega_{\mathbf{k}}^2} - \frac{m^3}{2\omega_{\mathbf{k}}^3}\Big)\frac{H^2}{\omega_{\mathbf{k}}^2}\\ &\quad \pm \frac{m}{24\omega_{\mathbf{k}}}\Big(1 - \frac{m}{\omega_{\mathbf{k}}}\Big)\frac{R}{\omega_{\mathbf{k}}^2} + \cdots\bigg].\end{aligned} \tag{A3}$$

[1]M. Castagnino, L. Chimento, and D. Harari, Phys. Rev. D **24**, 290 (1981).
[2]M. Castagnino and R. Weder, J. Math. Phys. **22**, 142 (1981).
[3]M. Castagnino, Ann. Inst. H. Poincaré **25**(1), 55 (1981).
[4]M. Castagnino, D. Harari, and C. Núñez, in *Proceedings of the Europhysics Study Conference, Erice (1982)*, edited by S. Ferrara and G. F. Ellis (Plenum, New York, 1983).
[5]L. Parker, Phys. Rev. D **3**, 346 (1971).
[6]A. Lichnerowicz, Bull. Soc. Math. France **92**, 11 (1964).
[7]B. S. DeWitt, *The Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965).
[8]T. S. Bunch and L. Parker, Phys. Rev. D **20**, 2499 (1979).
[9]L. Parker, in *Asymptotic Structure of Space–Time*, edited by F. Esposito and L. Witten (Plenum, New York, 1977).
[10]D. M. Chitre and J. B. Hartle, Phys. Rev. D **16**, 251 (1977).
[11]E. Calzetta and M. Castagnino, "On the Feynman Propagator in a Linear Expanding Universe," Phys. Rev. D **28**, 1298 (1983).

Superposition of solutions to Bäcklund transformations for the SU(n) principal σ-model[a)]

J. Harnad and Y. Saint-Aubin[b)]
Centre de recherche de mathématiques appliquées, Université de Montréal, C. P. 6128, Succ. A, Montréal, Québec, Canada H3C 3J7

S. Shnider[c)]
Department of Mathematics, McGill University, Montréal, Québec, Canada H3C 2T8

(Received 19 May 1983; accepted for publication 29 July 1983)

We show that the Bäcklund transformations for the SU(n) principal σ-model may be linearized using a geometrical interpretation of these equations involving the minimal orbit of SU(n, n) in the Grassmann manifold $G_n(\mathbb{C}^{2n})$. Linearization puts the equations in Zakharov–Mikhailov–Shabat (ZMS) form. Using this form of the equations, we prove inductively a nonlinear superposition law and a permutability theorem for iterated Bäcklund transformations analogous to known results in the theory of the sine–Gordon and KdV equations. From the superposition law we get an explicit form for multisoliton solutions to the σ-model.

PACS numbers: 11.10.Lm

1. INTRODUCTION

The two-dimensional principal U(n) σ-model is defined by the equations[1]

$$(g_{,\xi}\, g^{-1})_{,\eta} + (g_{,\eta}\, g^{-1})_{,\xi} = 0, \tag{1.1}$$

where $\xi = \frac{1}{2}(x+t)$ and $\eta = \frac{1}{2}(x-t)$ may be regarded as light cone coordinates in a two-dimensional Minkowski space and $g(\xi, \eta)$ is a U(n)-valued function. [We shall mainly be concerned with the case where $\det g = 1$, i.e., $g\in$SU(n) but refer part of the discussion to the general case $g\in$U(n).] This and similar models have been subject to considerable study in recent years, and it is known that they share many of the properties typical of completely integrable systems.[2–11] For example, there exist an infinity of conservation laws, both local and nonlocal, these being derivable either from Bäcklund transformations (BT) or from the linear equations of the inverse scattering method. The latter were shown by Zakharov, Mikhailov, and Shabat[6,7] (henceforth ZMS) to be solvable through the classical matrix Riemann problem, and, in particular, this method was used to derive multisoliton solutions.[7] However, for many other integrable systems, soliton solutions can be obtained more directly and simply through the use of Bäcklund transformations[12] (henceforth BT). Moreover, for those cases where a permutability theorem holds, the multisoliton solutions may be obtained recursively, giving rise to a sort of nonlinear superposition of individual solitons. It is our purpose to derive such a result for the BT of Eq. (1.1). To carry out this program we show, in the next section, how a natural geometric interpretation of the equations of the BT leads to a linearization, expressing their solution in terms of the solution to the ZMS equations. In Sec. 3 we use this linearized form to prove, recursively, that the solution to an iterated sequence of BT's is given by a nonlinear superposition law. The explicit formula implies a permutability theorem. Our methods depend on certain identities which only become clear when the equations are written in ZMS form, but the proofs are self-contained and do not use the ZMS theory.

a) Work supported in part by the National Science and Engineering Research Council of Canada and the Ministère de l'Education du Gouvernement du Québec.

b) Current address: Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139.

c) Current address: Department of Mathematics, Ben Gurion University of the Negev, P.O. Box 2053, Beersheva, Israel 84120.

2. GEOMETRICAL STRUCTURE OF BÄCKLUND TRANSFORMATIONS AND LINEARIZATION

Consider the BT's for Eq. (1.1) as given, e.g., (within slight modifications), in Ref. 10:

$$g_{,\xi} g^{-1} - g_{0,\xi} g_0^{-1} = -\lambda_0 (g g_0^{-1})_{,\xi} \tag{2.1}$$

$$g, g_0 \in \mathrm{U}(n),$$

$$g_{,\eta} g^{-1} - g_{0,\eta} g_0^{-1} = \lambda_0 (g g_0^{-1})_{,\eta} \tag{2.2}$$

with the constraint

$$\lambda_0 g g_0^{-1} + \bar{\lambda}_0 g_0 g^{-1} = \lambda_0 + \bar{\lambda}_0. \tag{2.3}$$

In fact, we shall only be concerned with solutions g_0 which lie in SU(n). As we shall see, the resulting g is, up to a constant phase factor, also in SU(n).

It is evident from Eqs. (2.1) and (2.2) that if g_0 satisfies Eq. (1.1), so does g and vice-versa. What is not evident by inspection is that if (2.1) and (2.2) are regarded as a system of equations for g, with given $g_0\in$U(n), the system is integrable, with $g\in$U(n), and that the nonlinear constraint (2.3) is compatible with Eqs. (2.1) and (2.2). To verify these facts and to simplify the analysis of the underlying equations, we introduce the new function

$$U \equiv g g_0^{-1} \tag{2.4}$$

in terms of which the system (2.1), (2.2), (2.3) becomes

$$U_\xi = (1/|1+\lambda_0|^2) \times \{ -\bar{\lambda}_0 A_0 + A_0 U - (1+\lambda_0+\bar{\lambda}_0) U A_0 + \lambda_0 U A_0 U \}, \tag{2.5a}$$

$$U_\eta = (1/|1-\lambda_0|^2) \times \{ -\bar{\lambda}_0 B_0 + B_0 U - (1-\lambda_0-\bar{\lambda}_0) U B_0 - \lambda_0 U B_0 U \} \tag{2.5b}$$

with the constraint

$$\lambda_0 U^2 - (\lambda_0 + \bar{\lambda}_0) U + \bar{\lambda}_0 = 0, \tag{2.6}$$

where

$$A_0 \equiv g_{0,\xi}\, g_0^{-1} = -A_0^{\dagger}, \quad B_0 \equiv g_{0,\eta} g_0^{-1} = -B_0^{\dagger}. \tag{2.7}$$

The requirement that g, and hence U, be unitary,

$$U^{\dagger} U = \mathbb{1}, \tag{2.8}$$

makes (2.6) equivalent to

$$\lambda_0 U + \bar{\lambda}_0 U^{\dagger} = \lambda_0 + \bar{\lambda}_0. \tag{2.9}$$

It is easily verified by cross differentiation that the integrability condition for (2.5a), (2.5b), for any λ_0, is that g_0 satisfy Eq. (1.1).

To study the constraints (2.8), (2.9), we make use of a geometrical interpretation of the matrix Riccati equations (2.5a), (2.5b), which at the same time shows how they may be linearized.[13] Let $G_n(\mathbb{C}^{2n})$ denote the Grassmann manifold of n dimensional subspaces of $\mathbb{C}^{2n}$. Homogeneous coordinates may be introduced, representing each subspace by the rank n rectangular matrix $\binom{X}{Y}$, $X, Y \in \mathbb{C}^{n\times n}$ whose columns span the space. The points of $G_n(\mathbb{C}^{2n})$ are identified with classes $\left[\begin{smallmatrix} X \\ Y \end{smallmatrix}\right]$ under the equivalence relation

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \begin{pmatrix} XT \\ YT \end{pmatrix}, \quad T \in \mathrm{Gl}(n, \mathbb{C}), \tag{2.10}$$

corresponding to a change of complex basis. On the affine subspace with $\det Y \neq 0$, we may introduce the affine coordinates identifying the point $\left[\begin{smallmatrix} X \\ Y \end{smallmatrix}\right]$ with the complex $n \times n$ matrix

$$U \equiv XY^{-1}. \tag{2.11}$$

Introducing a Hermitian structure on $\mathbb{C}^{2n}$ represented in the standard basis by the matrix

$$h = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix}, \tag{2.12}$$

we identify the submanifold $G_n^0(\mathbb{C}^{2n})$ of totally isotropic subspaces defined by

$$(X^{\dagger} Y^{\dagger}) h \begin{pmatrix} X \\ Y \end{pmatrix} = 0, \tag{2.13}$$

i.e.,

$$X^{\dagger} X = Y^{\dagger} Y.$$

The fact that $\binom{X}{Y}$ has rank n, together with (2.13), implies that Y is nonsingular, and hence the affine coordinates (2.11) are well defined on $G_n^0(\mathbb{C}^{2n})$. Equation (2.13) is then equivalent to the fact that, on $G_n^0(\mathbb{C}^{2n})$, U is unitary:

$$U^{\dagger} U = \mathbb{1}. \tag{2.14}$$

In fact, the resulting correspondence is easily shown to define a diffeomorphism $U(n) \sim G_n^0(\mathbb{C}^{2n})$.

The group $\mathrm{Sl}(2n, \mathbb{C})$ acts upon $G_n(\mathbb{C}^{2n})$ in the standard way induced by the linear action on $\mathbb{C}^{2n}$. In terms of affine coordinates, the action of an element

$$\begin{pmatrix} P & Q \\ R & S \end{pmatrix} \in \mathrm{Sl}(2n, \mathbb{C}), \quad P, Q, R, S \in \mathbb{C}^{n\times n}, \tag{2.15}$$

is given by the linear fractional transformations

$$\begin{pmatrix} P & Q \\ R & S \end{pmatrix}: U \mapsto (PU + Q)(RU + S)^{-1}. \tag{2.16}$$

The infinitesimal form of this action is defined by a Lie algebra homomorphism ϕ: $\mathrm{sl}(2n, \mathbb{C}) \mapsto \chi(G_n(\mathbb{C}^{2n}))$ to the algebra of vector fields (i.e., infinitesimal displacements) on $G_n(\mathbb{C}^{2n})$, defined in affine coordinates by

$$\phi: \begin{pmatrix} p & q \\ r & s \end{pmatrix} \mapsto -\{q - Us + pU - UrU\} \cdot \nabla_U,$$
$$\begin{pmatrix} p & q \\ r & s \end{pmatrix} \in \mathrm{sl}(2n, \mathbb{C}). \tag{2.17}$$

Given any curve

$$\begin{pmatrix} p(t) & q(t) \\ r(t) & s(t) \end{pmatrix} \quad \text{in } \mathrm{sl}(2n, \mathbb{C}),$$

the corresponding t-dependent vector field on $G_n(\mathbb{C}^{2n})$ defines the matrix Riccati equation

$$\dot{U}(t) = q + pU - Us - UrU, \tag{2.18}$$

whose integral curves are such that the tangent at $U(t)$, for any t, is

$$\phi\left(\begin{pmatrix} p(t) & q(t) \\ r(t) & s(t) \end{pmatrix}\right)\Big|_{U(t)}.$$

The same construction is applicable to systems of matrix Riccati partial differential equations, where the parameter t is replaced by the independent variables $\{t^i\}_{i=1,\dots,m}$, provided the equations satisfy appropriate integrability conditions. A system of integrable PDE's of the form

$$\begin{aligned} &\frac{\partial U}{\partial t^i} = q_i + p_i U - U S_i - U r_i U, \\ &U = U(\mathbf{t}), \\ &q_i = q_i(\mathbf{t}), \quad p_i = p_i(\mathbf{t}), \quad s_i = s_i(\mathbf{t}), \quad r_i = r_i(\mathbf{t}), \\ &\mathbf{t} = (t_1, \dots, t_m) \in \mathbb{R}^m, \end{aligned} \tag{2.19}$$

may be interpreted as defining (locally) a horizontal (covariant constant) section of the trivial bundle with fibre $G_n(\mathbb{C}^{2n})$ over $\mathbb{R}^m$ associated with the group action $\mathrm{Sl}(2n, \mathbb{C})$: $G_n(\mathbb{C}^{2n}) \mapsto G_n(\mathbb{C}^{2n})$ to the trivial principal bundle $\mathrm{Sl}(2n, \mathbb{C}) \times \mathbb{R}^m \mapsto \mathbb{R}^m$ with connection form

$$\omega|_{(g,\mathbf{t})} = -\mathrm{Ad}\, g^{-1} \begin{pmatrix} p_i(\mathbf{t}) & q_i(\mathbf{t}) \\ r_i(\mathbf{t}) & s_i(\mathbf{t}) \end{pmatrix} dt^i + g^{-1} dg. \tag{2.20}$$

The integrability condition is the vanishing of the corresponding curvature

$$\Omega \equiv d\omega + \tfrac{1}{2}[\omega, \omega] = 0. \tag{2.21}$$

The covariant constant cross sections defined by integration of (2.19) are the maximal integral manifolds of the horizontal distribution spanned by

$$X_i \equiv \frac{\partial}{\partial t^i} - \phi\left(\begin{pmatrix} p_i(\mathbf{t}) & q_i(\mathbf{t}) \\ r_i(\mathbf{t}) & s_i(\mathbf{t}) \end{pmatrix}\right). \tag{2.22}$$

The corresponding horizontal cross sections of the principal bundle σ: $\mathbf{t} \mapsto (G(\mathbf{t}), \mathbf{t})$ are defined by the $\mathrm{Sl}(2n, \mathbb{C})$-valued function

$$G(\mathbf{t}) = \begin{pmatrix} P(\mathbf{t}) & Q(\mathbf{t}) \\ R(\mathbf{t}) & S(\mathbf{t}) \end{pmatrix}, \tag{2.23}$$

satisfying the linear equations

$$\frac{\partial}{\partial t^i}\begin{pmatrix} P & Q \\ R & S \end{pmatrix} = \begin{pmatrix} p_i & q_i \\ r_i & s_i \end{pmatrix}\begin{pmatrix} P & Q \\ R & S \end{pmatrix} \tag{2.24}$$

with initial condition, e.g.,

$$G(\mathbf{t}_0) = \mathbb{1}. \tag{2.25}$$

The general solution to (2.19) is thus obtained by applying $G(\mathbf{t}_0)$ to the initial value U_0:

$$\begin{aligned} U(\mathbf{t}) &= X(\mathbf{t})Y^{-1}(\mathbf{t}) \\ &= (P(\mathbf{t})U_0 + Q(\mathbf{t}))(R(\mathbf{t})U_0 + S(\mathbf{t}))^{-1}, \end{aligned} \tag{2.26}$$

where

$$U_0 = X_0 Y_0^{-1} \tag{2.27}$$

and the homogeneous coordinates may be taken as

$$\begin{pmatrix} X(\mathbf{t}) \\ Y(\mathbf{t}) \end{pmatrix} = \begin{pmatrix} P(\mathbf{t}) & Q(\mathbf{t}) \\ R(\mathbf{t}) & S(\mathbf{t}) \end{pmatrix}\begin{pmatrix} X_0 \\ Y_0 \end{pmatrix}. \tag{2.28}$$

Restricting now to the the subgroup $\mathrm{SU}(n, n)$ consisting of those transformations preserving h, we have

$$\begin{aligned} P^\dagger P - R^\dagger R &= \mathbb{1}, \\ P^\dagger Q - R^\dagger S &= 0, \\ Q^\dagger Q - S^\dagger S &= \mathbb{1}. \end{aligned} \tag{2.29}$$

The infinitestimal form of these relations, defining the subalgebra $\mathrm{su}(n, n) \subset \mathrm{sl}(2n, \mathbb{C})$ is given by

$$\begin{aligned} p^\dagger &= -p, \\ r^\dagger &= q, \\ s^\dagger &= -s, \end{aligned} \tag{2.30}$$

together with

$$\mathrm{tr}(p + s) = 0. \tag{2.31}$$

This subgroup clearly preserves the submanifold $\mathrm{U}(n) \sim G_n^0(\mathbb{C}^{2n}) \subset G_n(\mathbb{C}^{2n})$ and it is easily verified that the action is transitive. That is, $G_n^0(\mathbb{C}^{2n}) \sim \mathrm{U}(n)$ is identifiable with a single orbit of $\mathrm{SU}(n, n)$ in $G_n(\mathbb{C}^{2n})$, the one of minimal dimension. It follows that if the coefficient matrices $\{p_i(\mathbf{t}), q_i(\mathbf{t}), r_i(\mathbf{t}), s_i(\mathbf{t})\}$ in (2.19) satisfy (2.30), (2.31), then the group-valued function $G(\mathbf{t})$ obtained by integrating (2.24), (2.25) will be in $\mathrm{SU}(n,n)$, thus preserving the Hermitian form h. Consequently, the isotropy condition (2.13) will be preserved by the solution $\binom{X(\mathbf{t})}{Y(\mathbf{t})}$ given by Eq. (2.28) provided it holds for $\binom{X_0}{Y_0}$, or equivalently, the unitarity condition (2.14) holds for $\mathrm{U}(\mathbf{t})$ provided it does for U_0.

Notice now that Eqs. (2.5a), (2.5b) are precisely of the form (2.19), with $\mathbf{t} = (\xi, \eta)$ and the coefficient matrices satisfy the relations (2.30) defining the $\mathrm{su}(n, n)$ subalgebra. The above considerations therefore prove that the unitarity constraint (2.8) is preserved. Moreover, the general solution is given by Eq. (2.26) or (2.28), with the group-valued function

$$G(\xi, \eta) = \begin{pmatrix} P(\xi, \eta) & Q(\xi, \eta) \\ R(\xi, \eta) & S(\xi, \eta) \end{pmatrix},$$

satisfying

$$\begin{aligned} \begin{pmatrix} P & Q \\ R & S \end{pmatrix}_{,\xi} &= \frac{1}{|1 + \lambda_0|^2} \\ &\times \begin{pmatrix} A_0 & \bar{\lambda}_0 A_0 \\ -\lambda_0 A_0 & (1 + \lambda_0 + \bar{\lambda}_0)A_0 \end{pmatrix}\begin{pmatrix} P & Q \\ R & S \end{pmatrix}, \end{aligned} \tag{2.32a}$$

$$\begin{aligned} \begin{pmatrix} P & Q \\ R & S \end{pmatrix}_{,\eta} &= \frac{1}{|1 - \lambda_0|^2} \\ &\times \begin{pmatrix} B_0 & -\bar{\lambda}_0 B_0 \\ \lambda_0 B_0 & (1 - \lambda_0 - \bar{\lambda}_0)B_0 \end{pmatrix}\begin{pmatrix} P & Q \\ R & S \end{pmatrix}. \end{aligned} \tag{2.32b}$$

The second constraint (2.9) may also be expressed in terms of the preservation of a Hermitian form, since it is equivalent to the relations

$$(X^\dagger\ Y^\dagger)\begin{pmatrix} 0 & \bar{\lambda}_0 \\ \lambda_0 & -(\lambda_0 + \bar{\lambda}_0) \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix} = 0, \tag{2.33}$$

defining the submanifold $\tilde{G}_n^0(\mathbb{C}^{2n})$ of maximal isotropic n-dimensional subspaces of $\mathbb{C}^{2n}$ under the Hermitian form

$$\tilde{h} = \begin{pmatrix} 0 & \bar{\lambda}_0 \\ \lambda_0 & -(\lambda_0 + \bar{\lambda}_0) \end{pmatrix}. \tag{2.34}$$

The intersection of $\mathrm{SU}(n, n)$ with the subgroup of $\mathrm{Gl}(2n, \mathbb{C})$ preserving $\tilde{h}$ consists of those elements commuting with the transformation

$$t \equiv h^{-1}\tilde{h} = \begin{pmatrix} 0 & \bar{\lambda}_0 \\ -\lambda_0 & \lambda_0 + \bar{\lambda}_0 \end{pmatrix}. \tag{2.35}$$

Since this transformation has two distinct eigenvalues $\lambda_0, \bar{\lambda}_0$, each corresponding to an n-dimensional space of eigenvectors, the subgroup of $\mathrm{Gl}(2n, \mathbb{C})$ commuting with it is equivalent to $\mathrm{Gl}(n, \mathbb{C}) \times \mathrm{Gl}(n, \mathbb{C})$. Making the appropriate change of basis diagonalizing t, this subgroup consists of elements of the form

$$\begin{pmatrix} P & Q \\ R & S \end{pmatrix} = T\begin{pmatrix} K & 0 \\ 0 & L \end{pmatrix}T^{-1}, \tag{2.36}$$

where

$$T = \begin{pmatrix} \mathbb{1} & \bar{\lambda}_0\mathbb{1} \\ \mathbb{1} & \lambda_0\mathbb{1} \end{pmatrix}. \tag{2.37}$$

The further condition that h be preserved implies

$$L = K^{\dagger -1}. \tag{2.38}$$

The condition

$$\det\begin{pmatrix} P & Q \\ R & S \end{pmatrix} = 1 \tag{2.39}$$

only implies $\det K$ is real; however, since we are only interested in the action of this group on $G_n(\mathbb{C}^{2n})$, we may take $\det K = 1$. This will be seen to follow from the block diagonalized form of Eqs. (2.32) without any renormalization of the K in (2.26), (2.36), provided $\det g_0$ is constant, and in particular for $g_0 \in \mathrm{SU}(n)$.

Thus, the underlying group which simultaneously preserves the two constraints (2.8) and (2.9) is the subgroup $\mathrm{Sl}(n, \mathbb{C}) \subset \mathrm{Sl}(2n, \mathbb{C})$ defined by the embedding

$$\begin{aligned} &K \mapsto T\begin{pmatrix} K & 0 \\ 0 & K^{\dagger -1} \end{pmatrix}T^{-1} \in \mathrm{Sl}(2n, \mathbb{C}), \\ &K \in \mathrm{Sl}(n, \mathbb{C}). \end{aligned} \tag{2.40}$$

The corresponding $\mathrm{sl}(n, \mathbb{C}) \subset \mathrm{sl}(2n, \mathbb{C})$ subalgebra is defined by

$$k \mapsto T \begin{pmatrix} k & 0 \\ 0 & -k^{\dagger} \end{pmatrix} T^{-1} \in \mathrm{sl}(2n, \mathbb{C}),$$
$$k \in \mathrm{sl}(n, \mathbb{C}). \tag{2.41}$$

To preserve the constraints (2.8) and (2.9), the algebra elements defining the matrix Riccati equations (2.19) must thus all be of the form

$$\begin{aligned} p &= \frac{1}{\lambda_0 - \bar{\lambda}_0} (\lambda_0 k + \bar{\lambda}_0 k^{\dagger}), \\ q &= \frac{-\bar{\lambda}_0}{\lambda_0 - \bar{\lambda}_0} (k + k^{\dagger}), \\ r &= \frac{\lambda_0}{\lambda_0 - \bar{\lambda}_0} (k + k^{\dagger}), \\ s &= -\frac{1}{\lambda_0 - \bar{\lambda}_0} (\bar{\lambda}_0 k^{\dagger} + \lambda_0 k). \end{aligned} \tag{2.42}$$

Comparison with Eqs. (2.5) shows that this is indeed the case, with

$$k = A_0/(1 + \lambda_0) \quad \text{for (2.5a)}, \tag{2.43a}$$
$$k = B_0/(1 - \lambda_0) \quad \text{for (2.5b)}. \tag{2.43b}$$

The significance of this construction is that it not only demonstrates the consistency of Eqs. (2.5) with the constraints (2.8), (2.9) but also indicates how (2.5) may be linearized. From Eqs. (2.36), (2.38), (2.42), (2.43a), and (2.43b), we see that $\left(\begin{smallmatrix} P & Q \\ R & S \end{smallmatrix}\right)$ may be expressed as

$$\begin{pmatrix} P & Q \\ R & S \end{pmatrix} = \begin{pmatrix} \lambda_0 K - \bar{\lambda}_0 K^{\dagger -1} & -\bar{\lambda}_0 (K - K^{\dagger -1}) \\ \lambda_0 (K - K^{\dagger -1}) & -\bar{\lambda}_0 K + \lambda_0 K^{\dagger -1} \end{pmatrix}, \tag{2.44}$$

where $K(\xi, \eta)$ satisfies

$$K_{\xi} = A_0 K/(1 + \lambda_0), \tag{2.45a}$$
$$K_{\eta} = B_0 K/(1 - \lambda_0). \tag{2.45b}$$

Notice that, provided $\det g_0$ is a constant, Eqs. (2.43) imply $\mathrm{tr}\, A_0 = \mathrm{tr}\, B_0 = 0$ and hence, from Eqs. (2.45), without loss of generality we may take $\det K = 1$ as previously stated. Thus, by Eq. (2.26), the general solution to (2.5a), (2.5b) is given in terms of $K(\xi, \eta)$ by

$$U = [(\lambda_0 K - \bar{\lambda}_0 K^{\dagger -1}) U_0 - \bar{\lambda}_0 (K - K^{\dagger -1})]$$
$$\times [\lambda_0 (K - K^{\dagger -1}) U_0 + (-\bar{\lambda}_0 K + \lambda_0 K^{\dagger -1})]^{-1}. \tag{2.46}$$

Equations (2.45) are precisely those of ZMS, which are the starting point of the inverse scattering approach. Here, they have been derived as a consequence of the Bäcklund transformations rather than vice-versa. To facilitate further comparison, we shall henceforth adopt the notation of Refs. 5–7, writing

$$\psi_{\xi}(\lambda_0) = A_0 \psi/(1 + \lambda_0), \tag{2.47a}$$
$$\psi_{\eta}(\lambda_0) = B_0 \psi/(1 - \lambda_0), \tag{2.47b}$$

with

$$K \equiv \psi(\lambda_0). \tag{2.48}$$

Note that, because $A_0, B_0 \in \mathrm{u}(n)$, we have

$$K^{\dagger -1} = \psi(\bar{\lambda}_0). \tag{2.49}$$

Furthermore, writing

$$U \equiv \mathbb{1} + [(\bar{\lambda}_0 - \lambda_0)/\lambda_0] P \tag{2.50}$$

the constraints (2.8) and (2.9) are equivalent to the conditions that P be an orthogonal projector

$$P^2 = P, \quad P = P^{\dagger}. \tag{2.51}$$

Substituting (2.48), (2.49) in (2.46) and solving (2.50) for P gives

$$P = \psi(\bar{\lambda}_0) \pi [\psi(\bar{\lambda}_0) \pi + \psi(\lambda_0)(1 - \pi)]^{-1}, \tag{2.52}$$

where

$$\pi = [\lambda_0/(\bar{\lambda}_0 - \lambda_0)][U_0 - \mathbb{1}] \tag{2.53}$$

is the value of P which corresponds to $U = U_0$. [Note that P and U need not necessarily assume these values at any initial point, since we have not required the initial value of $\psi(\lambda)$ to equal unity.] Since P is an orthogonal projector, it may be written

$$P = M(M^{\dagger} M)^{-1} M^{\dagger}, \tag{2.54}$$

where M is a rectangular $n \times k$ matrix whose columns span the image of P. Similarly expressing the initial value as

$$\pi = m(m^{\dagger} m)^{-1} m^{\dagger}, \tag{2.55}$$

Eq. (2.52) is equivalent to

$$M = \psi(\bar{\lambda}_0) m \tag{2.56}$$

precisely as in ZMS. [Notice that even if m is chosen as a unitary basis, M will not be because $\psi(\bar{\lambda}_0)$ for complex λ_0 is not unitary.] From Eq. (2.50) we see that

$$\det U = (\bar{\lambda}_0/\lambda_0)^k \tag{2.57}$$

and therefore $\det g$ differs from $\det g_0$ by this constant phase factor. Since rescaling preserves Eq. (1.1), we may interpret (2.1) and (2.2) as a BT for the $\mathrm{SU}(n)$ σ-model by modifying the normalization of g in (2.3) by $(\lambda_0/\bar{\lambda}_0)^{k/n}$. The modification thus depends on the rank of the projector determined by Eq. (2.3).

3. RECURSIVE SOLUTION OF BÄCKLUND TRANSFORMATIONS AND SUPERPOSITION FORMULA

We now turn to the problem of obtaining iterative solutions to the Bäcklund tranformations; that is, solving a sequence of equations like (2.1), (2.2) with the input g_0 determined as the solution for the previous step, a new value of the parameter λ_0 being introduced at each iteration. The defining equations of the sequence are thus

$$g_{i,\xi} g_i^{-1} - g_{i-1,\xi} g_{i-1}^{-1} = -\lambda_i (g_i g_{i-1}^{-1})_{,\xi}, \tag{3.1a}$$
$$g_{i,\eta} g_i^{-1} - g_{i-1,\eta} g_{i-1}^{-1} = \lambda_i (g_i g_{i-1}^{-1})_{,\eta}, \tag{3.1b}$$
$$\lambda_i g_i g_{i-1}^{-1} + \bar{\lambda}_i g_{i-1} g_i^{-1} = \lambda_i + \bar{\lambda}_i, \quad i = 1,\dots,l. \tag{3.2}$$

Applying the linearization procedure of the previous section to each step, we have

$$g_i = \left(1 + \frac{\bar{\lambda}_i - \lambda_i}{\lambda_i} P_i\right) g_{i-1} \equiv U_i g_{i-1}, \tag{3.3}$$

where

$$P_i = \widetilde{M}_i (\widetilde{M}_i^\dagger \widetilde{M}_i)^{-1} \widetilde{M}_i^\dagger, \tag{3.4}$$
$$\widetilde{M}_i \in \mathbb{C}^{n\times k_i}, \quad k_i = \operatorname{rank} P_i,$$

and

$$\widetilde{M}_i = \psi_{i-1}(\bar{\lambda}_i) m_i, \tag{3.5}$$
$$\widetilde{M}_i^\dagger = m_i^\dagger \psi_{i-1}^{-1}(\lambda_i), \quad m_i \in \mathbb{C}^{n\times k_i},$$

where

$$\psi_{i,\xi}(\lambda) = A_i \psi_i/(1+\lambda), \tag{3.6a}$$
$$\psi_{i,\eta}(\lambda) = B_i \psi_i/(1-\lambda), \tag{3.6b}$$

$$A_i = g_{i,\xi} g_i^{-1}, \quad B_i \equiv g_{i,\eta} g_i^{-1}, \quad i = 0,1,\ldots,l. \tag{3.7}$$

As pointed out in the previous section, these equations (ignoring the index i) are the starting point for the ZMS approach to the σ-model. In that approach the three matrix functions $\psi(\lambda)(\xi,\eta), A(\xi,\eta), B(\xi,\eta)$ may be regarded simultaneously as unknowns of the system, and the consistency of the system together with λ analyticity of ψ implies that $\psi(0) = g$ is a solution of the field equation (1.1). However, since this same system is the linearization of the BT, it provides, as we have seen in Eqs. (2.50), (2.52), and (2.4), a new solution to the field equation as well. The remarkable fact shown by ZMS is that there is a transformation, defined below in Eqs. (3.8) and (3.12) of the entire system $\psi_{i-1}(\lambda)$, A_{i-1}, B_{i-1} to a new system $\psi_i(\lambda), A_i, B_i$, of which the BT is just the $\lambda = 0$ special case. This fact implies that the solution to all iterated BT's should be expressible in terms of solutions to the first BT. In what follows we prove these results directly from the BT.

Let

$$\chi_i(\lambda) \equiv \left\{ 1 + \frac{\lambda_i - \bar{\lambda}_i}{\lambda - \lambda_i} P_i \right\}$$
$$= \frac{\lambda - \lambda_i U_i}{\lambda - \lambda_i}, \quad i = 1,\ldots,l, \tag{3.8}$$

such that

$$g_i = \chi_i(0) g_{i-1} = U_i g_{i-1}. \tag{3.9}$$

The U(n) valued function U_i satisfies the ith iteration of Eqs. (2.5):

$$U_{i,\xi} = (1/|1+\lambda_i|^2)[\bar{\lambda}_i A_{i-1} + A_{i-1} U_i$$
$$- (1 + \lambda_i + \bar{\lambda}_i) U_i A_{i-1} + \lambda_i U_i A_{i-1} U_i], \tag{3.10a}$$
$$U_{i,\eta} = (1/|1-\lambda_i|^2)[-\bar{\lambda}_i B_{i-1} + B_{i-1} U_i$$
$$- (1 - \lambda_i - \bar{\lambda}_i) U_i B_{i-1} - \lambda_i U_i B_{i-1} U_i], \tag{3.10b}$$
$$\lambda_i U_i + \bar{\lambda}_i U_i^\dagger = \lambda_i + \bar{\lambda}_i. \tag{3.11}$$

We now define recursively the series of matrix functions

$$\psi_i(\lambda) \equiv \chi_i(\lambda) \psi_{i-1}(\lambda) \tag{3.12}$$

with $\psi_0(\lambda)$ chosen as the $i = 0$ solution of Eqs. (3.6) and prove inductively that these are indeed the solutions for all i. Assuming the relations (3.6) to hold for $i - 1$, and, using Eqs. (3.8) and (3.9), we find that ψ_i satisfies the relations

$$(1+\lambda)\psi_{i,\xi}\psi_i^{-1} = \left(1 + \frac{\bar{\lambda}_i - \lambda_i}{1+\lambda_i} P_i\right) A_{i-1}$$
$$\times \left(1 + \frac{\lambda_i - \bar{\lambda}_i}{1+\bar{\lambda}_i} P_i\right), \tag{3.13a}$$
$$(1-\lambda)\psi_{i,\eta}\psi_i^{-1} = \left(1 + \frac{\lambda_i - \bar{\lambda}_i}{1-\lambda_i} P_i\right) B_{i-1}$$
$$\times \left(1 + \frac{\bar{\lambda}_i - \lambda_i}{1-\bar{\lambda}_i} P_i\right). \tag{3.13b}$$

Since the rhs is independent of λ and for $\lambda = 0$ the lhs becomes equal, by the inductive hypothesis and Eq. (3.9), to A_i and B_i, respectively, we see that Eqs. (3.6) are satisfied, proving by induction the result for all i. It should be mentioned that Eq. (3.6) alone determine ψ_i only up to right multiplication by a fixed matrix; however, this arbitrariness is entirely absorbed in the matrices m_i of Eq. (3.5). From Eqs. (3.8) and (3.13) we see that for all ψ_i to remain nonsingular, it is necessary to require $\lambda \neq \lambda_j, \bar{\lambda}_j, \forall j \geqslant i$. Defining the product

$$\hat{\chi}_l(\lambda) \equiv \prod_{i=1}^{l} \chi_i(\lambda) = \chi_l(\lambda) \cdots \chi_1(\lambda), \tag{3.14}$$

we have

$$\psi_l(\lambda) = \hat{\chi}_l(\lambda) \psi_0(\lambda) \tag{3.15}$$

and, in particular, for $\lambda = 0$,

$$g_l = \hat{\chi}_l(0) g_0. \tag{3.16}$$

This determines the solution to the lth iteration to the BT (3.1) without any integrations needed beyond the first step. Such a result also follows from the methods of ZMS, but we have derived it here directly from the structure of the BT.

Although Eq. (3.15) gives a recursive solution to the iterated problem, it does not express the solution directly in terms of solutions to the first step; that is, in terms of solutions to

$$\hat{g}_{i,\xi}\hat{g}_i^{-1} - g_{0,\xi} g_0^{-1} = -\lambda_i (\hat{g}_i g_0^{-1})_{,\xi}, \tag{3.17a}$$
$$\hat{g}_{i,\eta}\hat{g}_i^{-1} - g_{0,\eta} g_0^{-1} = \lambda_i (\hat{g}_i g_0^{-1})_{,\eta}, \tag{3.17b}$$
$$\lambda_i \hat{g}_i g_0^{-1} + \bar{\lambda}_i g_0 \hat{g}_i^{-1} = \lambda_i + \bar{\lambda}_i. \tag{3.18}$$

The solution to these equations is, by Eqs. (2.50), (2.54), and (2.55), of the form

$$\hat{g}_i = \left\{ 1 + \frac{(\bar{\lambda}_i - \lambda_i)}{\lambda_i} M_i (M_i^\dagger M_i)^{-1} M_i^\dagger \right\} g_0, \tag{3.19}$$

where

$$M_i = \psi_0(\bar{\lambda}_i) m_i \tag{3.20}$$

whereas Eqs. (3.14)–(3.16) are expressed in terms of $\widetilde{M}_i$, as defined by Eq. (3.5). It is possible, however, to develop the product in Eq. (3.14) and thereby derive a superposition formula explicitly expressing $\hat{\chi}_l(\lambda)$ and hence g_l in terms of the M_i's entering in the solution (3.19). We state the result as a proposition:

Proposition: Let all $\lambda_i \neq \bar{\lambda}_j$ for $i \neq j$ and $1 \leqslant i, j \leqslant l$. If $\lambda_i = \lambda_j$ for any $i \neq j$, let m_i and m_j be such that the vector subspaces $V_{\{m_i\}}$ and $V_{\{m_j\}}$ spanned by the column vectors of m_i and m_j, respectively, are linearly independent ($V_{\{m_i\}} \cap V_{\{m_j\}} = \{0\}$).

Under these restrictions and when $\lambda = \lambda_i$ for all $1 \leqslant i \leqslant l$, the matrix $\hat{\chi}_l(\lambda)$ defined by (3.14) and (3.8) can be written as

$$\hat{\chi}_l(\lambda) = \mathbb{1} + \sum_{i,j=1}^{l} \frac{M_i \gamma_{ij} M_j^\dagger}{\lambda - \lambda_j}, \tag{3.21}$$

where γ_{ij} is the $k_i \times k_j$ matrix whose components $\gamma_{ij}^{\alpha\beta}$ ($1 \leqslant \alpha \leqslant k_i$, $1 \leqslant \beta \leqslant k_j$) are defined as follows. Let

$$\Gamma_{ij} = \frac{M_i^\dagger M_j}{\lambda_i - \bar{\lambda}_j} \in \mathbb{C}^{k_i \times k_j}, \quad 1 \leqslant i,j \leqslant l. \tag{3.22}$$

Denoting the $(\alpha\beta)$ element as $\Gamma_{ij}^{\alpha\beta}$, we may regard these blocks as defining a $\kappa \times \kappa$ matrix Γ, where

$$\kappa = \sum_{i=1}^{l} k_i \tag{3.23}$$

indexed by the pair $(i\alpha)(j\beta)$ ($1 \leqslant i,j \leqslant l$, $1 \leqslant \alpha \leqslant k_i$, $1 \leqslant \beta \leqslant k_j$). Γ is invertible and γ is its inverse:

$$\gamma \equiv \Gamma^{-1}. \tag{3.24}$$

The $(\alpha\beta)$ component of γ_{ij} in (3.21) is the $(i\alpha)(j\beta)$ matrix element $\gamma_{ij}^{\alpha\beta}$.

To prove this result, we first make the assumption that Γ as defined by (3.22) is invertible, and show by induction on l that (3.21) is valid. We then prove, also by induction, that Γ defined in (3.22) is indeed invertible. The validity of (3.21), (3.22) for $l = 1$ follows from the definition (3.8) for $i = 1$ in view of (3.4) and the equality

$$\tilde{M}_1 = M_1. \tag{3.25}$$

For $l \geqslant 2$, we have, from Eqs. (3.5), (3.12), (3.14), and (3.20),

$$\tilde{M}_l = \hat{\chi}_{l-1}(\bar{\lambda}_l) M_l. \tag{3.26}$$

Assume now that (3.21) is valid for $l - 1$,

$$\hat{\chi}_{l-1} = \mathbb{1} + \sum_{i,j=1}^{l-1} \frac{M_i \tilde{\gamma}_{ij} M_j^\dagger}{\lambda - \lambda_j}, \tag{3.27}$$

where $\{\tilde{\gamma}_{ij}^{\alpha\beta}\}$ are the components of the $(\kappa - k_l) \times (\kappa - k_l)$ matrix inverse of the submatrix of Γ with components $\Gamma_{ij}^{\alpha\beta}$ ($1 \leqslant i,j \leqslant l-1$, $1 \leqslant \alpha \leqslant k_i$, $1 \leqslant \beta \leqslant k_j$). We thus have

$$\begin{aligned}\hat{\chi}_l(\lambda) &= \left(\mathbb{1} + \frac{\lambda_l - \bar{\lambda}_l}{\lambda - \lambda_l} P_l\right)\left(\mathbb{1} + \sum_{i,j=1}^{l} \frac{M_i \tilde{\gamma}_{ij} M_j^\dagger}{\lambda - \lambda_j}\right) \\ &= \mathbb{1} + \frac{\lambda_l - \bar{\lambda}_l}{\lambda - \lambda_l} P_l \left(\mathbb{1} - \sum_{i,j=1}^{l-1} \frac{M_i \tilde{\gamma}_{ij} M_j^\dagger}{\lambda_j - \lambda_l}\right) \\ &\quad + \sum_{i,j=1}^{l-1} \left\{\left(\frac{\lambda_l - \bar{\lambda}_l}{\lambda_j - \lambda_l} P_l + \mathbb{1}\right) \frac{M_i \tilde{\gamma}_{ij} M_j^\dagger}{\lambda - \lambda_j}\right\}.\end{aligned} \tag{3.28}$$

To prove the equality (3.21), it is thus sufficient to verify the relations

$$(\lambda_l - \bar{\lambda}_l) P_l \left(\mathbb{1} - \sum_{i,j=1}^{l-1} \frac{M_i \tilde{\gamma}_{ij} M_j^\dagger}{\lambda_j - \lambda_l}\right) = \sum_{i=1}^{l} M_i \gamma_{il} M_l^\dagger \tag{3.29}$$

and

$$\left(\mathbb{1} + \frac{\lambda_l - \bar{\lambda}_l}{\lambda_j - \lambda_l} P_l\right) \sum_{i=1}^{l-1} M_i \tilde{\gamma}_{ij} = \sum_{i=1}^{l} M_i \gamma_{ij}, \quad 1 \leqslant j \leqslant l-1. \tag{3.30}$$

Using Eqs. (3.4), (3.26), and the inductive hypothesis (3.27), we may replace (3.29) by

$$(\lambda_l - \bar{\lambda}_l) \hat{\chi}_{l-1}(\bar{\lambda}_l) M_l (\tilde{M}_l^\dagger \tilde{M}_l)^{-1} = \sum_{i=1}^{l} M_i \gamma_{il}. \tag{3.31}$$

Since Γ is assumed nonsingular, the set of relations (3.30), (3.31) are equivalent to the following, obtained by multiplying on the right by Γ:

$$\begin{aligned}&\sum_{i,j=1}^{l-1} \left(\mathbb{1} + \frac{\lambda_l - \bar{\lambda}_l}{\lambda_j - \lambda_l} P_l\right) M_i \tilde{\gamma}_{ij} \Gamma_{jk} \\ &\quad + (\lambda_l - \bar{\lambda}_l) \hat{\chi}_{l-1}(\bar{\lambda}_l) M_l (\tilde{M}_l^\dagger \tilde{M}_l)^{-1} \Gamma_{lk} = M_k, \\ &1 \leqslant k \leqslant l-1.\end{aligned} \tag{3.32}$$

The lhs of Eq. (3.32) may be written as

$$\begin{aligned}&\left\{\sum_{i,j=1}^{l}\left[-\frac{M_i \gamma_{ij} M_j^\dagger}{\bar{\lambda}_k - \lambda_j} + \frac{\lambda_l - \bar{\lambda}_l}{\bar{\lambda}_k - \lambda_l} P_l \right.\right. \\ &\qquad \left. \times \left(\frac{M_i \gamma_{ij} M_j^\dagger}{\lambda_l - \lambda_j} - \frac{M_i \gamma_{ij} M_j^\dagger}{\bar{\lambda}_k - \lambda_j}\right)\right] \\ &\qquad \left. + \frac{(\lambda_l - \bar{\lambda}_l)}{(\lambda_l - \bar{\lambda}_k)} P_l \hat{\chi}_{l-1}(\lambda_l)\right\} M_k \\ &= \left\{\mathbb{1} - \left(\mathbb{1} + \frac{\lambda_l - \bar{\lambda}_l}{\bar{\lambda}_k - \lambda_l} P_l\right) \hat{\chi}_{l-1}(\bar{\lambda}_k)\right\} M_k\end{aligned} \tag{3.33}$$

by the inductive hypothesis. However, the second term in the above vanishes since, for $1 \leqslant k \leqslant l-1$,

$$\begin{aligned}\hat{\chi}_{l-1}(\bar{\lambda}_k) M_k &= \chi_{l-1}(\bar{\lambda}_k) \cdots \chi_k(\bar{\lambda}_k) \chi_{k-1}(\bar{\lambda}_k) M_k \\ &= \chi_{l-1}(\bar{\lambda}_k) \cdots (1 - P_k) \tilde{M}_k \\ &= 0\end{aligned} \tag{3.34}$$

and, for $k = l - 1$, (3.32) becomes

$$\{\mathbb{1} - (\mathbb{1} - P_l) \chi_{l-1}(\bar{\lambda}_l)\} M_l = M_l, \tag{3.35}$$

proving the relation stated.

We turn now to the problem of the invertibility of Γ. As above, the proof is inductive. When Γ consists only of one block Γ_{11}, it is invertible since $\Gamma = \Gamma_{11} = (M_1^\dagger M_1)/(\lambda_1 - \bar{\lambda}_1)$ and M_1 is constructed from independent column vectors. Suppose now that $\tilde{\Gamma}$, the $(\kappa - k_l) \times (\kappa - k_l)$ submatrix of Γ built from the blocks Γ_{ij}, $1 \leqslant i,j \leqslant l-1$, is invertible. (Implicitly, we suppose that $\lambda_i \neq \bar{\lambda}_j$ for all $i \neq j$, $1 \leqslant i,j \leqslant l-1$.) We suppose further that each P_i ($1 \leqslant i \leqslant l$) is of rank 1. Since we allow possibly degenerate λ_i's (as long as the spaces spanned by the associated m_i's are all linearly independent); this involves no loss of generality. Thus, $k_i = 1$, $1 \leqslant i \leqslant l$, and $\tilde{\Gamma}$ and Γ are, respectively, $(l-1) \times (l-1)$ and $l \times l$ matrices. Since the m_i's are all $m \times 1$ matrices (i.e., column vectors), $\Gamma_{ij} = (M_i^\dagger M_j)/(\lambda_i - \bar{\lambda}_j)$ are simply complex numbers. The determinant of Γ is expressible in terms of the determinant of $\tilde{\Gamma}$:

$$\begin{aligned}\det \Gamma &= \det \tilde{\Gamma} \cdot \frac{M_l^\dagger M_l}{\lambda_l - \bar{\lambda}_l} + \sum_{i,j=1}^{l-1} \left(\frac{M_l^\dagger M_j}{\lambda_l - \bar{\lambda}_j}\right) \\ &\quad \times \left(\frac{M_i^\dagger M_l}{\lambda_i - \bar{\lambda}_l}\right) \mathrm{co}(\tilde{\Gamma}_{ij}),\end{aligned} \tag{3.36}$$

where $\mathrm{co}(\tilde{\Gamma}_{ij})$ denotes the cofactor of $\tilde{\Gamma}_{ij}$, namely $(-1)^{i+j+1}$ times the determinant of the $(l-2)\times(l-2)$ matrix obtained by removing the ith row and the jth column from $\tilde{\Gamma}$. But by definition of $\tilde{\gamma}$ as the inverse of $\tilde{\Gamma}$, these cofactors are related to the $\tilde{\gamma}_{ij}$'s by

$$\tilde{\gamma}_{ij} = \mathrm{co}(\tilde{\Gamma}_{ij})/\det\tilde{\Gamma}. \tag{3.37}$$

(By hypothesis, $\det\tilde{\Gamma}\neq 0$.) Thus $\det\Gamma$ is

$$\det\Gamma = \frac{\det\tilde{\Gamma}}{\lambda_l-\bar{\lambda}_l}\cdot M_l^\dagger \times\left\{\mathbb{1} - \sum_{i,j=1}^{l-1}\frac{M_j\tilde{\gamma}_{ji}M_i^\dagger(\lambda_l-\bar{\lambda}_l)}{(\lambda_l-\bar{\lambda}_j)(\lambda_i-\bar{\lambda}_l)}\right\}M_l. \tag{3.38}$$

We claim that the $n\times n$ matrix between the curly brackets is $\hat{\chi}_{l-1}^\dagger(\bar{\lambda}_l)\hat{\chi}_{l-1}(\bar{\lambda}_l)$. Indeed, since $\tilde{\Gamma}$ is invertible, the relation (3.21) can be used for $(l-1)$ solutions and

$$\begin{aligned}\hat{\chi}_{l-1}^\dagger(\bar{\lambda}_l)\hat{\chi}_{l-1}(\bar{\lambda}_l) &= \left(\mathbb{1}+\sum_{k,m=1}^{l-1}\frac{M_m\tilde{\gamma}_{km}M_k^\dagger}{\lambda_l-\bar{\lambda}_m}\right)\times\left(\mathbb{1}+\sum_{i,j=1}^{l-1}\frac{M_j\tilde{\gamma}_{ji}M_i^\dagger}{\bar{\lambda}_l-\lambda_i}\right)\\ &= \mathbb{1}-\sum_{i,j=1}^{l-1}\frac{M_j\tilde{\gamma}_{ji}M_i^\dagger}{\lambda_l-\bar{\lambda}_j}+\sum_{i,j=1}^{l-1}\frac{M_j\tilde{\gamma}_{ji}M_i^\dagger}{\bar{\lambda}_l-\lambda_i}\\ &\quad+\sum_{i,j,k,m=1}^{l-1}\frac{M_m\tilde{\gamma}_{mk}\tilde{\Gamma}_{kj}\tilde{\gamma}_{ji}(\lambda_k-\bar{\lambda}_j)M_i^\dagger}{(\lambda_l-\bar{\lambda}_m)(\lambda_i-\bar{\lambda}_l)},\end{aligned} \tag{3.39}$$

where we have used the fact that $\tilde{\Gamma}^\dagger = -\tilde{\Gamma}$ and hence $\tilde{\gamma}^\dagger = -\tilde{\gamma}$. Since

$$\sum_{k=1}^{l-1}\tilde{\gamma}_{mk}\tilde{\Gamma}_{kj} = \delta_{mj} \quad\text{and}\quad \sum_{j=1}^{l-1}\tilde{\Gamma}_{kj}\tilde{\gamma}_{ji} = \delta_{ki}, \tag{3.40}$$

the last term of (3.39) can be simplified and we have

$$\begin{aligned}\hat{\chi}_{l-1}^\dagger(\bar{\lambda}_l)\hat{\chi}_{l-1}(\bar{\lambda}_l) &= \mathbb{1}+\sum_{i,j=1}^{l-1}\left\{M_j\tilde{\gamma}_{ji}M_i^\dagger\times\left[\frac{1}{\bar{\lambda}_j-\lambda_l}+\frac{1}{\bar{\lambda}_l-\lambda_i}+\frac{\lambda_i-\bar{\lambda}_j}{(\lambda_l-\bar{\lambda}_j)(\lambda_i-\bar{\lambda}_l)}\right]\right\}\\ &= \mathbb{1}-\sum_{i,j=1}^{l-1}\frac{(\lambda_l-\bar{\lambda}_l)M_j\tilde{\gamma}_{ji}M_i^\dagger}{(\lambda_l-\bar{\lambda}_j)(\lambda_i-\bar{\lambda}_l)}.\end{aligned} \tag{3.41}$$

Hence

$$\det\Gamma = [\det\Gamma'/(\lambda_l-\bar{\lambda}_l)]\cdot M_l^\dagger\hat{\chi}_{l-1}^\dagger(\bar{\lambda}_l)\hat{\chi}_{l-1}(\bar{\lambda}_l)M_l. \tag{3.42}$$

The problem of the invertibility of Γ amounts then to the fact that the vector $\hat{\chi}_{l-1}(\bar{\lambda}_l)M_l$ is not identically zero. If none of the λ_i's, $1\leqslant i\leqslant l-1$, is equal to λ_l, then $\hat{\chi}_{l-1}(\bar{\lambda}_l)$ is nonsingular: each of the parentheses in the following expression

$$\hat{\chi}_{l-1}(\bar{\lambda}_l) = \left(\mathbb{1}+\frac{\lambda_{l-1}-\bar{\lambda}_{l-1}}{\bar{\lambda}_l-\lambda_{l-1}}P_{l-1}\right)\cdots\left(\mathbb{1}+\frac{\lambda_2-\bar{\lambda}_2}{\bar{\lambda}_l-\lambda_2}P_2\right)\left(\mathbb{1}+\frac{\lambda_1-\bar{\lambda}_1}{\bar{\lambda}_l-\lambda_1}P_1\right) \tag{3.43}$$

is a regular matrix. However, in the case where λ_l happens to be equal to some of the other λ_i's, $\hat{\chi}_{l-1}(\bar{\lambda}_l)$ is no longer a regular matrix. For example, if i is such that $\lambda_i = \lambda_l$, then

$$\mathbb{1}+[(\lambda_i-\bar{\lambda}_i)/(\bar{\lambda}_l-\lambda_i)]P_i = \mathbb{1}-P_i \tag{3.44}$$

is singular. In this case, the vector

$$\hat{\chi}_{l-1}(\bar{\lambda}_l)M_l = \left(\prod_{j=i+1}^{l-1}\chi_j(\bar{\lambda}_l)\right)(\mathbb{1}-P_i)\hat{\chi}_{i-1}(\bar{\lambda}_l)M_l \tag{3.45}$$

will be zero if

$$(\mathbb{1}-\mathrm{P}_i)\hat{\chi}_{i-1}(\bar{\lambda}_l)M_l = 0. \tag{3.46}$$

Multiplying by $\chi_{i-1}(\bar{\lambda}_l)^{-1}$, we obtain

$$(\mathbb{1}-\hat{\chi}_{i-1}(\bar{\lambda}_l)^{-1}P_i\hat{\chi}_{i-1}(\bar{\lambda}_l))M_l = 0, \tag{3.47}$$

that is,

$$M_l - M_i(\tilde{M}_i^+\tilde{M}_i)^{-1}\tilde{M}_i^+\tilde{M}_l = 0, \tag{3.48}$$

which is a dependency between M_l and M_i which contradicts the fact that m_i and m_l are linearly independent. If several λ_i's are equal to λ_l, the same argument shows that Γ will be invertible if and only if the different m_i's associated with equal λ_i's are linearly independent. The proposition is thus proven.

A corollary to this result is the fact that if any pair of parameters $\{\lambda_i\}$ is interchanged, together with the corresponding pair of input data $\{m_i\}$, the iterated Bäcklund transformations give the same result. Thus, for any sequence with parameters $\{\lambda_i\}$, another sequence exists for any permutation of the parameters which gives rise to the same result. This "permutability theorem" is precisely analogous to those known in other integrable systems, e.g., the sine–Gordon equation.[12]

The general multisoliton solution with parameters $\{\lambda_i\}$ and input data $\{m_i\}$ is obtained by substituting the single-soliton solutions determined by[5]:

$$g_0 = \exp(A_0\xi + B_0\eta) \tag{3.49}$$

$$M_i = \exp\left(\frac{A_0\xi}{1+\lambda_i}+\frac{B_0\eta}{1-\lambda_i}\right)m_i \tag{3.50}$$

$$A_0, B_0 = \text{const},\quad [\mathrm{A}_0, \mathrm{B}_0] = 0 \tag{3.51}$$

in (3.16) and (3.21). A more detailed analysis of the structure of such solutions will be the subject for further study.

ACKNOWLEDGMENTS

Part of this work was completed while one author (Steve Shnider) was at the Weizmann Institute of Science, Department of Mathematics and Physics. He wishes to thank his hosts for their kind hospitality. A scholarship from the Fonds F.C.A.C. (Ministère de l'Education du Québec) is gratefully acknowledged by Yvan Saint-Aubin.

[1] A. S. Bugadov and L. A. Takhtadzhyan, Dokl. Akad. Nauk SSSR **22**, 805 (1977) [Sov. Phys. Dokl. **22**, 428 (1977)].
[2] K. Pohlmeyer, Commun. Math. Phys. **46**, 207 (1976).
[3] M. Lüscher, Nucl. Phys. B **135**, 1 (1978).
[4] M. Lüscher and K. Pohlmeyer, Nucl. Phys. B **137**, 46 (1978).
[5] V. E. Zakharov and A. V. Mikhaĭlov, Zh. Eksp. Teor. Fiz. **74**, 1953 (1978) [Sov. Phys. JETP **47**, 1017 (1979)].
[6] V. E. Zakharov and A. B. Shabat, Funk. Anal. Pril. **13**, 13 (1979) [Func.

Anal. Appl. **13**, 166 (1979)].
[7]V. E. Zakharov and A. V. Mikhaĭlov, Pis'ma Zh. Eksp. Teor. Fiz. **27**, 47 (1978) [JETP Lett. **27**, 42 (1978)].
[8]H. Eichenherr, Nucl. Phys. B **146**, 215 (1978); Phys. Lett. B **90**, 121 (1980).
[9]H. Eichenherr and M. Forger, Nucl. Phys. B **155**, 381 (1979); B **164**, 528 (1980).
[10]A. T. Ogielski, M. K. Prasad, A. Sinha, and L. L. C. Wang, Phys. Lett. B **91**, 387 (1980).
[11]J. Harnad, S. Shnider, and Yvan Saint-Aubin, "Quadratic Pseudo-Potentials for Principal Sigma Models," CRMA 1075, 1982, Physica D (in press).
[12]R. M. Miura, Ed., *Bac̆klund Transformations*, Lecture Notes in Mathematics **515** (Springer-Verlag, Berlin, 1976).
[13]J. Harnad, P. Winternitz, and R. L. Anderson, J. Math. Phys. **24**, 1062 (1983).

Group theory of the dipole ghost

W. Heidenreich
Max–Planck Institut, D-813 Starnberg, Federal Republic of Germany

(Received 20 July 1983; accepted for publication 9 September 1983)

The conformal representation carried by the dipole ghost $\Box\Box\phi = 0$ is described in a discrete basis; it has dimension zero and contains the constant solution as an irreducible subspace. A free quantum field operator is constructed. The representation is decomposed with respect to the (3.2) de Sitter and the Poincaré groups. A model for coupling relativistically to an external source is given, in which only unitary mass 0 helicity 0 modes propagate.

PACS numbers: 11.10.Qr, 02.20. + b, 11.30.Cp

1. INTRODUCTION

Fourth-order field equations appeared recently in the gauge sector of conformal QED,[1] in renormalizable and conformal theories of gravity,[2,3] and are proposed for confining theories of strong interactions.[4] Progress in the application to gravity and strong interactions is hampered by the appearance of nonunitary ghosts in higher-order theories. So a careful approach is necessary, and taking the possible benefits into consideration, worthwhile.

A model for fourth-order theories is the dipole ghost[5] $\Box\Box\phi = 0$. It will be treated here with powerful group theoretic techniques: an equivalent formulation in conformal space is employed to find explicitly the (nonunitary) conformal representation carried by the dipole ghost. It suggests a conformally invariant indefinite metric field quantization.

The conformal representation is reduced to the (3.2) de Sitter and the Poincaré groups. It is possible to couple the dipole ghost to an external source in such a way that only a unitary representation of these groups propagates.

2. CONFORMAL PROPERTIES OF THE DIPOLE GHOST

Conformal space

Conformal space is a four-dimensional projective space

$$x^2 \equiv \eta^{ab} x_a x_b \equiv x_1^2 + x_2^2 + x_3^2 - x_4^2 + x_5^2 - x_6^2 = 0, \tag{1}$$

$$x_a \hat{=} \lambda x_a, \quad \lambda \neq 0.$$

It is a compactification of Minkowski space y_ν, $\nu = 1,...,4$. The two coordinate systems are connected by

$$y_\nu = x_\nu/(x_5 + x_6), \quad y_+ = x_5 + x_6, \quad B = x^2/(x_5 + x_6)^2; \tag{2}$$

$$B = 0, \quad y_+ \hat{=} \lambda y_+.$$

The dipole ghost in conformal space is a scalar field ϕ with degree of homogeneity 0,

$$(x\partial)\phi = 0, \tag{3}$$

and the field equation

$$(\partial^2)^2\phi = 0. \tag{4}$$

The operator $(\partial^2)^2$ is defined intrinsically only if (3) holds. Then it is in Minkowski coordinates

$$(\partial^2)^2 = y_+^{-4}\Box^2, \tag{5}$$

with

$$\Box = \left(\frac{\partial}{\partial y_1}\right)^2 + \left(\frac{\partial}{\partial y_2}\right)^2 + \left(\frac{\partial}{\partial y_3}\right)^2 - \left(\frac{\partial}{\partial y_4}\right)^2.$$

A basis of the Lie algebra of the conformal group $SO(4.2)/Z_2$ is for scalar fields

$$M_{ab} = -i(x_a\partial_b - x_b\partial_a). \tag{6}$$

Positive energy representations have a unique lowest weight in the decomposition with respect to the maximal compact subgroup $(SO(2)\times SO(4))/Z_2$. They are labeled by the eigenvalues of conformal energy M_{46} and the angular momenta of $SU(2)\times SU(2) \approx SO(4)$. The noncompact generators are conveniently grouped into energy raising and lowering operators

$$M_i^+ = M_{i6} + iM_{i4} = -2x_i\frac{\partial}{\partial(x_4 + ix_6)} - (x_4 - ix_6)\frac{\partial}{\partial x_i}, \tag{7}$$

$$M_i^- = M_{i6} - iM_{i4} = 2x_i\frac{\partial}{\partial(x_4 - ix_6)} + (x_4 + ix_6)\frac{\partial}{\partial x_i},$$

$$i = 1,2,3,5.$$

Solutions

A solution of Eqs. (3) and (4) is

$$\phi_0 = 1. \tag{8}$$

All Lie algebra elements give zero, $M_{ab}\phi_0 = 0$. So it is an irreducible trivial representation $D(0,0,0)$ of the conformal group.

Another solution with positive energy is

$$\phi_1 = x_i/(x_4 + ix_6). \tag{9}$$

It carries a $SO(2)\times SU(2)\times SU(2)$ representation (weight) $(1,\frac{1}{2},\frac{1}{2})$. Acting with the lowering operators M_j^- gives the constant solution. A relative lowest weight ϕ_1 leaks into the irreducible trivial representation. The positive energy modes of the dipole ghost carry the indecomposable representation

$$D(1,\tfrac{1}{2},\tfrac{1}{2}) \rightarrow D(0,0,0). \tag{10}$$

A basis of this solution space will be given later. In principle it could be obtained by acting with all polynomials of M_i^+ on the states (9).

 0022-2488/84/020376-04$02.50

The constant solution is the "pure gauge" of the solution space. For the construction of the conformal quantum field operator of the dipole ghost we need a Gupta–Bleuler triplet. The conjugate of the constant gauge field must be another trivial representation ϕ_s ("scalar field"), which leaks into the "physical" $D(1,\frac{1}{2},\frac{1}{2})$,

$$D(0,0,0) \rightarrow D(1,\tfrac{1}{2},\tfrac{1}{2}) \rightarrow D(0,0,0). \tag{11}$$

The conditions $M^-\phi_s = 0$ and $M^+\phi_s = \phi_1$ can only be satisfied (up to a factor and an additive constant) by

$$\phi_s = \ln(x_4 + ix_6). \tag{12}$$

The compact generators M_{46}, M_{ij} map this state into a linear combination of itself and the constant. ϕ_s fulfills the field equation $(\partial^2)^2\phi = 0$, but not Eq. (3),

$$(x\partial)\phi_s = 1. \tag{13}$$

So, Eq. (3) is the "Lorentz condition," which projects on the "physical" and the "gauge" solutions. The field equations which hold for the full Gupta–Bleuler triplet are Eq. (4) and

$$(x\partial)^2\phi = 0. \tag{14}$$

ϕ_s is an adjoint homogeneous function of degree 0. It is continuous on the universal covering of conformal space.

Homogeneous propagator

On conformal space, the only invariant regular distribution of x_a and $|x'_a|$ is $|xx'|^{-a}$. The limit $a \rightarrow 0$ yields 1, the homogeneous propagator of the trivial representation. To obtain the Gupta–Bleuler triplet, we can expand

$$\lim_{a\rightarrow 0} (|xx'|^{-a} - 1)/(-a) = \ln|xx'|. \tag{15}$$

This suggests for the dipole ghost of positive, respectively negative, energy appropriately-regularized distributions $\ln(xx')_\pm$. They are unique up to a factor and an additive constant, which corresponds to a change of gauge.

Decomposing the homogeneous propagators in energy- and angular-momentum eigenfunctions gives a discrete basis of the dipole ghost and its invariant (indefinite) norm. It is convenient to introduce the conformal time coordinate explicitly and use coordinates α, s_i, ρ,

$$x_i = \rho s_i,\ x_4 = \rho \sin\alpha,\ x_6 = \rho\cos\alpha;\ \rho \hat{=} \lambda\rho,\ s^2 = 1. \tag{16}$$

Then we get with $\tau = \alpha - \alpha'$,

$$\begin{aligned} &D^\pm(x,x') \\ &\equiv 1 + \ln(-2xx')_\pm \equiv 1 + \ln[\rho\rho'(2\cos(\tau \pm i\epsilon) - 2ss')] \\ &= 1 + \ln|\rho\rho'| \mp i\tau + \ln(1 - 2ss'e^{\pm i(\tau\pm i\epsilon)} + e^{\pm 2i(\tau \pm i\epsilon)}). \end{aligned} \tag{17}$$

The last logarithm is a generating function of the Gegenbauer polynomials $C_n^{(0)}$,

$$\ln(\cdots) = -\sum_{n=0}^{\infty} C_n^{(0)}(ss')e^{\pm in\tau}.$$

Using $C_n^{(0)} = (1/n)(C_n^{(1)} - C_{n-2}^{(1)})$ and the addition theorem for four-dimensional spherical harmonics, we get

$$\begin{aligned} D^+(x,x') = \ln|\rho\rho'| - i\tau - \sum_{j=1}^{\infty} \frac{2\pi^2}{j(j+1)} &\times \sum_{l,m} Y_{jlm}(s)Y_{jlm}(s')e^{ij\tau} \\ &+ \sum_{j=0}^{\infty} \frac{2\pi^2}{(j+2)(j+1)} \sum_{l,m} Y_{jlm}(s)Y_{jlm}(s')e^{i(j+2)\tau}. \end{aligned} \tag{18}$$

The Y are normalized by

$$\int d^3\Omega\, Y_{jlm}(\Omega)Y_{j'l'm'}(\Omega) = \delta_{jj'}\delta_{ll'}\delta_{mm'},$$

and $j \geqslant l \geqslant |m|$.

Expressed in the basis states

$$\begin{aligned} &s(x) = \ln(x_4 + ix_6) = \ln\rho + \ln(ie^{-i\alpha}), \\ &g(x) = 1, \\ &p_{jlm}(x) = \sqrt{2}\pi/(j(j+1))^{1/2} Y_{jlm} e^{ij\alpha}, \\ &q_{jlm}(x) = \sqrt{2}\pi/((j+2)(j+1))^{1/2} Y_{jlm} e^{i(j+2)\alpha}, \end{aligned} \tag{19}$$

we have

$$\begin{aligned} D^+(x,x') &= s^*(x')g(x) + g^*(x')s(x) \\ &\quad - \Sigma p^*(x')p(x) + \Sigma q^*(x')q(x). \end{aligned} \tag{20}$$

The "scalar" and the "gauge" modes are s and g; p and q are a basis of the nonunitary $D(1,\frac{1}{2},\frac{1}{2})$. The weight diagram of the solution space (19) is given in Fig. 1. Compared with a general spin-0 representation it is very restricted. This is possible as $D(4,0,0)$ is Weyl-equivalent to $D(0,0,0)$.

The relative sign between the two sums in the decomposition (18) shows explicitly that the invariant norm of $D(1,\frac{1}{2},\frac{1}{2})$ is indefinite.

Quantum field operator

We can define a free quantum field operator

$$\begin{aligned} \psi &= a_s s + a_g g + \Sigma a_n p_n + \Sigma b_n q_n \\ &\quad + a_s^+ s^* + a_g^+ g^* + \Sigma a_n^+ p_n^* + \Sigma b_n^+ q_n^* \\ &= \psi^{(+)} + \psi^{(-)}. \end{aligned} \tag{21}$$

The invariant commutation relation

$$[\psi^{(-)}(x), \psi^{(+)}(x')] = -D^+(x,x') \tag{22}$$

holds if the operators a, a^+ in the discrete-momentum basis fulfill the nonvanishing commutators

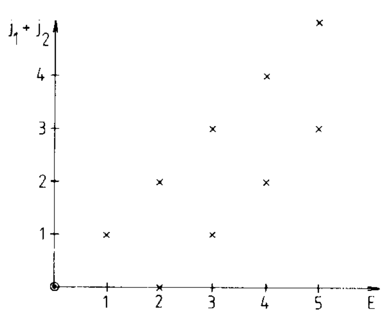

FIG. 1. Weight diagram of the dipole ghost. $\times$ denotes the weights of $D(1,\frac{1}{2},\frac{1}{2})$, dot and circle the weights of "gauge" and "scalar" modes.

$$[a_g,a_s^+]=1,\ [a_s,a_g^+]=1,\ [a_n,a_{n'}^+]=+\delta_{nn'},$$
$$[b_n,b_{n'}^+]=-\delta_{nn'}. \tag{23}$$

An indefinite many-dipole ghosts-space is obtained by acting with the operators a^+, b^+ on a vacuum $|0\rangle$,

$$a_s|0\rangle=a_g|0\rangle=a_{jlm}|0\rangle=b_{jlm}|0\rangle=0. \tag{24}$$

The a-modes have positive, but the b-modes have negative probability. A theory with only positive propagating modes is not conformally invariant. This does not exclude the possibility that a physical interpretation becomes possible in theories with a smaller symmetry group.

Therefore, I decompose the dipole ghost with respect to the (3.2) de Sitter and the Poincaré groups.

3. de SITTER PROPERTIES OF THE DIPOLE GHOST

The weight diagram of the dipole ghost (Fig. 1) can easily be decomposed in a diagram with respect to SO(2)×SO(3), the maximal-compact subgroup of the de Sitter group SO(3.2). Collecting de Sitter representations gives

$$D(0,0,0)\to D(1,\tfrac12,\tfrac12)\to D(0,0,0)|_{SO(3.2)}$$
$$=(D(0,0)\to D(1,1)\to D(0,0))\oplus D(1,0)\oplus D(2,0). \tag{25}$$

The first part is a Gupta–Bleuler triplet with a nonunitary $D(1,1)$; it is the dipole ghost in (2 + 1)-dimensional Minkowski space. The other part describes a massless particle. The representations in the first part have a 2nd-order Casimir eigenvalue $Q=0$, the massless representations have $Q=-2$.

The field equation of the dipole ghost in (3.2) de Sitter space is

$$\hat{Q}(\hat{Q}+2)\phi=0. \tag{26}$$

4. THE DIPOLE GHOST IN MINKOWSKI SPACE

The "pure gauge" of the dipole ghost is the trivial representation $g=1$. In Minkowski space it is $g(y_\nu)=1$, a trivial representation of the Poincaré group. The "scalar" state $s(x)$ becomes

$$s'(y_\nu,y_+)=\ln|y_+|+\ln(y_4+\tfrac12 i(1+y^2)). \tag{27}$$

y_+ is invariant under Poincaré transformations, but not under dilations D and special conformal transformations K_ν. The generators of the Lorentz group and translations are explicitly

$$M_{\mu\nu}=-i(y_\mu\partial_\nu-y_\nu\partial_\mu),\quad P_\nu=-i\partial_\nu, \tag{28}$$

and

$$D=-i(y_\nu\partial^\nu-y_+\partial_+),$$
$$K_\nu=-i(2y_\nu(y_\mu\partial^\mu-y_+\partial_+)-y^2\partial_\nu). \tag{29}$$

Acting on $s'(y_\nu,y_+)$ we have

$$y_+\partial_+s'(y_\nu,y_+)=1, \tag{30}$$

while for all other states of the dipole ghost

$$y_+\partial_+\psi(y_\nu,y_+)=0. \tag{31}$$

The field has dimension 0.

Now we can define

$$s(y_\nu)\equiv s'(y_\nu,y_+)-\ln|y_+|, \tag{32}$$

on which, e.g., the dilatation operator acts like $Ds=-i(y_\nu\partial^\nu s-1)$.

The existence of these two trivial representations g and s in the dipole ghost complicates the infrared limit in an expansion in plane waves. Specifically $s(y_\nu)$ is not a momentum eigenfunction, but leaks into $D(1,\frac12,\frac12)$.

Except for the limit $p_\nu=0$ we get a plane-wave basis of the dipole ghost by Fourier transformation of the homogeneous propagator (17).

In Minkowski coordinates it is

$$D^\pm(x,x')=\ln(y_+y'_+)+1+\ln((y-y')^2_\pm). \tag{33}$$

The first term comes from $s'(y_\nu,y_+)$ and can be subtracted as in Eq. (32). Then we have

$$D^\pm(y_\nu,y'_\nu)\equiv D^\pm(x,x')-\ln(y_+y'_+)$$
$$=1+\ln((\mathbf{y}-\mathbf{y}')^2-(t-t'\pm i\epsilon)^2). \tag{34}$$

The Pauli–Jordan commutation function is

$$D(y_\nu)=D^+(y_\nu)-D^-(y_\nu)=-\epsilon(t)\theta(-y^2). \tag{35}$$

This expression was first given by Narnhofer and Thirring.[5] Using the Fourier transformation of $\ln(x+i0)$[6] we get

$$\int e^{iky}D^+(y)d^4y$$
$$=(2\pi)^3/k\,(\theta(k_0)/k_0+\ln|k_0|\delta(k_0)-\gamma\delta(k_0))$$
$$\times(\delta'(k+k_0)-\delta'(k_0-k))-4\pi^3/$$
$$k\delta(k_0)((k_0+k)^{-2}-(k_0-k)^{-2})+\delta^4(k)$$
$$=-(2\pi)^3/$$
$$(k_0k)\theta(k_0)\delta'(k_0-k)+\text{terms at }k_0=0. \tag{36}$$

For $k_0\neq0$ we have the "plane wave" expansion

$$D^+(y)|_{k_0\neq0}=(2\pi)^{-1}\int e^{i(kt-\mathbf{k}\mathbf{y})}(itk^{-2}-k^{-3})d^3\mathbf{k}. \tag{37}$$

(Invariance under finite Lorentz transformations can be shown by converting terms of the form $\mathbf{y}e^{\cdots}$ into $\partial/\partial\mathbf{k}e^{\cdots}$ and partial integration.)

A possible set of "plane waves" for the dipole ghost at $k_0\neq0$ is

$$|\mathbf{k}\rangle=k^{-1/2}e^{i(kt-\mathbf{k}\mathbf{y})},$$
$$|\mathbf{k}'\rangle=(itk^{-3/2}-k^{-5/2})e^{i(kt-\mathbf{k}\mathbf{y})}. \tag{38}$$

The mass operator $M^2=-\Box$ gives

$$\Box|\mathbf{k}\rangle=0,$$
$$\Box|\mathbf{k}'\rangle=k^{-1/2}e^{i(kt-\mathbf{k}\mathbf{y})}=|\mathbf{k}\rangle,$$

and then $\Box^2|\mathbf{k}'\rangle=0$.

The states $|\mathbf{k}\rangle$ carry a mass 0 helicity 0 representation of the Poincaré group. The states $|\mathbf{k}'\rangle$ leak into $|\mathbf{k}\rangle$ and have mass 0, helicity 0 also. Taking this information at $k_0\neq0$ together with the infrared states s and g, we find the Poincaré representation carried by the dipole ghost:

$$T \to D(m = 0, \lambda = 0) \to D(m = 0, \lambda = 0) \to T. \tag{39}$$

Here T is the trivial representation.

To repeat: "plane wave" states $|\mathbf{k}\rangle$ and $|\mathbf{k}'\rangle$ leak into the constant solution g. This is an irreducible trivial representation. To quantize with the Gupta–Bleuler technique, we need a "scalar state" $s(y_\nu)$. This detailed study of the infrared behavior of the dipole ghost was only possible in the discrete basis.

It is feasible to interpret the invariant subspace $D(m = 0, \lambda = 0) \to T$ in Eq. (39) as a gauge freedom, like the pure gauge in electrodynamics.[7] I refrain from this possibility as the dipole ghost would contain "gauge" and "scalar" modes only, without any "physical" states.

5. UNITARY THEORY WITH THE DIPOLE GHOST

The "nonunitarity" of higher-order field theories shows itself group theoretically in the appearance of nonunitary representations. Yet this by itself does not induce a nonunitary S-matrix, as can be seen in ordinary relativistic electrodynamics. In Gupta–Bleuler quantization, the Poincaré representation carried by the field potentials is indecomposable and thus nonunitary. Nevertheless the resulting S-matrix is unitary: the propagating modes carry a unitary representation.

In a relativistic interacting theory with the dipole ghost, we could have the Lagrangian

$$L = \psi^* \Box^2 \psi + \psi j, \tag{40}$$

where j is an external scalar source. The resulting field equation

$$\Box^2 \psi = j \tag{41}$$

propagates only the unitary mass and helicity zero modes if j is of the form

$$j \equiv \Box s. \tag{42}$$

Then $\Box\psi - s$ is a free field,

$$\Box(\Box\psi - s) = 0. \tag{43}$$

The initial condition $(\Box\psi - s)(t = -\infty) = 0$ is true for all times due to the field equation. So $\Box\psi = 0$ vanishes wherever s vanishes, i.e., in empty space the ghost does not propagate.

A similar technique for exorcising a nonunitary ghost was used in conformal gravity.[3] As the field has dimension 0, any polynomial self-interaction requires a dimensional coupling constant.

ACKNOWLEDGMENTS

I thank C. Fronsdal and L. Castell for discussions and criticism. I further thank the Physics Dept. of UCLA, where part of this work was done, for hospitality and the DAAD for a fellowship.

[1] B. Binegar, C. Fronsdal, and W. Heidenreich, "Conformal QED," J. Math. Phys. **24**, 2828 (1983).

[2] K. S. Stelle, "Classical gravity with higher derivatives," Gen. Relativ. Gravit. **9**, 353–371 (1978).

[3] B. Binegar, C. Fronsdal, and W. Heidenreich, "Linear conformal quantum gravity," Phys. Rev. D **27**, 2249–2261 (1983).

[4] S. K. Kaufman, "Quarks and glue: equations of motion," Nucl. Phys. B **87**, 133–144 (1975).

[5] N. Narnhofer and W. Thirring, "The taming of the dipole ghost," Phys. Lett. B **76**, 428–432 (1978).

[6] I. M. Gel'fand and G.E. Shilov, *Generalized Functions*, Vol. 1 (Academic, New York, 1964).

[7] A. Z. Capri, G. Grübl, and R. Kobes, *Fock Space Construction of the Massless Dipole Field*, Preprint Alberta-Thy-9-82.

Unified gauge theory for electromagnetism and gravitation based on twistor bundles

C. P. Luehr and M. Rosenbaum[a)]
Centro de Estudios Nucleares, Universidad Nacional Autonoma de México, Circuito Exterior, C. U. 04510 México, D.F., Mexico

(Received 28 December 1982; accepted for publication 23 September 1983)

A unified gauge theory of the combined gravitational and electromagnetic fields is obtained by two different procedures using twistors as a starting point for the construction of the appropriate bundles. One of these formalisms is obtained by relaxing the conditions on the structure of a twistor bundle theory previously developed by the authors for the Poincaré group as the structure group. The other formalism is based on a tensor product bundle and can be readily extended to include structure groups involving direct products of nonabelian groups with the Poincaré group. The results of the theory are compared with those obtained in projective theories of the generalized Jordan–Kaluza–Klein type, and some of the essential differences are pointed out.

PACS numbers: 11.15. − q, 04.50. + h, 04.40. + c

I. INTRODUCTION

With the growing importance of gauge fields in physics and the need to unify them with general relativity, theories of the Kaluza–Klein type[1] and their generalization to nonabelian gauge fields have acquired renewed interest.

Some of the modern approaches[2] to this subject advocate the idea that the extra dimensions which are introduced in these theories should be regarded as microscopic and physically real new degrees of freedom, and a suggestive attempt to combine "spontaneous compactification" and supergravity, based on an 11-dimensional theory, has been presented by Witten.[3] An extensive review of the recent work on Kaluza–Klein theories has been given by Salam and Strathdee.[4]

Also, since the differential geometric methods of fiber bundles provide a very convenient framework for a coordinate-free discussion of gauge theories[5–7] several papers have appeared lately[8–10] where Kaluza–Klein theories are formulated in that language. All these approaches have the common feature that they start with a principal G bundle P, and a base space which is already the space-time endowed with a Riemannian metric. Thus, the structural group G is a compact Lie group which is used for gauging the additional internal degrees of freedom.

It is our belief that a truly unified field theory should result from the simultaneous gauging of the compact (internal) and noncompact (external) groups which characterize the theory. Consequently, a unified theory of gravitation and other fields should result from the simultaneous gauging of the Poincaré group and the other symmetry groups involved.

A straightforward attempt to extend the fiber bundle techniques to theories with noncompact group symmetries, such as the gauge theory of the Poincaré group, to obtain gravitational theories has led, however, to some essential difficulties.[11]

One procedure to resolve these problems has been proposed by the authors in Ref. 11, where the Poincaré group is dealt with as an internal gauge group acting on fibers of an appropriately constructed vector bundle, thus leading to an unambiguous gauge theory of gravitation. The authors have also shown[12] that twistors provide a very natural framework for the construction and geometrical interpretation of the five-dimensional representation space of the Poincaré group used as a typical fiber in Ref. 11.

The present paper is part of a program intended to combine the twistor language and the fiber bundle techniques for the purpose of constructing, in a unified manner gauge theories combining both compact (internal) and noncompact group symmetries. Specifically, we show in this paper that the twistor bundle formalism developed in Ref. 12 has already built-in the group U(1) as a normal subgroup and that, by an adequate relaxation of the conditions on the structure of the typical fiber of the bundle, unified field theories for electromagnetism and gravitation are obtained.

We further show how our procedure can be readily generalized to obtain a natural formalism, in terms of tensor product bundles, for the construction of other gauge theories where the structure group is the direct product of a nonabelian internal group with the Poincaré group.

As a continuation of our program, we plan in a future paper to exhibit a procedure, based on an enlargement of our twistor spaces, for the obtainment of supergravity and its coupling to Yang–Mills type fields.

The presentation is organized as follows:

In Sec. II we review some of the basic notation and essential features of the twistor structures and twistor bundles which were amply discussed in Ref. 12 for a construction of a gauge theory for the Poincaré group from an axiomatic, coordinate-free, and component-free point of view. The purpose of this section is to allow the paper to be as self-contained as possible without making it unduly long. Recourse to Secs. II and III of Ref. 12 will readily provide any additional material required for conceptual and notational clarifications.

In Sec. III we develop a tensor product bundle formalism which permits the construction of a gauge theory for the structure group $G = \mathrm{U}(1)\times\mathscr{P}$. Contrary to other fiber bun-

[a)] Supported in part by International Scientific Program Grants, National Science Foundation OIP75-09783A01 and Consejo Nacional de Ciencia y Technologia No. 955.

dle theories of the Kaluza–Klein type,[8–10] the fiber bundle in our theory is a vector bundle where the typical fiber is constructed as a representation space of G, and an element of this space is given by the tensor product of a twistor with a basis element of the representation space for U(1).

Some important differences which result from our approach are that there do not occur the additional Brans–Dicke scalar fields which are contained in projective theories of the generalized Jordan–Kaluza–Klein type,[13] and also there do not appear any undesirable cosmological constant terms.

Further, although the theory admits connections with nonvanishing torsion, no components of torsion associated with a fifth dimension are present, and there is no need to account for new physical effects such as those that have been predicted in other theories[8] where these additional components appear.

There is one other interesting result of our theory. It shows that the electromagnetic field does not couple to torsion. This implication is not obvious *a priori*. In fact, one of the general prescriptions of gauge theory consists of replacing derivation operators X by covariant derivatives D_X in the Lagrangian. In the case of the Einstein–Cartan theory of gravitation in the presence of charged fields, as well as other more general theories with nonzero torsion, this minimal coupling principle implies that the electromagnetic field tensor, defined as the covariant curl of the electromagnetic potential, contains torsion terms due to the nonsymmetry of the connection coefficients. As Hehl *et al.*[14] have pointed out, this definition would lead to breaking of gauge invariance under the usual gauge transformation $A_\mu \to A'_\mu = A_\mu + \partial_\mu \phi$, where ϕ is a scalar function, and therefore conclude that gauge invariance forbids the application of the minimal coupling procedure to the Maxwell field.

However, if one accepts the general principle that spinning particles both generate and react to torsion,[14] then it would appear reasonable to expect that photons should also couple to torsion. There have been some attempts[15] to make minimal coupling compatible with local gauge invariance (in a modified form), but these seem to be in disagreement with experimental evidence.[16]

The results in this paper [see in particular Eqs. (III.18)–(III.22)] provide an alternate approach to resolve the problem. They show that minimal coupling does in fact apply to the electromagnetic field, but an additional term containing torsion occurs in such a way in the expressions where the electromagnetic field tensor appears that it cancels out exactly the torsion contributions from the nonsymmetric connection coefficients, leaving only the contribution due to the standard connection (covariant derivatives expressed only in terms of Christoffel symbols); thus the gauge invariance of the electromagnetic field tensor is also preserved.

The formulation presented in Sec. III of a gauge theory for the group $G = \mathrm{U}(1) \times \mathscr{P}$ via a tensor product bundle is rather convenient for extension to other gauge theories where the structure group is a direct product of a nonabelian internal group with the Poincaré group. However, the twistor formalism developed in Ref. 12 has already built-in in the presence of U(1). Section IV of this paper is intended to show how this presence can be made explicit by an adequate relaxation of the conditions on the structure of the typical fibers.

II. SUMMARY OF THE GAUGE THEORY FOR THE POINCARÉ GROUP $\mathscr{P}$

In the formulation[12] of a gauge theory for $\mathscr{P}$, we used the twistor space $\mathscr{U} \equiv (\mathscr{U}, \mathbf{\Lambda}, \mathbf{I})$ as a representation space for $\mathscr{P}$. As part of the structure for $\mathscr{U}$, $\mathbf{\Lambda} \in \mathscr{U}^{\wedge 4}$ is a given normalized (i.e., satisfying the requirement $\hat{\mathbf{\Lambda}} {}^{\circ\circ}_{\circ\circ} \mathbf{\Lambda} = \epsilon_{\alpha\beta\gamma\delta}\epsilon^{\alpha\beta\gamma\delta} = 4!$) totally antisymmetric twist tensor $(\epsilon^{\alpha\beta\gamma\delta})$, and $\mathbf{I} \in \mathscr{E} \subset \mathscr{U}^{\wedge 2}$ is a given null element in the space $\mathscr{E}$ of real twist tensors. The element $\mathbf{I}$ is simple and invariant under the action of $\mathscr{P}$ and can be identified, therefore, with the vertex of the null cone at infinity, i.e., $\mathbf{I}$ is the so-called infinity twistor or metric twistor.

An origin element $\mathbf{O} \in \mathscr{E}$ is any chosen (simple) null element satisfying $\mathbf{O} \odot \mathbf{I} = 2$, where the inner product "$\odot$" [signature $(+\ +\ ----)$] is defined by $\mathbf{A} \odot \mathbf{B} = \frac{1}{2} A^{\alpha\beta}\epsilon_{\alpha\beta\gamma\delta} B^{\gamma\delta}$ (in the Penrose component notation) for arbitrary twist tensors $\mathbf{A}, \mathbf{B} \in \mathscr{U}^{\wedge 2}$. The space $\mathscr{F} \equiv \mathscr{W}_0 \subset \mathscr{E}$ consists of all elements orthogonal to both $\mathbf{I}$ and $\mathbf{O}$.

Starting with the four-dimensional base manifold $\mathscr{M}$, we constructed the bundles $\mathscr{U}(\mathscr{M}), \mathscr{U}^{\wedge 2}(\mathscr{M}), \mathscr{U}^{\wedge 4}(\mathscr{M})$, and $\mathscr{E}(\mathscr{M})$ with $\mathscr{U}, \mathscr{U}^{\wedge 2}, \mathscr{U}^{\wedge 4}$, and $\mathscr{E}$ as typical fibers. The cross sections $\mathbf{\Lambda} \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 4}(\mathscr{M}))$ and $\mathbf{I} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$ are given as part of the structure of $\mathscr{U}(\mathscr{M})$. To any choice of an origin twist tensor field $\mathbf{O} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$, there corresponds a unique bundle $\mathscr{F}(\mathscr{M})$ with $\mathscr{F}$ as typical fiber.[17]

A connection $\mathbf{D}$ defined on $\mathscr{U}(\mathscr{M})$ satisfies the usual axioms of a connection given in Eqs. (3.2) of Ref. 12, and also the requirements of Eqs. (3.3)–(3.5) in that paper, which we repeat here:

$$X(\langle \mathbf{u}|\mathbf{v}\rangle) = \langle D_X \mathbf{u}|\mathbf{v}\rangle + \langle \mathbf{u}|D_X \mathbf{v}\rangle, \tag{II.1}$$

$$D_X \mathbf{\Lambda} = 0, \tag{II.2}$$

$$D_X \mathbf{I} = 0, \tag{II.3}$$

where $\mathbf{u}, \mathbf{v} \in \Gamma(\mathscr{M}, \mathscr{U}(\mathscr{M}))$, and $\langle \mathbf{u}|\mathbf{v}\rangle$ denotes the nondegenerate Hermitian inner product, antilinear in the twistor $\mathbf{u}$ and linear in the twistor $\mathbf{v}$, with signature $(+\ +\ -\ -)$. The group SU(2,2) is defined as the set of unimodular transformations under which this inner product is invariant.

The connection $\mathbf{D}$ on $\mathscr{U}(\mathscr{M})$ gives rise to a connection $\mathbf{D}$ on the bundles $\mathscr{U}^{\wedge 2}(\mathscr{M}), \mathscr{U}^{\wedge 4}(\mathscr{M})$, and $\mathscr{E}(\mathscr{M})$. Equation (II.3), together with the compatibility of the covariant derivative with the inner product $\odot$ on the fiber $\mathscr{E}_q$, guarantees that D_X will be compatible with the local action of the Poincaré group.

On $\mathscr{E}(\mathscr{M})$, the connection $\mathbf{D}$ is projected to give the connection $\mathbf{D}^{\mathscr{F}}$ on $\mathscr{F}(\mathscr{M})$:

$$D_X^{\mathscr{F}} \mathbf{V} = \mathbf{I}_{\mathscr{F}} \odot (D_X \mathbf{V}) \tag{II.4}$$

for $\mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{F}(\mathscr{M}))$, where the $\mathscr{F}_q \otimes \mathscr{F}_q$ valued field

$$\mathbf{I}_{\mathscr{F}} = \tfrac{1}{2}\mathbf{\Lambda} - \tfrac{1}{2}\mathbf{I} \otimes \mathbf{O} - \tfrac{1}{2}\mathbf{O} \otimes \mathbf{I} \tag{II.5}$$

is the unit tensor field for $\mathscr{F}(\mathscr{M})$.

We will also need to consider the connection $\mathbf{D}^{\mathscr{E}}$, which is a connection on $\mathscr{E}(\mathscr{M})$ defined as

$$D^{\mathscr{E}}_X \mathbf{V} = D_X \mathbf{V} + \tfrac{1}{2}\, \mathbf{I}(D_X \mathbf{O}) \odot \mathbf{V}. \tag{II.6}$$

for $\mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$, and which satisfies

$$D^{\mathscr{E}}_X \mathbf{O} = D_X \mathbf{O}, \tag{II.7}$$

$$D^{\mathscr{E}}_X \mathbf{I} = D_X \mathbf{I} = 0, \tag{II.8}$$

$$D^{\mathscr{E}}_X \mathbf{V} = D^{\mathscr{F}}_X \mathbf{V}, \tag{II.9}$$

for $\mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{F}(\mathscr{M}))$. Note that $\mathbf{D}^{\mathscr{E}}$ also acts as a connection on the complexification $\mathscr{C}\mathscr{E}(\mathscr{M}) = \mathscr{U}^{\wedge 2}(\mathscr{M})$ of $\mathscr{E}(\mathscr{M})$. Furthermore, for the purpose of being able to map cross sections $\mathbf{x} \in \Gamma(\mathscr{M}, \mathscr{T}(\mathscr{M})) \rightarrow \mathbf{x} \circ \mathbf{J} \in \Gamma(\mathscr{M}, \mathscr{F}(\mathscr{M}))$ and construct suitable Lagrangians, the $\mathscr{T}'_q \otimes \mathscr{F}_q$ valued field $\mathbf{J} = \mathbf{D} \otimes \mathbf{O}$ was introduced in Ref. 12 ($\mathscr{T}_q$ and $\mathscr{T}'_q$ denote the tangent and cotagent fibers above q, respectively).

The field $\mathbf{J}$ was also used in constructing maps of other objects defined on $\mathscr{F}(\mathscr{M})$, such as inner products, connections, and curvature tensors to give corresponding objects defined on $\mathscr{T}(\mathscr{M})$. With the additional assumption that $\mathbf{J}(q)$ for each q is nonsingular, this map is a bijection of $\mathscr{T}_q$ on $\mathscr{F}_q$ and a unique $\mathscr{F}_q \otimes \mathscr{T}_q$ valued field $\mathbf{F}$ acting as the inverse operation to the map $\mathbf{x} \rightarrow \mathbf{x} \circ \mathbf{J}$ was defined.

One of the central problems of Poincaré gauge field theory is the identification of the translation potentials with the soldering form of the tangent bundle. We deal with this problem in our theory by first interpreting the origin twist tensor field as the points at which the fibers are tied to the base manifold. Noting next that the effect of a translation acting on $\mathbf{O}$ can be envisaged geometrically as a rotation of the origin twist tensor around the null cone with its arrow being moved to a different point of the surface $\mathscr{W} = \{\mathbf{P}, \mathbf{P} \in \mathscr{E}, \mathbf{P} \odot \mathbf{P} = 0;\ \mathbf{I} \odot \mathbf{P} = 2\}$ (see Fig. 1 in Ref. 12), and since $\mathscr{W}$ is in one-to-one correspondence with the elements of Minkowski space-time, we see that our translation acting on the fibers corresponds to a change of the origin in Minkowski space. Furthermore, a change in the origin twist tensor implies a change of the vector space $\mathscr{F}$ tangent to $\mathscr{W}$.

It is important to stress that up to this stage no *a priori* assumptions have been made regarding a metric structure and connection on the tangent bundle. A natural isomorphism can be achieved, however, by means of which structures originating in the fibers can be mapped onto the tangent bundle, by selecting a given origin field and introducing its covariant gradient $\mathbf{J} = \mathbf{D} \otimes \mathbf{O}$ as a means of carrying out in a unique manner the mappings referred to above.

Note, finally, that the selection of an origin twist tensor field imposes no special restriction on the theory. Any twist tensor field which serves as an origin field can be transformed by a local Poincaré translation into any of the other possible choices of an origin field.

The same Poincaré translation will also transform the connection $\mathbf{D}$ on the bundle and the subspace $\mathscr{F}$ in such a way that the theory with the new origin twist tensor field and new connection is equivalent to the original theory. Indeed, the tensor field $\mathbf{J}$ will also be transformed in such a way that the new maps with the new theory will induce exactly the same metric structure and connection on the tangent bundle as were obtained with the original theory with the original $\mathbf{J}$.

To complete this summary of background material, we recall that in Ref. 12 we defined the curvature tensors $\mathbf{R}_{\mathscr{E}}$ for the $\mathscr{E}(\mathscr{M})$ connection $\mathbf{D}$, $\mathbf{R}^{\mathscr{E}}_{\mathscr{E}}$ for the $\mathscr{E}(\mathscr{M})$ connection $\mathbf{D}^{\mathscr{E}}$, and $\mathbf{R}_{\mathscr{F}}$ for the $\mathscr{F}(\mathscr{M})$ connection $D^{\mathscr{F}}$. The tensors $\mathbf{R}_{\mathscr{E}}$ and $\mathbf{R}^{\mathscr{E}}_{\mathscr{E}}$ are $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{E}_q \otimes \mathscr{E}_q$ valued, and $\mathbf{R}_{\mathscr{F}}$ is $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{F}_q \otimes \mathscr{F}_q$ valued. They are related by the equations

$$\mathbf{R}_{\mathscr{E}} = \mathbf{R}_{\mathscr{F}} + \tfrac{1}{2}\,[1 - (\mathbf{34})]\,\mathbf{T}_{\mathscr{F}} \otimes \mathbf{I}, \tag{II.10}$$

$$\mathbf{R}^{\mathscr{E}}_{\mathscr{E}} = \mathbf{R}_{\mathscr{F}} + \tfrac{1}{2}\,\mathbf{T}_{\mathscr{F}} \otimes \mathbf{I}, \tag{II.11}$$

where

$$\mathbf{T}_{\mathscr{F}} = \mathbf{R}_{\mathscr{E}} \odot \mathbf{O} = \mathbf{R}^{\mathscr{E}}_{\mathscr{E}} \odot \mathbf{O} \tag{II.12}$$

is $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{F}_q$ valued, and the symbol $(\mathbf{34})$ denotes the transposition which exchanges the 3rd and 4th twist-tensor files.

These curvature tensors all have the same invariant

$$\begin{aligned} (R_{\mathscr{F}})_s &\equiv C(\mathbf{13};\ \mathbf{24};\ \odot)\,[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{R}_{\mathscr{F}}] \\ &= (R^{\mathscr{E}}_{\mathscr{E}})_s \equiv C(\mathbf{13};\ \mathbf{24};\ \odot)\,[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{R}^{\mathscr{E}}_{\mathscr{E}}] \\ &= (R_{\mathscr{E}})_s \equiv C(\mathbf{13};\ \mathbf{24};\ \odot)\,[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{R}_{\mathscr{E}}]. \end{aligned} \tag{II.13}$$

[The symbol $C(\mathbf{13};\ \mathbf{24};\ \odot)$ denotes contraction of the 1st with the 3rd and 2nd with the 4th twist-tensor files via the $\odot$ operation, and the notation $(\mathbf{F}\circ)_k$ is used to define a linear map acting to the right on the k th file of a tensor.]

This invariant is used in the construction of a Lagrangian from which field equations follow by an action principle.

III. GAUGE THEORY FOR THE GROUP $G = \mathrm{U}(1) \times \mathscr{P}$. FORMULATION VIA TENSOR PRODUCT BUNDLE

We now enlarge the gauge group from $\mathscr{P}$ to $G = \mathrm{U}(1) \times \mathscr{P}$. The group U(1) is the set of unitary transformations (i.e., metric-preserving) on the one-dimensional complex space $\mathscr{U}_1$ with a positive-definite inner product $\langle \theta | \phi \rangle$, antilinear in $\theta \in \mathscr{U}_1$, and linear in $\phi \in \mathscr{U}_1$. As a representation space for G we use $\mathscr{V} = \mathscr{U}_1 \otimes \mathscr{U}$.

For the development of the theory we need the bundles $\mathscr{U}_1(\mathscr{M})$, $\mathscr{V}(\mathscr{M})$, and $\mathscr{V}^{\wedge 2}(\mathscr{M})$ with the typical fibers $\mathscr{U}_1$, $\mathscr{V}$, and $\mathscr{V}^{\wedge 2} \equiv \mathscr{U}_1 \otimes \mathscr{U}_1 \otimes \mathscr{E} = \mathscr{U}_1 \otimes \mathscr{U}_1 \otimes \mathscr{U}^{\wedge 2}$, respectively.

As a basis choose any $e \in \Gamma(\mathscr{M}, \mathscr{U}_1(\mathscr{M}))$ such that $\langle e | e \rangle = 1$. Then an arbitrary element $\psi \in \Gamma(\mathscr{M}, \mathscr{V}(\mathscr{M}))$ can be expressed in the form

$$\psi = e \otimes \mathbf{u} \tag{III.1}$$

with $\mathbf{u} \in \Gamma(\mathscr{M}, \mathscr{U}(\mathscr{M}))$. Also, an arbitrary element $\mathbf{W} \in \Gamma(\mathscr{M}, \mathscr{V}^{\wedge 2}(\mathscr{M}))$ can be expressed in the form

$$\mathbf{W} = g \otimes \mathbf{V}, \tag{III.2}$$

where

$$g = e \otimes e \tag{III.3}$$

and

$$\mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 2}(\mathscr{M})).$$

Connections: A connection $\mathbf{D}'$ given on $\mathscr{U}_1(\mathscr{M})$ satisfies the usual axioms given in Eqs. (3.2) of Ref. 12, and also the requirement

$$X(\langle \theta | \phi \rangle) = \langle D'_X \theta | \phi \rangle + \langle \theta | D'_X \phi \rangle \tag{III.4}$$

for $\theta, \phi \in \Gamma(\mathscr{M}, \mathscr{U}_1(\mathscr{M}))$.

Moreover, since $\mathscr{U}_1$ is one-dimensional, we can define $\mathbf{A} \equiv \mathbf{A}^{(e)} \in \Gamma(\mathscr{M}, \mathscr{T}'(\mathscr{M}))$ by the equation

$$D'_X e = i(\mathbf{x} \circ \mathbf{A}) e. \tag{III.5}$$

The connections $\mathbf{D}'$ on $\mathscr{U}_1(\mathscr{M})$ and $\mathbf{D}$ on $\mathscr{U}(\mathscr{M})$ give rise to a connection $\mathbf{D}^G$ on $\mathscr{V}(\mathscr{M})$ satisfying

$$D^G_X(\theta \otimes \mathbf{u}) = (D'_X \theta) \otimes \mathbf{u} + \theta \otimes (D_X \mathbf{u}), \tag{III.6}$$

for $\theta \in \Gamma(\mathscr{M}, \mathscr{U}_1(\mathscr{M}))$ and $\mathbf{u} \in \Gamma(\mathscr{M}, \mathscr{U}(\mathscr{M}))$. For $\boldsymbol{\psi} = e \otimes \mathbf{u}$ as given above, we have

$$D^G_X \boldsymbol{\psi} = (i\mathbf{x} \circ \mathbf{A} e) \otimes \mathbf{u} + e \otimes (D_X \mathbf{u}) = e \otimes (D_X \mathbf{u} + i\mathbf{x} \circ \mathbf{A} \mathbf{u}). \tag{III.7}$$

This connection $\mathbf{D}^G$ on $\mathscr{V}(\mathscr{M})$ can be used to define a connection $\mathbf{D}^G$ on $\mathscr{V}^{\wedge 2}(\mathscr{M})$. For the above given $\mathbf{W} = g \otimes \mathbf{V}$, we have

$$\begin{aligned} D^G_X \mathbf{W} &= D^G_X(g \otimes \mathbf{V}) = (D'_X g) \otimes \mathbf{V} + g \otimes (D_X \mathbf{V}) \\ &= (2i\mathbf{x} \circ \mathbf{A} g) \otimes \mathbf{V} + g \otimes (D_X \mathbf{V}) \\ &= g \otimes (D_X \mathbf{V} + 2i\mathbf{x} \circ \mathbf{A} \mathbf{V}). \end{aligned} \tag{III.8}$$

Also, since the connection $\mathbf{D}'$ on $\mathscr{U}_1(\mathscr{M})$ gives rise to a connection $\mathbf{D}'$ on $\mathscr{U}_1^{\otimes 2}(\mathscr{M})$ which satisfies

$$D'_X(\theta \otimes \phi) = (D'_X \theta) \otimes \phi + \theta \otimes (D'_X \phi) \tag{III.9}$$

for $\theta, \phi \in \Gamma(\mathscr{M}, \mathscr{U}_1(\mathscr{M}))$, we can use $\mathbf{D}'$ on $\mathscr{U}_1^{\otimes 2}(\mathscr{M})$ and $\mathbf{D}^{\mathscr{E}}$ on $\mathscr{E}(\mathscr{M})$ or $\mathscr{U}^{\wedge 2}(\mathscr{M})$ to define a connection $\mathbf{D}^{G\mathscr{E}}$ on $\mathscr{V}^{\wedge 2}(\mathscr{M})$ such that it satisfies

$$D^{G\mathscr{E}}_X(\Phi \otimes \mathbf{V}) = (D'_X \Phi) \otimes \mathbf{V} + \Phi \otimes (D^{\mathscr{E}}_X \mathbf{V}) \tag{III.10}$$

for $\Phi \in \Gamma(\mathscr{M}, \mathscr{U}_1^{\otimes 2}(\mathscr{M}))$ and $\mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 2}(\mathscr{M}))$.

Curvatures: For the $\mathbf{D}^{G\mathscr{E}}$ connection on $\mathscr{V}^{\wedge 2}(\mathscr{M})$, define the curvature tensor $\mathbf{R}^{G\mathscr{E}}$ as a

$$\mathscr{C}(\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{E}_q \otimes \mathscr{E}_q) = \mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{U}_q^{\wedge 2} \otimes \mathscr{U}_q^{\wedge 2}$$

valued field by

$$g \otimes (\mathbf{x}\mathbf{y}{}^{\circ}_{\circ} \mathbf{R}^{G\mathscr{E}} \odot \mathbf{V}) = (D^{G\mathscr{E}}_X D^{G\mathscr{E}}_Y - D^{G\mathscr{E}}_Y D^{G\mathscr{E}}_X - D^{G\mathscr{E}}_{[X,Y]})(g \otimes \mathbf{V}), \tag{III.11}$$

where $\mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 2}(\mathscr{M}))$. A simple calculation gives

$$\begin{aligned} g \otimes (\mathbf{x}\mathbf{y}{}^{\circ}_{\circ} \mathbf{R}^{G\mathscr{E}} \odot \mathbf{V}) &= [(D'_X D'_Y - D'_Y D'_X - D'_{[X,Y]}) g] \otimes \mathbf{V} \\ &\quad + g \otimes [(D^{\mathscr{E}}_X D^{\mathscr{E}}_Y - D^{\mathscr{E}}_Y D^{\mathscr{E}}_X - D^{\mathscr{E}}_{[X,Y]}) \mathbf{V}]. \end{aligned} \tag{III.12}$$

But

$$[(D^{\mathscr{E}}_X D^{\mathscr{E}}_Y - D^{\mathscr{E}}_Y D^{\mathscr{E}}_X - D^{\mathscr{E}}_{[X,Y]}) \mathbf{V} = \mathbf{x}\mathbf{y}{}^{\circ}_{\circ} \mathbf{R}^{\mathscr{E}}_{\mathscr{E}} \odot \mathbf{V}]. \tag{III.13}$$

Also

$$\begin{aligned} (D'_X D'_Y - D'_Y D'_X - D'_{[X,Y]}) g &= [(D'_X D'_Y - D'_Y D'_X - D'_{[X,Y]}) e] \otimes e \\ &\quad + e \otimes [(D'_X D'_Y - D'_Y D'_X - D'_{[X,Y]}) e] \\ &= 2i\{X(\mathbf{y} \circ \mathbf{A}) - Y(\mathbf{x} \circ \mathbf{A}) - [\mathbf{x}, \mathbf{y}] \circ \mathbf{A}\} e \otimes e. \end{aligned} \tag{III.14}$$

If we now make use of the following equations which are derived from the theory[18] of exterior differential forms (the symbol $\mathscr{L}$ denotes the Lie derivative),

$$\begin{aligned} &\mathscr{L}_X f = Xf = \mathbf{x} \circ (\mathbf{d} f), \\ &\mathbf{d}(\mathbf{y} \circ \mathbf{A}) + \mathbf{y} \circ (\mathbf{d} \wedge \mathbf{A}) = \mathscr{L}_Y \mathbf{A}, \end{aligned} \tag{III.15}$$

we get

$$X(\mathbf{y} \circ \mathbf{A}) - Y(\mathbf{x} \circ \mathbf{A}) - [\mathbf{x}, \mathbf{y}] \circ \mathbf{A} = \mathbf{x}\mathbf{y}{}^{\circ}_{\circ} (\mathbf{d} \wedge \mathbf{A}), \tag{III.16}$$

where $\mathbf{d} \wedge \mathbf{A}$ is the exterior derivative of $\mathbf{A}$.

From (III.14) and (III.16) we then have

$$(D'_X D'_Y - D'_Y D'_X - D'_{[X,Y]}) g = 2i\mathbf{x}\mathbf{y}{}^{\circ}_{\circ} (\mathbf{d} \wedge \mathbf{A}) g. \tag{III.17}$$

Substituting (III.13) and (III.17) into (III.12) yields

$$g \otimes (\mathbf{x}\mathbf{y}{}^{\circ}_{\circ} \mathbf{R}^{G\mathscr{E}} \odot \mathbf{V}) = g \otimes \{\mathbf{x}\mathbf{y}{}^{\circ}_{\circ} [\mathbf{R}^{\mathscr{E}}_{\mathscr{E}} + 2i(\mathbf{d} \wedge \mathbf{A}) \otimes \mathbf{I}_{\mathscr{E}}] \odot \mathbf{V}\}, \tag{III.18}$$

where $\mathbf{I}_{\mathscr{E}} = \frac{1}{2} \mathbf{\Lambda}$ is the $\mathscr{E}_q \otimes \mathscr{E}_q$ valued unit tensor. Hence we can write

$$\mathbf{R}^{G\mathscr{E}} = \mathbf{R}^{\mathscr{E}}_{\mathscr{E}} + 2i(\mathbf{d} \wedge \mathbf{A}) \otimes \mathbf{I}_{\mathscr{E}}. \tag{III.19}$$

We next show that $\mathbf{R}^{G\mathscr{E}}$ is independent of the choice of $g = e \otimes e$. To this end, consider two basis fields $e = e_1$ and $e = e_2$, where $e_2 = e^{i\phi} e_1$ and ϕ is a real scalar field. Corresponding to $e = e_1$ and $e = e_2$, respectively, we have $g = g_1 = e_1 \otimes e_1$ and $g = g_2 = e_2 \otimes e_2$ with $g_2 = e^{2i\phi} g_1$. Also we have the corresponding $\mathbf{A} = \mathbf{A}^{(1)}$ and $\mathbf{A} = \mathbf{A}^{(2)}$, where $\mathbf{A}^{(2)} = \mathbf{A}^{(1)} + \mathbf{d}\phi$. Since $\mathbf{d} \wedge \mathbf{d}\phi = 0$, we have $\mathbf{d} \wedge \mathbf{A}^{(1)} = \mathbf{d} \wedge \mathbf{A}^{(2)}$; consequently, $\mathbf{R}^{G\mathscr{E}}$ is independent of the basis e.

The electromagnetic field tensor defined in terms of exterior derivatives is independent of torsion. This can be seen most easily by reexpressing the left-hand member in (III.16) in terms of any connection ∇_X on the tangent bundle. Thus, we have

$$\begin{aligned} &X(\mathbf{y} \circ \mathbf{A}) - Y(\mathbf{x} \circ \mathbf{A}) - [\mathbf{x}, \mathbf{y}] \circ \mathbf{A} \\ &= (\nabla_X \mathbf{y}) \circ \mathbf{A} + \mathbf{y} \circ (\nabla_X \mathbf{A}) - (\nabla_Y \mathbf{x}) \circ \mathbf{A} - \mathbf{x} \circ (\nabla_Y \mathbf{A}) - [\mathbf{x}, \mathbf{y}] \circ \mathbf{A} \\ &= \mathbf{x}\mathbf{y}{}^{\circ}_{\circ} \mathbf{T}_{\mathscr{T}} \circ \mathbf{A} + \mathbf{x}\mathbf{y}{}^{\circ}_{\circ} (\nabla \wedge \mathbf{A}). \end{aligned} \tag{III.20}$$

Combining (III.16) and (III.20), we get

$$\mathbf{d} \wedge \mathbf{A} = \nabla \wedge \mathbf{A} + \mathbf{T}_{\mathscr{T}} \circ \mathbf{A}, \tag{III.21}$$

where $\mathbf{T}_{\mathscr{T}}$ denotes the torsion tensor in the tangent bundle.

But[19]

$$\nabla \wedge \mathbf{A} = \overset{s}{\nabla} \wedge \mathbf{A} - \mathbf{T}_{\mathscr{T}} \circ \mathbf{A}, \tag{III.22}$$

where $\overset{s}{\nabla}$ is the standard connection (i.e., the unique connection which is symmetric and compatible with the inner product in the tangent bundle). Therefore,

$$\mathbf{d} \wedge \mathbf{A} = \overset{s}{\nabla} \wedge \mathbf{A}, \tag{III.23}$$

i.e., since the connection coefficients from $\overset{s}{\nabla}$ are given by the Christoffel symbols, the electromagnetic field, as appearing in the curvature tensor in (III.18), does not couple to torsion—the torsion term appearing in (III.20) is cancelled exactly by a term with the opposite sign which occurs in (III.21).

Torsion: If $\mathbf{O}$ is used as the origin tensor field, the torsion tensor $\mathbf{T}^G_{\mathscr{E}}$ for $\mathbf{D}^G$ is

$$\mathbf{T}^G_{\mathscr{E}} = \mathbf{R}^{G\mathscr{E}} \odot \mathbf{O} = \mathbf{T}_{\mathscr{E}} + 2i(\mathbf{d} \wedge \mathbf{A}) \otimes \mathbf{O}, \tag{III.24}$$

where $\mathbf{T}_{\mathscr{E}}$ is the torsion tensor for the connection $\mathbf{D}$.

Lagrangian densities: By appropriate contractions on $\mathbf{R}^{G\mathscr{E}}$ as given by (III.18), we can set up free Lagrangian densities for the gauge fields $(D^{\,G}_{X})$ as scalar functionals of $\mathbf{R}^{G\mathscr{E}}$.

In what follows, we derive a unified Lagrangian density for gravitation and electromagnetism in the Einstein–Cartan formalism, although the theory is more general as it allows for other types of Lagrangians to be constructed such as those containing quadratic combinations of the Riemann and the electromagnetic field tensors.

The gravitational Lagrangian density $\mathscr{L}_g$, expressed in terms of quantities on the $\mathscr{F}(\mathscr{M})$ bundle, can be rewritten in terms of $\mathbf{R}^{G\mathscr{E}}$ as

$$\begin{aligned} \mathscr{L}_g &\equiv C(\mathbf{13};\, \mathbf{24};\, \odot)[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{R}_{\mathscr{F}}] \\ &= C(\mathbf{13};\, \mathbf{24};\, \odot)[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{R}^{\mathscr{E}}_{\mathscr{E}}] \\ &= C(\mathbf{13};\, \mathbf{24};\, \odot)[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{R}^{G\mathscr{E}}]. \end{aligned} \tag{III.25}$$

The electromagnetic Lagrangian density $\mathscr{L}_{\text{em}}$, expressed in terms of quantities on the $\mathscr{F}(\mathscr{M})$ bundle, can be rewritten in terms of $\mathbf{R}^{G\mathscr{E}}$ as

$$\begin{aligned} \mathscr{L}_{\text{em}} &\equiv [(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2(\mathbf{d}\wedge\mathbf{A})] \overset{\odot}{\odot} [(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2(\mathbf{d}\wedge\mathbf{A})] \\ &= \tfrac{1}{36}\{[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2(\mathbf{d}\wedge\mathbf{A}) \otimes \mathscr{I}_{\mathscr{E}}] \overset{\odot}{\odot} \mathscr{I}_{\mathscr{E}}\} \overset{\odot}{\odot} \\ &\qquad \{[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2(\mathbf{d}\wedge\mathbf{A}) \otimes \mathscr{I}_{\mathscr{E}}] \overset{\odot}{\odot} \mathscr{I}_{\mathscr{E}}\} \\ &= -\tfrac{1}{144}\{[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{R}^{G\mathscr{E}}] \overset{\odot}{\odot} \mathscr{I}_{\mathscr{E}}\} \overset{\odot}{\odot} \\ &\qquad \{[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{R}^{G\mathscr{E}}] \overset{\odot}{\odot} \mathscr{I}_{\mathscr{E}}\}. \end{aligned} \tag{III.26}$$

So we now have the total Lagrangian density $\mathscr{L} = \mathscr{L}_g + \mathscr{L}_{\text{em}}$ expressed in terms of the curvature tensor $\mathbf{R}^{G\mathscr{E}}$ for the connection $\mathbf{D}^{G\mathscr{E}}$. Thus the gauge theory of the group G allows a Lagrangian that includes both the gravitational and electromagnetic parts, which are equivalent to the gravitational and electromagnetic parts of the usual Lagrangian expressed in terms of quantities on the tangent bundle $\mathscr{T}(\mathscr{M})$.

It should be rather evident from the formalism in this section, how it can be extended to include structure groups involving direct products of nonabelian internal groups with the Poincaré group: One merely replaces the one-dimensional representation space for U(1), with basis vector e, by an n-dimensional representation space for the internal group, with basis vectors $e_{(i)}$ [where $(i) = 1,2,\ldots,n$ denote the internal degrees of freedom], and, instead of Eq. (III.5) for defining the gauge field, we substitute

$$D'_X e_{(i)} = i\mathbf{x}\circ\mathbf{A}^{(j)}{}_{(i)}\, e_{(j)}, \tag{III.27}$$

where $\mathbf{A}^{(i)}{}_{(j)} \in \Gamma(\mathscr{M}, \mathscr{T}'(\mathscr{M}))$, i.e., it is a covariant vector in the cotangent bundle and a tensor with respect to the internal degrees of freedom. In addition, an inner product for the basis vectors has to be introduced which is determined by the nature of the internal group.

IV. FORMULATION OF THE GAUGE THEORY EMPLOYING THE ALREADY BUILT-IN PRESENCE OF U(1) AS A NORMAL SUBGROUP

In our gauge theory[12] for $\mathscr{P}$, we used the representation space $\mathscr{U} \equiv (\mathscr{U}, \mathbf{\Lambda}, \mathbf{I})$ where $\mathbf{\Lambda}$, $\mathbf{I}$ were given as part of the structure preserved by the action of $\mathscr{P}$. In the gauge theory that follows, the pair $\mathbf{\Lambda}$, $\mathbf{I}$ are not part of the invariant structure because they are not preserved by the action of U(1). We shall first present some background material, and then describe the structure for $\mathscr{U}$ that is appropriate for the group G generated by the two normal subgroups U(1) and $\mathscr{P}$. After that we shall set up the bundles and connections, and use these in the construction of the gauge theory for G.

A U(1) transformation U_ϕ on $\mathscr{U}$ for ϕ real is defined as

$$U_\phi \mathbf{u} = e^{i\phi}\mathbf{u} \tag{IV.1}$$

for $\mathbf{u} \in \mathscr{U}$. It also acts on twist tensor spaces $\mathscr{U}^{\otimes r}$ for positive integer values of r as

$$U_\phi \mathbf{N} = e^{ir\phi}\mathbf{N} \tag{IV.2}$$

for $\mathbf{N} \in \mathscr{U}^{\otimes r}$.

On $\mathscr{U}^{\wedge 2}$, the dual ${}^*\mathbf{V}$, the inner product $\mathbf{V}\odot\mathbf{W}$, and the definition of reality all depend on the normalized element $\mathbf{\Lambda} \in \mathscr{U}^{\wedge 4}$. In the gauge theory for G, the element $\mathbf{\Lambda}$ will be variable but remain normalized. We may specify the $\mathbf{\Lambda}$ dependence by writing ${}^{*(\mathbf{\Lambda})}\mathbf{V}$ and $\mathbf{V}\odot_{\mathbf{\Lambda}}\mathbf{W}$, and describe an element $\mathbf{V} \in \mathscr{U}^{\wedge 2}$ as being "$\mathbf{\Lambda}$-real" when it satisfies the definition of reality based on this $\mathbf{\Lambda}$. Furthermore, $\mathscr{E} \equiv \mathscr{E}^{(\mathbf{\Lambda})} \subset \mathscr{U}^{\wedge 2}$ is the space of $\mathbf{\Lambda}$-real elements. The $\odot_{\mathbf{\Lambda}}$ operation gives a real inner product in $\mathscr{E}^{(\mathbf{\Lambda})}$.

The following properties for U(1) transformations are easily verified. For a normalized $\mathbf{\Lambda}_1 \in \mathscr{U}^{\wedge 2}$, let

$$\mathbf{\Lambda}_2 = U_\phi \mathbf{\Lambda}_1. \tag{IV.3}$$

Then for $\mathbf{V}, \mathbf{W} \in \mathscr{U}^{\wedge 2}$ we have

$$(U_\phi\mathbf{V})\odot_{\mathbf{\Lambda}_2}(U_\phi\mathbf{W}) = \mathbf{V}\odot_{\mathbf{\Lambda}_1}\mathbf{W}, \tag{IV.4}$$

$$[{}^{*(\mathbf{\Lambda}_2)}(U_\phi\mathbf{V})]\hat{} = U_\phi[{}^{*(\mathbf{\Lambda}_1)}\mathbf{V}]\hat{}. \tag{IV.5}$$

But if $\mathbf{V}$ is $\mathbf{\Lambda}_1$-real, then it follows from the right-hand term in (IV.5) that $U_\phi[{}^{*(\mathbf{\Lambda}_1)}\mathbf{V}]\hat{} = U_\phi\mathbf{V}$, so $[{}^{*(\mathbf{\Lambda}_2)}(U_\phi\mathbf{V})]\hat{} = U_\phi\mathbf{V}$, i.e.,

$$\mathbf{V} \text{ is } \mathbf{\Lambda}_1\text{-real} \;\leftrightarrow\; U_\phi\mathbf{V} \text{ is } \mathbf{\Lambda}_2\text{-real}. \tag{IV.6}$$

We also have

$$\mathscr{E}^{(\mathbf{\Lambda}_2)} = U_\phi\mathscr{E}^{(\mathbf{\Lambda}_1)} \equiv \{U_\phi\mathbf{V} \,|\, \mathbf{V} \in \mathscr{E}^{(\mathbf{\Lambda}_1)}\}. \tag{IV.7}$$

Now, let a normalized element $\mathbf{\Lambda}_0 \in \mathscr{U}^{\wedge 4}$ and a null element $\mathbf{I}_0 \in \mathscr{E}^{(\mathbf{\Lambda}_0)}$ be assumed as given. Suppose we have also chosen a null element $\mathbf{O}_0 \in \mathscr{E}^{(\mathbf{\Lambda}_0)}$ satisfying $\mathbf{O}_0\odot_{\mathbf{\Lambda}_0}\mathbf{I}_0 = 2$. We can then define the sets

$$\mathscr{I} = \{U_\phi\mathbf{I}_0 \,|\, \phi \in \mathbb{R}\}, \tag{IV.8}$$

$$\mathscr{O} = \{U_\phi\mathbf{O}_0 \,|\, \phi \in \mathbb{R}\}. \tag{IV.9}$$

The set $\mathscr{I}$ (the infinity set) is preserved under U(1), and it will be used as part of the structure of $\mathscr{U}$ which is preserved by the structure group G. To indicate this, we can write $\mathscr{U} \equiv (\mathscr{U}, \mathscr{I})$. The set $\mathscr{O}$ (the origin set) is also preserved under U(1) but not under $\mathscr{P}$.

The infinity element $\mathbf{I} \equiv \mathbf{I}^{(\mathbf{\Lambda})}$ for $\mathscr{E}^{(\mathbf{\Lambda})}$ is now defined as being one of the two elements of $\mathscr{I}$ which is Λ-real. The origin element $\mathbf{O} = \mathbf{O}^{(\Lambda)}$ for $\mathscr{E}^{(\Lambda)}$ is equivalently defined as the unique element of $\mathscr{O}$ which is Λ-real and satisfies $\mathbf{O}^{(\Lambda)} \odot_\Lambda \mathbf{I}^{(\Lambda)} = 2$. Note also that the maps $\Lambda \rightarrow \mathbf{I}^{(\Lambda)}$ and $\Lambda \rightarrow \mathbf{O}^{(\Lambda)}$ are double valued.

Furthermore, for each normalized Λ, the set $\mathscr{F} \equiv \mathscr{F}^{(\Lambda)}$ is the subset of $\mathscr{E} \equiv \mathscr{E}^{(\Lambda)}$ consisting of elements orthogonal to both $\mathbf{I}^{(\Lambda)}$ and $\mathbf{O}^{(\Lambda)}$. Under the transformation U_ϕ, where $\Lambda_2 = U_\phi \Lambda_1$, we have the following properties

$$\mathbf{I}^{(\Lambda_2)} = U_\phi \mathbf{I}^{(\Lambda_1)}, \tag{IV.10}$$

$$\mathbf{O}^{(\Lambda_2)} = U_\phi \mathbf{O}^{(\Lambda_1)}, \tag{IV.11}$$

$$\mathscr{F}^{(\Lambda_2)} = U_\phi \mathscr{F}^{(\Lambda_1)} \equiv \{ U_\phi \mathbf{V} | \mathbf{V} \in \mathscr{F}^{(\Lambda_1)} \}, \tag{IV.12}$$

the last one being a consequence of (IV.7).

For notational simplification, we shall from now on sometimes drop the superscripts or subscripts Λ, which indicate the dependence on Λ, when there will be no danger of confusion.

Using the base manifold $\mathscr{M}$, we can now set up the bundles $\mathscr{U}(\mathscr{M})$, $\mathscr{U}^{\wedge 2}(\mathscr{M})$, $\mathscr{U}^{\wedge 4}(\mathscr{M})$ with $\mathscr{U} \equiv (\mathscr{U}, \mathscr{I})$, $\mathscr{U}^{\wedge 2}$, and $\mathscr{U}^{\wedge 4}$ as typical fibers. For each normalized cross section $\Lambda \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 4}(\mathscr{M}))$, we have the corresponding bundle $\mathscr{E}(\mathscr{M}) \equiv \mathscr{E}^{(\Lambda)}(\mathscr{M})$ with $\mathscr{E} \equiv \mathscr{E}^{(\Lambda)}$ as typical fiber. We also have the bundles (not vector bundles) $\mathscr{I}(\mathscr{M})$ and $\mathscr{O}(\mathscr{M})$ with the sets $\mathscr{I}$ and $\mathscr{O}$ as typical fibers, respectively, and each fiber $\mathscr{I}_q, \mathscr{O}_q$ above $q \in \mathscr{M}$ has a structure isomorphic to $\mathscr{I}$ and $\mathscr{O}$. As was mentioned previously, the bundle $\mathscr{I}(\mathscr{M})$ will be taken as part of the structure of the bundle $\mathscr{U}(\mathscr{M})$.

Note that, for each normalized cross section $\Lambda \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 4}(\mathscr{M}))$, we have the corresponding infinity and origin cross sections $\mathbf{I} \equiv \mathbf{I}^{(\Lambda)} \in \Gamma(\mathscr{M}, \mathscr{E}^{(\Lambda)}(\mathscr{M}))$ and $\mathbf{O} \equiv \mathbf{O}^{(\Lambda)} \in \Gamma(\mathscr{M}, \mathscr{E}^{(\Lambda)}(\mathscr{M}))$ such that $\mathbf{I}^{(\Lambda)}(q) \in \mathscr{I}_q$ for each $q \in \mathscr{M}$, $\mathbf{I}^{(\Lambda)}$ is Λ-real, $\mathbf{O}^{(\Lambda)}(q) \in \mathscr{O}_q$ for each $q \in \mathscr{M}$, $\mathbf{O}^{(\Lambda)}$ is Λ-real, and $\mathbf{O}^{(\Lambda)} \odot \mathbf{I}^{(\Lambda)} = 2$. Finally, we have the bundle $\mathscr{F}(\mathscr{M}) \equiv \mathscr{F}^{(\Lambda)}(\mathscr{M})$ with $\mathscr{F} \equiv \mathscr{F}^{(\Lambda)}$ as typical fiber.

Connections: Let $\mathbf{D}^G$ be a structure preserving connection on $\mathscr{U}(\mathscr{M})$. It satisfies the usual axioms of a connection given by Eqs. (3.2) in Ref. 12, as well as Eq. (II.1). However, Eq. (II.2) is not satisfied, and Eq. (II.3) will be modified in a way to be specified later.

Theorem IV. 1: For a normalized $\Lambda \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 4}(\mathscr{M}))$, there exists $\mathbf{A} = \mathbf{A}^{(\Lambda)} \in \Gamma(\mathscr{M}, \mathscr{T}'(\mathscr{M}))$ such that

$$D_X^G \Lambda = 4i(\mathbf{x} \circ \mathbf{A})\Lambda. \tag{IV.13}$$

Proof: Since $\mathscr{U}^{\wedge 4}$ is a one-dimensional space, $D_X^G \Lambda$ must be proportional to Λ. Furthermore, the coefficient of Λ must be pure imaginary due to the fact the properties assumed for the connection imply that

$$\mathrm{Re}[(D_X^G \Lambda)^{\circ\circ}_{\circ\circ} \hat{\Lambda}] = 0. \tag{IV.14}$$

Note that Eq. (IV.13) is equivalent to saying that, for the connection $\mathbf{D} \equiv \mathbf{D}^{(\Lambda)}$ on $\mathscr{U}(\mathscr{M})$ defined as

$$D_X \mathbf{u} = (D_X^G - i\mathbf{x} \circ \mathbf{A})\mathbf{u} \tag{IV.15}$$

for $\mathbf{u} \in \Gamma(\mathscr{M}, \mathscr{U}(\mathscr{M}))$, its operation on Λ as a twist-tensor connection is

$$D_X \Lambda = (D_X^G - 4i\mathbf{x} \circ \mathbf{A})\Lambda = 0. \tag{IV.16}$$

Theorem IV. 2: $\mathbf{D} \equiv \mathbf{D}^{(\Lambda)}$ is a connection on $\mathscr{E}(\mathscr{M}) \equiv \mathscr{E}^{(\Lambda)}(\mathscr{M})$, i.e., for $\mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$ we have $D_X \mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$.

Proof: The proof of this theorem follows immediately from Eq. (IV.16).

Now we write the axiom that replaces Eq. (II.3);

Axiom: For each normalized $\Lambda \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 4}(\mathscr{M}))$ and corresponding $\mathbf{A} \equiv \mathbf{A}^{(\Lambda)} \in \Gamma(\mathscr{M}, \mathscr{T}'(\mathscr{M}))$ defined by Eq. (IV.13), the field $\mathbf{I} \equiv \mathbf{I}^{(\Lambda)}$ satisfies

$$D_X \mathbf{I} \equiv (D_X^G - 2i\mathbf{x} \circ \mathbf{A})\mathbf{I} = 0. \tag{IV.17}$$

It can be shown that if Eq. (IV.17) is true for one normalized Λ, it is true for all other normalized Λ.

In the case of the gauge theory[12] for $\mathscr{P}$, we were able to establish a unique procedure for inducing a metric structure and connection on the tangent bundle via a tensor $\mathbf{J} = \mathbf{D} \otimes \mathbf{O}$. We will show in what follows that a similar approach is possible for the structure group G. For this purpose we shall require one more theorem and prove some additional properties of the newly defined connection $\mathbf{D} \equiv \mathbf{D}^{(\Lambda)}$.

Theorem IV. 3: Given any normalized $\Lambda \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 4}(\mathscr{M}))$, and corresponding $\mathbf{A} \equiv \mathbf{A}^{(\Lambda)} \in \Gamma(\mathscr{M}, \mathscr{T}'(\mathscr{M}))$ defined by Eq. (IV.13), the field $\mathbf{O} \equiv \mathbf{O}^{(\Lambda)}$ satisfies

$$D_X \mathbf{O} \equiv (D_X^G - 2i\mathbf{x} \circ \mathbf{A})\mathbf{O} = \mathbf{P}_X, \tag{IV.18}$$

where $\mathbf{P}_X \equiv \mathbf{P}_X{}^{(\Lambda)} \in \Gamma(\mathscr{M}, \mathscr{F}^{(\Lambda)}(\mathscr{M}))$ and is linear in X.

Proof: The proof of this theorem follows directly from

$$D_X(\mathbf{I} \odot \mathbf{O}) = D_X 2 = 0, \quad \text{and} \quad D_X(\mathbf{O} \odot \mathbf{O}) = D_X 0 = 0.$$

Thus the field

$$\mathbf{J} \equiv \mathbf{J}^{(\Lambda)} = \mathbf{D}^{(\Lambda)} \otimes \mathbf{O}^{(\Lambda)} \tag{IV.19}$$

is $\mathscr{T}'_q \otimes \mathscr{F}_q{}^{(\Lambda)}$ valued, and can be used to map cross sections $\mathbf{x} \in \Gamma(\mathscr{M}, \mathscr{T}(\mathscr{M})) \rightarrow \mathbf{x} \circ \mathbf{J}^{(\Lambda)} \in \Gamma(\mathscr{M}, \mathscr{F}^{(\Lambda)}(\mathscr{M}))$. If we make the additional assumption that the map $\mathbf{x}(q) \in \mathscr{T}_q \rightarrow \mathbf{P}_X(q) \in \mathscr{F}_q$ is nonsingular for each q, a unique $\mathscr{F}_q^{(\Lambda)} \otimes \mathscr{T}_q$ valued field $\mathbf{F} \equiv \mathbf{F}^{(\Lambda)}$ can be constructed which acts as an inverse map

$$\mathbf{P}_X \in \Gamma(\mathscr{M}, \mathscr{F}^{(\Lambda)}(\mathscr{M})) \rightarrow \mathbf{P}_X \odot \mathbf{F}^{(\Lambda)} \in \Gamma(\mathscr{M}, \mathscr{T}(\mathscr{M})).$$

These fields can also be used in the construction of maps of other objects on $\mathscr{F}(\mathscr{M}) \equiv \mathscr{F}^{(\Lambda)}(\mathscr{M})$, such as inner products, connections, and curvature tensors, to give corresponding objects on $\mathscr{T}(\mathscr{M})$.

Note that if for a normalized $\Lambda_1 \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 4}(\mathscr{M}))$, we let

$$\Lambda_2 = U_\phi \Lambda_1, \tag{IV.20}$$

where $\phi \in \Gamma(\mathscr{M}, \mathbb{R}(\mathscr{M}))$, then it follows from Eq. (IV.13) that

$$\mathbf{A}^{(\Lambda_2)} = \mathbf{A}^{(\Lambda_1)} + \mathbf{d}\phi, \tag{IV.21}$$

where $\mathbf{d}\phi$ is the gradient of ϕ. From this we can derive the result

$$D_X^{(\Lambda_2)} \mathbf{u} = U_\phi D_X^{(\Lambda_1)} U_\phi^{-1} \mathbf{u}, \tag{IV.22}$$

$$D_X^{(\Lambda_2)} \mathbf{N} = U_\phi D_X^{(\Lambda_1)} U_\phi^{-1} \mathbf{N}, \tag{IV.23}$$

for $\mathbf{u} \in \Gamma(\mathscr{M}, \mathscr{U}(\mathscr{M}))$ and $\mathbf{N} \in \Gamma(\mathscr{M}, \mathscr{U}^{\otimes r}(\mathscr{M}))$. Eqs. (IV.20)–(IV.23) show how the vector potential and connections, parametrized by different Λ's, are related by U(1) transformations.

It follows, in particular, from (IV.23) that if the equations

$$D_X^{(\Lambda)}\Lambda = 0, \quad \text{(IV.24)}$$

$$D_X^{(\Lambda)}\mathbf{I}^{(\Lambda)} = 0, \quad \text{(IV.25)}$$

$$D_X^{(\Lambda)}\mathbf{O}^{(\Lambda)} = \mathbf{P}_X{}^{(\Lambda)}, \quad \text{(IV.26)}$$

are satisfied for $\Lambda = \Lambda_1$, then the same equations with $\Lambda = \Lambda_2$ are also satisfied, along with the following results

$$\mathbf{P}_X^{(\Lambda_2)} = U_\phi \mathbf{P}_X^{(\Lambda_1)}, \quad \text{(IV.27)}$$

$$\mathbf{I}^{(\Lambda_2)} = U_\phi \mathbf{I}^{(\Lambda_1)}, \quad \text{(IV.28)}$$

$$\mathbf{O}^{(\Lambda_2)} = U_\phi \mathbf{O}^{(\Lambda_1)}. \quad \text{(IV.29)}$$

Consequently, the structure preserving conditions [Eqs. (IV.24) and (IV.25)] on the fiber bundle are gauge-independent. Note also that, by virtue of (IV.26), the $\mathbf{J}_1(q) \equiv \mathbf{J}^{(\Lambda_1)}(q)$ maps $\mathscr{T}_q$ onto $\mathscr{F}_q^{(\Lambda_1)}$ for each q, and the $\mathbf{J}_2(q) \equiv \mathbf{J}^{(\Lambda_2)}(q)$ maps $\mathscr{T}_q$ onto $\mathscr{F}_q^{(\Lambda_2)}$ for each q, but inner products are preserved under the U_ϕ map from $\mathscr{F}_q^{(\Lambda_1)}$ to $\mathscr{F}_q^{(\Lambda_2)}$ (provided "$\odot_{\Lambda_1}$" and "$\odot_{\Lambda_2}$" are used for the inner product in $\mathscr{F}_q^{(\Lambda_1)}$ and $\mathscr{F}_q^{(\Lambda_2)}$, respectively). Thus, the inner product in $\mathscr{F}_q^{(\Lambda)}$ is also independent of gauge. Furthermore, the $\mathbf{J}_1(q)$ will induce an inner product in $\mathscr{T}_q$ from $\mathscr{F}_q^{(\Lambda_1)}$ and the $\mathbf{J}_2(q)$ will also induce an inner product in $\mathscr{T}_q$ from $\mathscr{F}_q^{(\Lambda_2)}$. However, these two induced inner products in $\mathscr{T}_q$ can be shown to be identical and completely equivalent theories ensue from the different choices of gauge.

We have now reached a stage at which the basic constituents of our gauge theory for the structure group G have been made formally isomorphic to those used in the previously developed theory for $\mathscr{P}$, and which have been summarized in Sec. II. The procedure for constructing curvatures and Lagrangians can be matched step by step with the one followed there, recalling only that the proper connection to use here is $\mathbf{D} \equiv \mathbf{D}^{(\Lambda)}$. Thus, the connection $\mathbf{D} \equiv \mathbf{D}^{(\Lambda)}$ on $\mathscr{E}(\mathscr{M}) \equiv \mathscr{E}^{(\Lambda)}(\mathscr{M})$ can be projected [cf. Eqs. (II.4) and (II.5)] to give a connection $\mathbf{D}^{\mathscr{F}} \equiv \mathbf{D}^{\mathscr{F}(\Lambda)}$ on $\mathscr{F}(\mathscr{M}) \equiv \mathscr{F}^{(\Lambda)}(\mathscr{M})$. This connection is also a connection on $\mathscr{C}\mathscr{F}(\mathscr{M}) \equiv \mathscr{C}\mathscr{F}^{(\Lambda)}(\mathscr{M})$. In analogy to what we did in Sec. II [Eqs. (II.6)–(II.9)], a connection $\mathbf{D}^{\mathscr{E}} \equiv \mathbf{D}^{\mathscr{E}(\Lambda)}$ on $\mathscr{E}(\mathscr{M}) \equiv \mathscr{E}^{(\Lambda)}(\mathscr{M})$ and on $\mathscr{C}\mathscr{E}(\mathscr{M}) = \mathscr{U}^{\wedge 2}(\mathscr{M})$ is additionally defined.

The curvature tensors that can be constructed with these different connections are the same as those given in Sec. II with the proviso that here they are Λ-dependent. Moreover, these curvature tensors all have the same scalar invariant which can be shown to be independent of Λ and is given by (II.13).

Specifically, making use of (IV.15) in the definition of the curvature tensor $\mathbf{R}_{\mathscr{E}}^{(\Lambda)}$ given by

$$\mathbf{x}\mathbf{y}{}_\circ^\circ \mathbf{R}_{\mathscr{E}}^{(\Lambda)} \odot \mathbf{V} = (D_X^{(\Lambda)} D_Y^{(\Lambda)} - D_Y^{(\Lambda)} D_X^{(\Lambda)} - D_{[X,Y]}^{(\Lambda)})\mathbf{V} \quad \text{(IV.30)}$$

for $\mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 2}(\mathscr{M}))$, yields

$$\begin{aligned}
\mathbf{x}\mathbf{y}{}_\circ^\circ \mathbf{R}^G \odot \mathbf{V} &= (D_X^G D_Y^G - D_Y^G D_X^G - D_{[X,Y]}^G)\mathbf{V} \\
&= \mathbf{x}\mathbf{y}{}_\circ^\circ \mathbf{R}_{\mathscr{E}}^{(\Lambda)} \odot \mathbf{V} + 2i\{X(\mathbf{y}\circ\mathbf{A}) \\
&\quad - Y(\mathbf{x}\circ\mathbf{A}) - [\mathbf{x},\mathbf{y}]\circ\mathbf{A}\}\mathbf{V} \\
&= \mathbf{x}\mathbf{y}{}_\circ^\circ \mathbf{R}_{\mathscr{E}}^{(\Lambda)} \odot \mathbf{V} + 2i\mathbf{x}\mathbf{y}{}_\circ^\circ (\mathbf{d} \wedge \mathbf{A})\mathbf{V}, \quad \text{(IV.31)}
\end{aligned}$$

after resorting to (III.15). (Note that $\mathbf{R}^G$ has a dependence on Λ, but only because the product $\odot_\Lambda$ was used in its definition). This last result implies that

$$\mathbf{R}^G = \mathbf{R}_{\mathscr{E}}^{(\Lambda)} + 2i(\mathbf{d} \wedge \mathbf{A}) \otimes \mathbf{I}_{\mathscr{E}}. \quad \text{(IV.32)}$$

In a similar way we can obtain the curvature tensor $\mathbf{R}^{G\mathscr{F}} \equiv \mathbf{R}^{G(\Lambda)\mathscr{F}}$ for the $\mathbf{D}^{G\mathscr{F}} \equiv \mathbf{D}^{G(\Lambda)\mathscr{F}}$ connection on $\mathscr{C}\mathscr{F}(\mathscr{M}) \equiv \mathscr{C}\mathscr{F}^{(\Lambda)}(\mathscr{M})$ defined by

$$D_X^{G\mathscr{F}}\mathbf{V} = D_X^{\mathscr{F}}\mathbf{V} + 2i\mathbf{x}\circ\mathbf{A}\mathbf{V}$$

for $\mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{C}\mathscr{F}(\mathscr{M}))$, and also the curvature tensor $\mathbf{R}^{G\mathscr{E}} \equiv \mathbf{R}^{G(\Lambda)\mathscr{E}}$ for the $\mathbf{D}^{G\mathscr{E}} \equiv \mathbf{D}^{G(\Lambda)\mathscr{E}}$ connection on $\mathscr{C}\mathscr{E}^{(\Lambda)}(\mathscr{M}) \equiv \mathscr{U}^{\wedge 2}(\mathscr{M})$ defined by

$$D_X^{G\mathscr{E}}\mathbf{V} = D_X^{\mathscr{E}}\mathbf{V} + 2i\mathbf{x}\circ\mathbf{A}\mathbf{V}$$

for $\mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 2}(\mathscr{M}))$. They are given by

$$\mathbf{R}^{G\mathscr{F}} = \mathbf{R}_{\mathscr{F}} + 2i(\mathbf{d} \wedge \mathbf{A}) \otimes \mathbf{I}_{\mathscr{F}}, \quad \text{(IV.33)}$$

and

$$\mathbf{R}^{G\mathscr{E}} = \mathbf{R}_{\mathscr{E}}^{\mathscr{E}} + 2i(\mathbf{d} \wedge \mathbf{A}) \otimes \mathbf{I}_{\mathscr{E}}. \quad \text{(IV.34)}$$

The curvature tensor given by Eq. (IV.34) is identical to that obtained in our tensor product bundle formalism described in Sec. III [cf. Eqs. (III.18)]. Hence the Lagrangian densities constructed there will apply equally well here and are furthermore independent of the choice of Λ.

As a final remark note that the scalar element of volume $d\rho = d\rho^{(\Lambda)}$ on $\mathscr{M}$ given in Eq. (4.17) of Ref. 12 now appears, by definition, to depend on Λ. In fact, however, the expression is actually independent of Λ. Therefore, since the Lagrangian densities and the scalar element of volume are independent of Λ, the Lagrangian is also independent of Λ and both the formalisms here and in Sec. III lead to the same theory.

ACKNOWLEDGMENT

One of the authors (C.P.L.) wishes to express appreciation for the hospitality of the Mathematics Department at Oregon State University.

[1]For a review of some of the classical work on the Kaluza–Klein theory, as well as other projective theories see, e.g., P. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, New York, 1942); and D. K. Sen, *Fields and/or Particles* (Ryerson, Toronto, 1968).

[2]J. Scherk and J. H. Schwartz, Phys. Lett. B **57**, 463 (1975); E. Cremmer and J. Scherk, Nucl. Phys. B **103**, 393 (1976); **108**, 409 (1976); J. F. Luciani, Nucl. Phys. B **135**, 111 (1978); N. Manton, Nucl. Phys. B **158**, 141 (1979).

[3]E. Witten, Nucl. Phys. B **186**, 412 (1981).

[4]A. Salam and J. Strathdee, "On Kaluza–Klein Theory," Preprint IC/81/211, ICTP, Trieste, 1981.

[5]A. Trautman, Rep. Math. Phys. **1**, 29 (1970).

[6]W. Drecksler and M. E. Mayer, *Fiber Bundle Techniques in Gauge Theories*, Lecture Notes in Physics, Vol. 67 (Springer-Verlag, Berlin, 1977).

[7]See, e.g., T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66** (6), 213 (1980), and references therein.
[8]M. W. Kalinowski, Int. J. Theor. Phys. **20**, 563 (1981).
[9]T. Bradfield and R. Kantowski, J. Math. Phys. **23**, 129 (1982).
[10]R. Percarci and S. Randjbar-Daemi, "Kaluza-Klein theories on bundles with homogeneous fibers," Preprint IC/82/18 ICTP, Trieste, 1982.
[11]See, e.g., C. P. Luehr and M. Rosenbaum, J. Math. Phys. **21**, 1432 (1980) and references therein.
[12]C. P. Luehr and M. Rosenbaum, J. Math. Phys. **23**, 1471 (1982).
[13]Y. M. Cho, J. Math. Phys. **16**, 2029 (1975); L. N. Chang, K. I. Macrae, and F. Mansouri, Phys. Rev. D **13**, 235 (1976); Y. M. Cho and P. G. O. Freund, Phys. Rev. D **16**, 1711 (1975).
[14]F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, Rev. Mod. Phys. **48**, 393 (1976).
[15]S. Hojman, M. Rosenbaum, M. P. Ryan, and L. C. Shepley, Phys. Rev. D **17**, 3141 (1978).
[16]W. -T. Ni, Phys. Rev. D **19**, 2260 (1979).
[17]Note that our formalism may readily be related to the usual approach in gauge theories, which start from a principal bundle on which connections are defined, by regarding our fiber bundles as vector bundles $P \times_\rho \mathscr{U}(\mathscr{M})$, $P \times_\rho \mathscr{U}^{\wedge 2}(\mathscr{M})$, $P \times_\rho \mathscr{U}^{\wedge 4}(\mathscr{M})$, and $P \times_\rho \mathscr{E}(\mathscr{M})$, associated with a principal G bundle (or frame bundle) P, where ρ is a representation of the structural group G on the vector spaces $\mathscr{U}$, $\mathscr{U}^{\wedge 2}$, $\mathscr{U}^{\wedge 4}$, or $\mathscr{E}$. Thus, for example, the associated vector bundle $P \times_\rho \mathscr{U}(\mathscr{M})$ can be defined (see, e.g., Ref. 7) by the equivalence relation on $P \times \mathscr{U}$:

$$(p,\rho(g)\circ u) \simeq (p\cdot g,u) \quad \forall p \in P,\ u \in \mathscr{U},\ g \in G;$$

the transition functions on this bundle are then given by the representation $\rho(\phi)$ applied to the transition functions ϕ on P. Moreover, if ρ is the identity representation on the fiber $\mathscr{U}$, then $P \times_\rho \mathscr{U}(\mathscr{M}) = \mathscr{U}(\mathscr{M})$. We can therefore pass from the vector bundle $\mathscr{U}(\mathscr{M})$ to its associated principal bundle P and vice versa by changing the space on which the transition functions act from the vector space to the group manifold and back. Similar arguments apply to the other vector bundles used in our theory.
[18]Y. Choquet-Bruhat, C. de Witt-Morette, and M. Dillard-Bleide, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1977).
[19]C. P. Luehr and M. Rosenbaum, J. Math. Phys., **15**, 1120 (1974); C. P. Luehr, M. Rosenbaum, M. P. Ryan, and L. C. Shepley, J. Math. Phys., **18**, 965 (1977).

The U(6) ∧ SU(3) hidden symmetry in collective excitations of many-body systems

O. Castaños and A. Frank
Centro de Estudios Nucleares, UNAM, Circuito Exterior, C.U., Delegación Coyoacán, 04510 México, D.F., Mexico

(Received 20 August 1982; accepted for publication 13 January 1983)

We analyze the problem of an A-nucleon system interacting through harmonic oscillator forces in terms of variables which separate collective and noncollective aspects of the Hamiltonian. To study the symmetry group of the collective Hamiltonian in the limit when A is very large, we carry out a group contraction on the dynamical group generators for the system, which permits the identification of this group as U(6) ∧ SU(3). In addition, an explicit realization for the symmetry group generators is given in terms of Bohr- and Mottelson-like variables.

PACS numbers: 24.90. + d, 02.20. + b, 11.30.Jw

1. INTRODUCTION

One of the most important unsolved problems in nuclear structure physics is that of extracting from the microscopic A-nucleon system a collective Hamiltonian and the subsequent determination of its eigenstates. There are many different approaches in the literature[1] all striving towards this goal. In several recent works,[2–7] the collective Hamiltonian is projected out from the A-body Hamiltonian by restricting it to a definite irreducible representation (irrep) of the orthogonal group O(n) associated to the $n = A - 1$ Jacobi vectors. In particular, Vanagas[2] has considered the scalar irrep of O(n), where the collective states are those invariant under transformations of this group. Studying the problem of an A-nucleon system interacting through harmonic oscillator forces, Deenen and Quesne[5] were able to prove that the symmetry group associated to the O(n)-scalar collective Hamiltonian is a U(6) group, which is very suggestive, given the phenomenological success of the interacting boson model. However, one should be aware that the O(n)-scalar collective Hamiltonian is, in general, inconsistent with the Pauli principle,[2] and a more realistic problem is that of projecting the many-nucleon problem onto a definite irrep of O(n) which is consistent with the latter. This problem is more difficult to tackle, and in this paper we consider an A-nucleon system interacting through harmonic-oscillator forces in the case where $A \gg 1$, following closely the results of a previous publication,[8] where the present authors and their collaborators constructed a complete set of basis states for this limiting situation. In this work, we turn our attention to the study of the symmetry group associated to the oscillator collective Hamiltonian for $A \gg 1$, where the latter is defined consistently with the Pauli principle.

The translationally invariant Hamiltonian for a system of A particles interacting through harmonic oscillator forces is given by

$$H = \frac{1}{2} \sum_{s=1}^{n} \sum_{i=1}^{3} (P_{is}^2 + X_{is}^2), \tag{1.1}$$

where X_{is} are the Jacobi coordinates, P_{is} their corresponding momenta, and we chose units in which $\hbar$, the mass of the nucleon, and an appropriate frequency are 1.

The group theoretical structure of Hamiltonian (1.1) has been studied thoroughly and we know that a possible decomposition of the symmetry group U($3n$) of the Hamiltonian (1.1) is the following:

$$\begin{array}{ccccccc} \mathrm{U}(n) & \supset & \mathrm{O}(n) & \supset & S_{n+1} & & \\ \times & & (w_1,w_2,w_3) & & \{f\} & & \\ \mathrm{U}(3) & \supset & \mathrm{SU}(3) & \supset & \mathscr{O}(3) & \supset & \mathscr{O}(2) \\ [h_1, h_2, h_3] & & (\lambda, \mu) & & L & & M \end{array} \tag{1.2}$$

Underneath each group we write down the quantum numbers characterizing its irreps. For U(3) we have the partition $[h_1, h_2, h_3]$, where $h_1 + h_2 + h_3 = N$ is the total number of quanta; $\lambda = h_1 - h_2$, $\mu = h_2 - h_3$ characterize the irrep of SU(3) and L,M the irreps of $\mathscr{O}(3)$ and $\mathscr{O}(2)$, respectively. The U(n) irrep is the same as that of O(3), due to the complementarity of these groups, while that of O(n) can have also at most three rows[2,8] (w_1,w_2,w_3). The symmetric group S_{n+1} is characterized by the partition $\{f\} = \{f_1,f_2,...,f_n,f_{n+1}\}$ of $n+1$, while its row is specified by the Yamanouchi symbol $r = (r_{n+1},r_n,...,r_2,r_1)$. Thus, the eigenstates of H can be denoted by[8]

$$|N(\lambda\mu)\, qLM\Omega\, (w_1w_2w_3)\, \delta\, \{f\}(r)\rangle, \tag{1.3}$$

where Ω corresponds to the set of indices required to distinguish between repeated irreps (w_1,w_2,w_3) of O(n) appearing in a given irrep $[h_1, h_2, h_3]$ of U(n); δ plays the same role for repeated irreps $\{f\}$ of S_{n+1} appearing in a given irrep (w_1,w_2,w_3) of O(n), and q is an extra index to classify repeated irreps of $\mathscr{O}(3)$ in a given one of SU(3).

The collective part of the Hamiltonian (1.1) is obtained by projecting on a definite irrep $(w_1w_2w_3)$ of O(n), which is determined by constructing the set of "compact states," i.e., the lowest possible energy states that satisfy the restrictions imposed by the Pauli principle.[8] Following Sabaliauskas,[9] one can take a linear combination of Slater determinants corresponding to these compactly filled states with Z protons and $A - Z$ neutrons such that it is characterized by the number of quanta N and the irrep (λ_m, μ_m) corresponding to the maximal eigenvalue of the quadratic Casimir operator of the SU(3) group. Furthermore, this state is also character-

 0022-2488/84/020388-08$02.50

ized[10] by the chain of groups (1.2), and by construction it is also characterized by a definite irrep $(w_1 w_2 w_3)$ of $O(n)$, which coincides with the irrep $[h_1, h_2\, h_3]$ of $U(n)$. This fact is a consequence of the compact filling of the oscillator levels. Thus, when we apply the $O(n)$ scalar quadratic functions of the annihilation operators $\xi_{is} = (1/\sqrt{2})(X_{is} + iP_{is})$ of the form

$$\hat{B}_{ij} = \sum_{s=1}^{n} \xi_{is}\, \xi_{js} \tag{1.4}$$

to the ground state, we get zero, as we cannot obtain states with a lower number of quanta that satisfy the Pauli principle. It is then straightforward to find the relations[8]

$$w_1 = (N + 2\lambda_m + \mu_m)/3, \tag{1.5a}$$

$$w_2 = (N + \mu_m - \lambda_m)/3, \tag{1.5b}$$

$$w_3 = (N - 2\mu_m - \lambda_m)/3. \tag{1.5c}$$

By using the tables[9] for (λ_m, μ_m) of Sabaliauskas, it is an easy matter to find the irrep $(w_1 w_2 w_3)$ of $O(n)$ consistent with the Pauli principle.

To analyze the group theoretical structure of this many-body system in the case when the number of particles is very large we proceed as follows. In the next section, we introduce the Zickendraht–Dzublik transformation[11] and then turn our attention to the construction of the generators of the symplectic group in six dimensions, Sp(6) in these coordinates.

In Sec. 3 we make the projection of the generators of Sp(6) onto a definite irrep (w_1, w_2, w_3) of $O(n)$.

We analyze in Sec. 4 the collective Hamiltonian in the limit where A is very large, by means of a contraction of the generators introduced in the previous section and discuss the symmetry group associated with the problem for this limiting situation.

Finally, in the last section we summarize our results and make some concluding remarks.

2. GENERATOR OF THE Sp(6) GROUP AND THE ZICKENDRAHT–DZUBLIK TRANSFORMATION

We now turn our attention to the Sp(6) group, which plays the role of a dynamical group for the collective excitations of an A-nucleon system interacting through harmonic oscillator forces. This is due to the complementarity of the groups Sp(6), $O(n)$, and $Sp(6n)$ in the sense that for the irreducible representations $[\frac{1}{2}^{3n}]\ [\frac{1}{2}^{3n-1}\frac{3}{2}]$ of $Sp(6n)$, which contain all totally symmetric irreps of the symmetry group $U(3n)$, the irreps of $O(n)$ and Sp(6) are in one-to-one correspondence.[12] Thus, the Sp(6) generators will connect different states of the system without changing the $O(n)$ irrep, i.e., without leaving the collective subspace of the many-body space.

The generators of Sp(6) can be written in terms of the Jacobi coordinates and momenta as[8]

$$\hat{C}_{ij} = \tfrac{1}{2}\{\hat{q}_{ij} + \hat{T}_{ij} + i\hat{L}_{ij}\}, \tag{2.1a}$$

$$\hat{B}^{+}_{ij} = \tfrac{1}{2}\{\hat{q}_{ij} - \hat{T}_{ij} - n\delta_{ij} - i\hat{S}_{ij}\}, \tag{2.1b}$$

$$\hat{B}_{ij} = \tfrac{1}{2}\{\hat{q}_{ij} - \hat{T}_{ij} + n\delta_{ij} + i\hat{S}_{ij}\}, \tag{2.1c}$$

where the operators $\hat{q}_{ij}$, $\hat{S}_{ij}$, $\hat{T}_{ij}$, and $\hat{L}_{ij}$ are given by

$$\hat{q}_{ij} = \sum_{s=1}^{n} X_{is} X_{js}, \quad i,j = 1,2,3, \tag{2.2a}$$

$$\hat{S}_{ij} = \sum_{s=1}^{n} (X_{is} P_{js} + X_{js} P_{is}), \tag{2.2b}$$

$$\hat{T}_{ij} = \sum_{s=1}^{n} P_{is} P_{js}, \tag{2.2c}$$

$$\hat{L}_{ij} = \sum_{s=1}^{n} (X_{is} P_{js} - X_{js} P_{is}). \tag{2.2d}$$

The operators (2.2d) are the components of the total angular momentum in the frame of reference fixed in space. We now introduce the Zickendraht–Dzublik (ZD) transformation, by means of which the separation of the space in collective and noncollective parts is more naturally discussed, as well as the large A limit that we analyze in Sec. 4.

The ZD transformation[11] is given by

$$X_{is} = \sum_{k=1}^{3} \rho_k\, D^{1_3}_{ki}(\vartheta_j)\, D^{1_n}_{n-3+k,s}(\phi), \tag{2.3}$$

where $D^{1_3}_{ki}$, $D^{1_n}_{n-3+k,s}$ are the fundamental matrix representations of the $\mathscr{O}(3)$ and $O(n)$ groups, respectively. The ϑ_i, $i = 1,2,3$ are the standard Euler angles and it would seem that there are $n(n-1)/2$ angles ϕ. However, since in (2.3) one only needs the last three rows of the representations, it is possible to define the angles so that only $3A - 9\,\phi$'s appear in them.[2]

The physical meaning of the ρ_k, with $k = 1,2,3$, can be understood through the definition of the inertia tensor[12]

$$I_{ij} = \rho^2 \delta_{ij} - q_{ij}, \quad i,j = 1,2,3 \tag{2.4a}$$

with

$$\rho^2 = \sum_{i=1}^{3} q_{ii}. \tag{2.4b}$$

Using the coordinate transformation (2.3) and the orthogonal properties of the matrices $\| D^{1_n}_{ts}(\phi)\|$ we see that

$$\|I\| = \|\tilde{D}^{1_3}(\vartheta_i)\| \times \begin{pmatrix} \rho^2 - \rho_1^2 & 0 & 0 \\ 0 & \rho^2 - \rho_2^2 & 0 \\ 0 & 0 & \rho^2 - \rho_3^2 \end{pmatrix} \|D^{1_3}(\vartheta_i)\|, \tag{2.5}$$

where $\sim$ indicates the transpose. It is clear from (2.5) that the ϑ_i's define the orientation of an intrinsic or "body fixed" frame in which the principal moments of inertia become $\rho_2^2 + \rho_3^2$, $\rho_1^2 + \rho_3^2$, and $\rho_1^2 + \rho_2^2$.

The expressions for $\hat{q}_{ij}$, $\Sigma_{s=1}^{n} X_{is} P_{js}$, $\hat{T}_{ij}$, and $\hat{L}_{ij}$ in terms of the variables ρ_k's, ϑ_i's, and ϕ's were given in Appendix A of Ref. 8. We note, in particular, in the expression for $\hat{T}_{ij}$, Eq. (A.21), the appearance of the operator

$$\sum_{t=1}^{n-3} \sum_{k,k'=1}^{3} \frac{D^{1_3}_{ki} D^{1_3}_{k'j}}{\rho_k \rho_{k'}} \mathscr{L}'_{kt} \mathscr{L}'_{k't}, \tag{2.6}$$

where $\mathscr{L}'_{s,t}$ are the generators of the orthogonal group $O(n)$ in the frame of reference "fixed in the body".[8] These operators have been studied by Rowe[4] and by Buck and Biedenharn.[6] In particular, the operators $\hat{\mathscr{L}}'_{n-1,n}$, $\hat{\mathscr{L}}'_{n,n-2}$,

$\hat{\mathscr{L}}'_{n-2,n-1}$ have been identified by these authors as the components of a vortex spin for the system.

Through the commutation relations between the $\hat{\mathscr{L}}'_{st}$,[8] the last expression can be written in the more convenient form

$$\sum_{k=1}^{3} D^{1_3}_{ki} D^{1_3}_{kj} \frac{1}{\rho_k^2}\left(\sum_{t=1}^{n-3} \hat{\mathscr{L}}'^2_{\kappa t}\right) + \frac{i(n-3)}{2} \sum_{k \neq k'=1}^{3} \frac{D^{1_3}_{ki} D^{1_3}_{k'j}}{\rho_k \rho_{k'}} \hat{\mathscr{L}}'_{\kappa'\kappa} + \sum_{k \neq k'=1}^{3} \frac{D^{1_3}_{ki} D^{1_3}_{k'j}}{2\rho_k \rho_{k'}} \times\left[\sum_{t=1}^{n-3} (\hat{\mathscr{L}}'_{\kappa t} \hat{\mathscr{L}}'_{\kappa' t} + \hat{\mathscr{L}}'_{\kappa' t} \hat{\mathscr{L}}'_{\kappa t})\right], \tag{2.7}$$

with $\kappa = n-3+k$ and $\kappa' = n-3+k'$.

The factor in the parentheses in the first term can be related to the Casimir operators in the chain of groups $\mathrm{O}(n) \supset \mathrm{O}(n-1) \supset \cdots \supset \mathrm{O}(3) \supset \mathrm{O}(2)$ by means of the relations

$$\sum_{t=1}^{n-3} \hat{\mathscr{L}}'^2_{n-2,t} = G'(n-2) - G'(n-3), \tag{2.8a}$$

$$\sum_{t=1}^{n-3} \hat{\mathscr{L}}'^2_{n-1,t} = G'(n-1) - G'(n-2) - \hat{\mathscr{L}}'^2_3, \tag{2.8b}$$

$$\sum_{t=1}^{n-3} \hat{\mathscr{L}}'^2_{n,t} = G'(n) - G'(n-1) - \hat{\mathscr{L}}'^2_1 - \hat{\mathscr{L}}'^2_2, \tag{2.8c}$$

where we have defined

$$\hat{\mathscr{L}}'_k \equiv \tfrac{1}{2} \sum_{k',k''=1}^{3} \epsilon_{kk'k''} \hat{\mathscr{L}}'_{n-3+k',n-3+k''}, \tag{2.9}$$

with $\epsilon_{kk'k''}$ being the antisymmetric tensor and where $G'(r) \equiv \frac{1}{2} \sum_{s,t=1}^{r} \hat{\mathscr{L}}'^2_{s,t}$ is the Casimir operator[13] of an orthogonal group $\mathrm{O}(r)$. If the group $\mathrm{O}(r)$ is characterized by the partition $[\lambda_1, \lambda_2, \ldots, \lambda_{[r/2]}]$ its eigenvalue is given by $\sum_{s=1}^{[r/2]} \lambda_s(\lambda_s + n - 2s)$.[13]

In similar form, the factor in the brackets in the last term of (2.7) can be written as

$$\sum_{t=1}^{n-3} (\hat{\mathscr{L}}'_{n-1,t} \hat{\mathscr{L}}'_{n-2,t} + \hat{\mathscr{L}}'_{n-2,t} \hat{\mathscr{L}}'_{n-1,t}) = \frac{1}{i} [G'(n-2), \hat{\mathscr{L}}'_3], \tag{2.10a}$$

$$\sum_{t=1}^{n-3} (\hat{\mathscr{L}}'_{n,t} \hat{\mathscr{L}}'_{n-1,t} + \hat{\mathscr{L}}'_{n-1,t} \hat{\mathscr{L}}'_{n,t}) = \frac{1}{i} [G'(n-1), \hat{L}_i] + \hat{\mathscr{L}}'_2 \hat{\mathscr{L}}'_3 + \hat{\mathscr{L}}'_3 \hat{\mathscr{L}}'_2, \tag{2.10b}$$

$$\sum_{t=1}^{n-3} (\hat{\mathscr{L}}'_{n,t} \hat{\mathscr{L}}'_{n-2,t} + \hat{\mathscr{L}}'_{n-2,t} \hat{\mathscr{L}}'_{n,t}) = -\frac{1}{i} [G'(n-2), \hat{\mathscr{L}}'_2], \tag{2.10c}$$

where $[A, B]$ denotes the commutator between A and B. Formulas (2.6)–(2.10) are very useful when projecting the Sp(6) generators onto a definite irrep of $\mathrm{O}(n)$.

To simplify the contraction procedure that we carry out in Sec. 4, we modify slightly the operators (2.2) through the transformation

$$\mathscr{O}^{-1} \hat{q}_{ij} \mathscr{O}, \quad \mathscr{O}^{-1} \hat{L}_{ij} \mathscr{O}, \quad \mathscr{O}^{-1} \hat{S}_{ij} \mathscr{O}, \quad \mathscr{O}^{-1} \hat{T}_{ij} \mathscr{O}, \tag{2.11a}$$

where

$$\mathscr{O} = (\rho_1 \rho_2 \rho_3)^{-(n-3)/2}, \tag{2.11b}$$

which corresponds to a change of volume element.

Considering (2.11) together with Eqs. (2.8) and (2.10), we have

$$\hat{q}_{ij} = \sum_{k=1}^{3} \rho_k^2 D^{1_3}_{ki} D^{1_3}_{kj}, \tag{2.12a}$$

$$\hat{L}_{ij} = \sum_{k \neq l=1}^{3} D^{1_3}_{ki} D^{1_3}_{lj} \hat{\mathscr{L}}'_{kl}, \tag{2.12b}$$

$$\begin{aligned} \mathscr{O}^{-1} \hat{S}_{ij} \mathscr{O} = {} & i(n-3)\delta_{ij} - 2i \sum_{k=1}^{3} D^{1_3}_{ki} D^{1_3}_{kj} \rho_k \frac{\partial}{\partial \rho_k} + \sum_{k<k'=2}^{3} \frac{\rho_k^2 + \rho_{k'}^2}{\rho_k^2 - \rho_{k'}^2} [D^{1_3}_{ki} D^{1_3}_{k'j} + D^{1_3}_{kj} D^{1_3}_{k'i}] \hat{L}'_{kk'} \\ & + \sum_{k<k'=2}^{3} \frac{\rho_k \rho_{k'}}{\rho_k^2 - \rho_{k'}^2} \times [D^{1_3}_{ki} D^{1_3}_{k'j} + D^{1_3}_{kj} D^{1_3}_{k'i}] \hat{\mathscr{L}}'_{\kappa\kappa'} \end{aligned} \tag{2.12c}$$

with $\kappa = n-3+k$, $\kappa' = n-3+k'$, and

$$\begin{aligned} \mathscr{O}^{-1} \hat{T}_{ij} \mathscr{O} = {} & \sum_{k=1}^{3} D^{1_3}_{ki} D^{1_3}_{kj} \times\left[-\frac{\partial^2}{\partial \rho_k^2} + \frac{(A-4)(A-6)}{4\rho_k^2}\right] \\ & - \sum_{k \neq k'} \frac{D^{1_3}_{ki} D^{1_3}_{kj}}{\rho_k^2 - \rho_{k'}^2} \left(\rho_k \frac{\partial}{\partial \rho_k} - \rho_{k'} \frac{\partial}{\partial \rho_{k'}}\right) \\ & - i \sum_{k \neq k'} \frac{D^{1_3}_{ki} D^{1_3}_{k'j}}{\rho_k^2 - \rho_{k'}^2} \left(\rho_k \frac{\partial}{\partial_k} + \rho_{k'} \frac{\partial}{\partial_{k'}}\right) \hat{L}'_{kk'} \\ & - i \sum_{k \neq k'} \frac{D^{1_3}_{ki} D^{1_3}_{k'j}}{\rho_k^2 - \rho_{k'}^2} \left(\rho_{k'} \frac{\partial}{\partial \rho_k} + \rho_k \frac{\partial}{\partial \rho_{k'}}\right) \hat{\mathscr{L}}'_{\kappa\kappa'} \\ & + \hat{R}_{ij} + \hat{Q}_{ij} + \sum_{\substack{k \neq k' \\ k \neq \bar{k}'}} \frac{D^{1_3}_{k'i} D^{1_3}_{\bar{k}'j}}{(\rho_k^2 - \rho_{k'}^2)(\rho_k^2 - \rho_{\bar{k}'}^2)} \times [\rho_k \hat{L}'_{kk'} + \rho_{k'} \hat{\mathscr{L}}'_{\kappa\kappa'}][\rho_k \hat{L}'_{k\bar{k}'} + \rho_{\bar{k}'} \hat{\mathscr{L}}'_{\kappa\bar{\kappa}'}] \\ & + i \sum_{\bar{k} \neq k \neq k'} \frac{D^{1_3}_{k'i} D^{1_3}_{\bar{k}'j}}{(\rho_k^2 - \rho_{k'}^2)(\rho_{k'}^2 - \rho_{\bar{k}'}^2)} \times [\rho_{k'}^2 \hat{L}'_{k'\bar{k}'} + \rho_{k'} \rho_{\bar{k}'} \hat{\mathscr{L}}'_{\kappa'\bar{\kappa}'}], \end{aligned} \tag{2.12d}$$

where

$$\begin{aligned} \hat{R}_{ij} = {} & \frac{D^{1_3}_{1i} D^{1_3}_{1j}}{\rho_1^2} (G'(n-2) - G'(n-3)) + \frac{D^{1_3}_{2i} D^{1_3}_{2j}}{\rho_2^2} (G'(n-1) - G'(n-2) - \hat{\mathscr{L}}'^2_3) \\ & + \frac{D^{1_3}_{3i} D^{1_3}_{3j}}{\rho_3^2} (G'(n) - G'(n-1) - \hat{\mathscr{L}}'^2_1 - \hat{\mathscr{L}}'^2_2), \end{aligned} \tag{2.12e}$$

and

$$\hat{Q}_{ij} = \frac{D^{1_3}_{2i} D^{1_3}_{2j} + D^{1_3}_{3i} D^{1_3}_{2i}}{2i\rho_2\rho_3}$$
$$\times[G'(n-1)\hat{\mathscr{L}}'_1 - \hat{\mathscr{L}}'_1 G'(n-1)$$
$$+ i(\hat{\mathscr{L}}'_2 \hat{\mathscr{L}}'_3 + \hat{\mathscr{L}}'_3 \hat{\mathscr{L}}'_2)] + \frac{1}{2i\rho_1\rho_2}$$
$$\times(D^{1_3}_{1i} D^{1_3}_{2j} + D^{1_3}_{2i} D^{1_3}_{1j})$$
$$\times(G'(n-2)\hat{\mathscr{L}}'_3 - \hat{\mathscr{L}}'_3 G'(n-2)) - \frac{1}{2i\rho_1\rho_3}$$
$$\times(D^{1_3}_{1i} D^{1_3}_{3j} + D^{1_3}_{1j} D^{1_3}_{3i})$$
$$\times(G'(n-2)\hat{\mathscr{L}}'_2 - \hat{\mathscr{L}}'_2 G'(n-2)). \tag{2.12f}$$

By means of Eqs. (2.1) and (2.12), it is straightforward to construct the generators of Sp(6) in terms of the ZD coordinates.

In the next section, we carry out the projection of the collective part of these operators and construct a matrix representation for them.

3. THE COLLECTIVE GENERATORS OF Sp(6)

It is clear from Eq. (1.2) that the Hamiltonian (1.1) is invariant under both the $\mathscr{O}(3)$ and O(n) orthogonal groups. Thus, to find explicitly the eigenstates (1.30) it is convenient to pass from the coordinates X_{is} to the six collective variables ρ_k, ϑ_k, $k = 1,2,3$ and the $3n-6$ angular coordinates ϕ, as indicated in (2.3). This is achieved using Eqs. (2.12a) and (2.12c), i.e.,

$$H_0 = \mathscr{O}^{-1}H\mathscr{O} = \frac{1}{2}\sum_{i=1}^{3}\{\mathscr{O}^{-1}\hat{T}_{ii}\,\mathscr{O} + \hat{q}_{ii}\}$$
$$= \frac{1}{2}\Bigg\{\sum_{k=1}^{3}\Big(-\frac{\partial^2}{\partial\rho_k^2} + \frac{(A-4)(A-6)}{4\rho_k^2}\Big)$$
$$- \sum_{k>k'=1}^{3}\frac{2}{\rho_k^2-\rho_{k'}^2}$$
$$\times\Big(\rho_k\frac{\partial}{\partial\rho_k} - \rho_{k'}\frac{\partial}{\partial\rho_{k'}}\Big)$$
$$+\rho^2 + \sum_{k>k'=1}^{3}\frac{\rho_k^2+\rho_{k'}^2}{(\rho_k^2-\rho_{k'}^2)^2}$$
$$\times[\hat{L}'^2_{kk'} + \hat{\mathscr{L}}'^2_{\kappa\kappa'}] + 4\sum_{k>k'=1}^{3}\frac{\rho_k\rho_{k'}}{(\rho_k^2-\rho_{k'}^2)^2}$$
$$\times L'_{kk'}\hat{\mathscr{L}}'_{\kappa\kappa'} + \frac{1}{\rho_1^2}[G'(n-1) - G'(n-3)]$$
$$+\frac{1}{\rho_2^2}[G'(n-1) - G'(n-2) - \hat{\mathscr{L}}'^2_3]$$
$$+\frac{1}{\rho_3^2}[G'(n) - G'(n-1) - \hat{\mathscr{L}}'^2_1 - \hat{\mathscr{L}}'^2_2]\Bigg\}. \tag{3.1}$$

Due to the structure of (3.1), we can propose the following form for its eigenstates:

$$\Psi = \sum_{\tau_1,\tau_2,\zeta_1} f^{(w_1w_2w_3)}_{\tau_1,\tau_2,\zeta_1}(\rho_1\rho_2\rho_3;\vartheta_k)\,D^{(w_1w_2w_3)}_{\tau_1\tau_2,\zeta_1;\delta\{f\}(r)}(\phi), \tag{3.2}$$

where $D^{(w_1w_2w_3)}_{\tau_1\tau_2,\zeta_1;\delta\{f\}(r)}$ are the irreps of O(n) and $f^{(w_1w_2w_3)}$ is still to be determined and contains the collective structure of the system.

The only operators appearing in (3.1) that are related with the O(n) group are the generators $\hat{\mathscr{L}}'_k$, $k = 1,2,3$ of (2.9) and the Casimir operator $G'(r)$, $r = n, n-1, n-2, n-3$, both of them in the frame of reference "fixed in the body".[8] Thus, the collective wave function $\mathbf{f}^{(w_1w_2w_3)}$ satisfies the equation

$$\mathbf{H}_{\text{coll}}\,\mathbf{f}^{(w_1w_2w_3)} = E\,\mathbf{f}^{(w_1w_2w_3)}, \tag{3.3}$$

where $\mathbf{H}_{\text{coll}}$ is the matrix representation of the operator H_0 given by (3.1) in which $\hat{\mathscr{L}}'_k$, $G'(r)$ are replaced by their matrices $\mathbf{L}'_k$, $\mathbf{G}'(r)$ with respect to the Gel'fand states[14] of O(n),

$$\left|\begin{matrix} w_1 & & w_2 & & w_3 & & 0 & & \cdots & & 0 \\ & \tau_1 & & \tau_2 & & 0 & & \cdots & & 0 & \\ & & \zeta_1 & & 0 & & \cdots & & 0 & & \\ & & & 0 & & \cdots & & 0 & & & \\ & & & & & \vdots & & & & & \\ & & & & & 0 & & & & & \end{matrix}\right\rangle \equiv \left|\begin{matrix} w_1 & & w_2 & & w_3 \\ & \tau_1 & & \tau_2 & \\ & & \zeta_1 & & \end{matrix}\right\rangle. \tag{3.4}$$

This particular form of the Gel'fand–Zetlin states is due to the coordinate transformation (2.3), where the matrix representation $D^{1_n}_{ts}$ of O(n) only depends on the last three rows.[2,8]

The matrices $\mathbf{G}'(r)$ are well known,[13] while the $\mathscr{L}'_k$ were given in Ref. 8, Eqs. (4.4)–(4.8). The latter were taken from the analysis of Pang and Hecht[15] and the phase convention selected so that it agrees with standard results when we restrict ourselves to a group O(3). Thus the collective generators of Sp(6), $\mathbf{B}^+_{ij}$, $\mathbf{B}_{ij}$, and $\mathbf{C}_{ij}$ are obtained by replacing in $\mathscr{O}^{-1}B^+_{ij}\mathscr{O}$, $\mathscr{O}^{-1}B_{ij}\mathscr{O}$, and $\mathscr{O}^{-1}C_{ij}\mathscr{O}$ the operators $\hat{\mathscr{L}}'_k$, $G'(r)$ by their matrices $\mathbf{L}'_k$, $\mathbf{G}'(r)$.

For later convenience, we introduce these generators in the form of an irreducible tensor of $\mathscr{O}(3)$, i.e.,

$$\mathbf{B}^+_{lm} = \sum_{i,j} (11ij|\ lm)\, \mathbf{B}^+_{ij}, \tag{3.5}$$

where $(11ij|\ lm)$ is a mixed coefficient related to the standard spherical Clebsch–Gordan coefficient by means of the relation

$$(11ij|\ lm) = \sum_{\mu,\nu} \langle 1\mu 1\nu|\ lm\rangle\, \Delta_{\mu i}\Delta_{\mu' j}, \tag{3.6}$$

where $\|\Delta_{\mu i}\|$ are the unitary matrices that connect the spherical and Cartesian bases.[16] For $\mathbf{B}_{lm}$ and $\mathbf{C}_{lm}$, we have relations analogous to (3.5).

4. CONTRACTION PROCEDURE OF THE COLLECTIVE Sp(6) GENERATORS

Up to this point we have been concerned with the construction of the generators of the Sp(6) group associated with the collective excitations of the A-nucleon problem interacting through harmonic oscillator forces. We now turn our attention to the large A limit for these operators and first of all discuss the form of the collective Hamiltonian $\mathbf{H}_{\rm coll}$ in this limit. To this end, it is convenient to introduce an additional transformation[8] that relates the collective degrees of freedom ρ_k, $k = 1,2,3$, with those appearing in phenomenological collective models of the nuclei.[17]

We first express the ρ_k, $k = 1,2,3$, in terms of three new variables ρ, b, c through[15]

$$\rho_k^2 = \tfrac{1}{3}\rho^2\{1 + 2b\cos(c - 2\pi k/3)\}. \tag{4.1}$$

In order for the coordinate transformation (2.3) to be bijective one needs the additional conditions[18] $0 \leqslant \rho_2 \leqslant \rho_1 \leqslant \rho_3 \leqslant \infty$. This translates into restrictions for b and c that limit them to the lined triangle in Fig. 1, where $x = b\cos c$, $y = b\sin c$. The restrictions on c, i.e., $0 \leqslant c \leqslant \pi/3$ are the same as those of γ in the Bohr–Mottelson model[19] but, as we see from Fig. 1, b can not exceed 1 so its range is not that of β, which is in the interval $0 \leqslant \beta \leqslant \infty$. Furthermore, instead of ρ we prefer to introduce a variable $\bar{\alpha}$ with the range $-\infty \leqslant \bar{\alpha} \leqslant \infty$. This transformation is defined by[8]

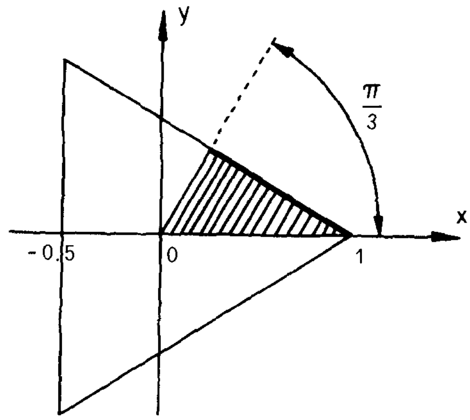

FIG. 1. The variables b, c appearing in Eqs. (4.1) and (4.2) are restricted to the lined triangle, where the coordinates are given by $x = b\cos c$ and $y = b\sin c$.

$$\gamma = c, \tag{4.2a}$$

$$1 + 2\beta^2/\sigma^2 = (1 - b^2)(1 - 3b^2 + 2b^3\cos 3c)^{-1}, \tag{4.2b}$$

$$\bar{\alpha} = \sqrt{2}\,\sigma\ln(\rho/\sigma), \tag{4.2c}$$

where the parameter $\sigma^2 \equiv w_1 + w_2 + w_3 + \frac{3}{2}(n-4)$ is related to the energy of the ground state of the system, and from its construction,[8,9] one can easily prove that to highest order in A, $\sigma^2 \approx A^{4/3}$ so $A \gg 1$ implies $\sigma \gg 1$.

Using Eqs. (4.1) and (4.2), we get for $\sigma \gg 1$ that

$$\begin{aligned}\rho_k^2 = \sigma^2/3 &+ \left[(\sqrt{2}/3)\,\bar{\alpha} + \tfrac{2}{3}\beta\cos(\gamma - 2\pi k/3)\right] \\ &+ \bar{\alpha}^2/3 + (\beta^3/3)\cos 3\gamma\cos(\gamma - 2\pi k/3) \\ &+ (4/3\sqrt{2})\bar{\alpha}\beta\cos(\gamma - 2\pi k/3) + \mathscr{O}(1/\sigma),\end{aligned} \tag{4.3}$$

where $\mathscr{O}(1/\sigma)$ denotes terms of order $1/\sigma$ or smaller.

In order to analyze the matrix representation $\hat{\mathscr{L}}'_k$ in this limit, one needs to write down the corresponding expressions[15] as a power series in the parameter σ. As shown in Ref. 8, one then compares the $\sigma \to \infty$ result with the matrix elements of the generators $\mathbb{C}_{ij}$, $i,j = 1,2,3$ of a $\mathbb{U}(3)$ group, in the unitary Gel'fand and Zetlin basis[14] $\mathbb{U}(3) \supset \mathbb{U}(2) \supset \mathbb{U}(1)$,

$$\left(\begin{matrix} w_1 & & w_2 & & w_3 \\ & \tau'_1 & & \tau'_2 & \\ & & \xi'_1 & & \end{matrix}\ \middle|\ \mathbb{C}_{ij}\ \middle|\ \begin{matrix} w_1 & & w_2 & & w_3 \\ & \tau_1 & & \tau_2 & \\ & & \xi_1 & & \end{matrix}\right) \equiv \mathbb{C}_{ij} \tag{4.4}$$

and finds the correspondence

$$\hat{\mathscr{L}}'_1 \underset{\sigma\to\infty}{\longrightarrow} i(\mathbb{C}_{23} - \mathbb{C}_{32}) \equiv \mathbb{L}'_1, \tag{4.5a}$$

$$\hat{\mathscr{L}}'_2 \underset{\sigma\to\infty}{\longrightarrow} i(\mathbb{C}_{31} - \mathbb{C}_{13}) \equiv \mathbb{L}'_2, \tag{4.5b}$$

$$\mathscr{L}'_3 \underset{\sigma\to\infty}{\longrightarrow} i(\mathbb{C}_{12} - \mathbb{C}_{21}) \equiv \mathbb{L}'_3, \tag{4.5c}$$

where we denote by boldface letters $\mathbf{L}'_k$, $\mathbb{L}'_k$ the matrices associated with the operators $\hat{\mathscr{L}}'_k$, $\mathbb{L}'_k$ in their respective bases (3.4), (4.4). Thus, for arbitrary A one has the expansion[8]

$$\mathbf{L}'_k = \mathbb{L}'_k + (\bar{\mathbb{L}}'_k/\sigma^2), \quad k = 1,2,3. \tag{4.6}$$

By means of Eqs. (4.3) and (4.6), the collective Hamiltonian $\mathbf{H}_{\rm coll}$ can be expressed in a power series of σ:

$$\mathbf{H}_{\rm coll} = \Big(-\frac{\partial^2}{\partial \overline{\alpha}^2} + \overline{\alpha}^2 \Big) \mathbb{I} + \Big(-\frac{1}{\beta^4}\frac{\partial}{\partial\beta}\beta^4\frac{\partial}{\partial\beta} - \frac{1}{\beta^2 \sin 3\gamma}\frac{\partial}{\partial\gamma}\sin 3\gamma \frac{\partial}{\partial\gamma} + \beta^2 \Big)\mathbb{I} + \sum_{k=1}^{3} \frac{1}{4\beta^2 \sin^2(\gamma - 2\pi k/3)} \mathbf{J}_k'^2 + \mathscr{O}(1/\sigma), \tag{4.7}$$

where $\mathbb{I} = \| \delta_{\tau_1'\tau_1}\, \delta_{\tau_2'\tau_2}\, \delta_{\varsigma_1'\varsigma_1} \|$ is the unit matrix and

$$\mathbf{J}_k' = L_k' \,\mathbb{I} + \mathbb{L}_k', \quad k = 1,2,3. \tag{4.8}$$

The Hamiltonian (4.7) has the structure of a one-dimensional oscillator in the variable $\overline{\alpha}$ plus a five-dimensional oscillator of the Bohr–Mottelson type. However, instead of the usual $\hat{L}_k'$ angular momentum components of the Bohr–Mottelson equation, one encounters a "total angular momentum" $\mathbf{J}_k'$ composed of the latter and an intrinsic operator $\mathbb{L}_k'$ that behaves as an additional spin for the system.

Clearly $\mathbf{H}_{\rm coll}$ has not, in general, a U(6) symmetry group, but has an additional degeneracy brought about by the presence of $\mathbb{L}_k'$.[20]

For the unphysical situation where $(w_1 w_2 w_3) = (0,0,0)$ for the O(n) irrep, i.e., for the scalar representation, or for the case of closed shells, $(w_1,w_2,w_3) = (w,w,w)$, $\mathbb{L}_k'$ would play no role, and indeed a U(6) symmetry group would appear. These cases would correspond effectively to $\mathbf{J}_k' = \hat{L}_k'\mathbb{I}$ in (4.8), and thus to no additional degeneracy for the states. Moreover, for these cases, U(6) is the symmetry group of $\mathbf{H}_{\rm coll}$ for any A, as shown by several authors.[5,21,22]

For the case of an arbitrary O(n) irrep, it remains to determine the symmetry group associated to this additional degeneracy. To investigate this point, in the remainder of this section, we analyze the behavior of the collective Sp(6) generators (3.5) in the same large A situation and then search for the hidden symmetry group for the Hamiltonian (4.7). To this end, we expand the generators (3.5) in a power series in the parameter σ and then proceed to the limiting situation, which, in effect, is a contraction procedure.[23]

Using the representation matrix $\mathbf{B}_{lm}^+$ together with Eqs. (4.3) and (4.6), we find

$$\mathbf{B}_{00}^+ = \sqrt{\tfrac{2}{3}}\,\sigma \Big(-\overline{\alpha} + \frac{\partial}{\partial \overline{\alpha}} \Big)\mathbb{I} + \mathscr{O}(\sigma^0) \tag{4.9a}$$

and

$$\mathbf{B}_{2m}^+ = \sqrt{\tfrac{2}{3}}\,\sigma \Big\{ \Big[F_{2m}(\gamma,\vartheta_i)\Big(-\frac{\partial}{\partial\beta} + \beta \Big) + F_{2m}(\gamma - \pi/2, \vartheta_i)\frac{1}{\beta}\frac{\partial}{\partial\gamma}\Big]\mathbb{I} + \frac{1}{2\sqrt{2}\beta} \times \Big[(D^2_{2m} - D^2_{-2m})\frac{\mathbf{J}_3'}{s_3} + i(D^2_{1m} - D^2_{-1m})\frac{\mathbf{J}_2'}{s_2} - (D^2_{1m} + D^2_{-1m})\frac{\mathbf{J}_1'}{s_1}\Big]\Big\} + \mathscr{O}(\sigma^0), \tag{4.9b}$$

where $s_k \equiv \sin(\gamma - 2\pi k/3)$, $k = 1,2,3$, $D^L_{mm'}$ is the rotation matrix of three dimensions in spherical components, and

$$F_{2m}(\gamma, \vartheta_i) \equiv \cos\gamma\, D^2_{0m} + \frac{1}{\sqrt{2}}\sin\gamma (D^2_{2m} + D^2_{-2m}). \tag{4.9c}$$

To obtain the $\mathbf{B}^{lm}$ with $l = 0,2$ we just evaluate the adjoint of $\mathbf{B}_{lm}^+$ with $l = 0,2$, Eqs. (4.9a) and (4.9b). Similarly, for the $\mathbf{C}_{lm}$ we get

$$\mathbf{C}_{00} = -(1/\sqrt{3})\,\sigma^2\,\mathbb{I} - (1/\sqrt{3})\,\mathbf{H}_{\rm coll} + \mathscr{O}(1/\sigma), \tag{4.10a}$$

$$\mathbf{C}_{1m} = -(1/\sqrt{2})\,\hat{L}_m\,\mathbb{I}, \tag{4.10b}$$

where $\hat{L}_m$, $m = 1,0,-1$ are the spherical components of the orbital angular momentum fixed in space, which is independent of σ. Finally, the corresponding expression for the quadrupole generators $\mathbf{C}_{2m}$ is given by

$$\mathbf{C}_{2m} = -\frac{1}{\sqrt{6}}F_{2m}(-2\gamma,\vartheta_i)\Big[\frac{\partial^2}{\partial\beta^2} - \frac{1}{\beta}\frac{\partial}{\partial\beta} - \frac{1}{\beta^2}\frac{\partial^2}{\partial\gamma^2} - \beta^2\Big]\mathbb{I} + \sqrt{\tfrac{2}{3}}\,F_{2m}(-2\gamma - \pi/2, \vartheta_i)\Big[\frac{1}{\beta}\frac{\partial^2}{\partial\beta\,\partial\gamma} - \frac{1}{\beta^2}\frac{\partial}{\partial\gamma}\Big]\mathbb{I} + \frac{2}{\sqrt{3}}F_{2m}(\gamma - \pi/2,\vartheta_i)\frac{1}{\beta}\frac{\partial^2}{\partial\gamma\,\partial\overline{\alpha}}\mathbb{I}$$
$$-\frac{2}{\sqrt{3}}F_{2m}(\gamma,\vartheta_i)\Big[\frac{\partial^2}{\partial\overline{\alpha}\,\partial\beta} - \overline{\alpha}\beta - \frac{3}{2\sqrt{2}\beta^2\sin 3\gamma}\frac{\partial}{\partial\gamma}\Big]\mathbb{I} - \frac{1}{8}\sqrt{\tfrac{3}{2}}\,D^2_{0m}\frac{\mathbf{J}_3'^2}{\beta^2\sin^2\gamma} + \frac{1}{8\sqrt{6}\beta^2}D^2_{0m}\sum_{k=1}^{3}\frac{J_k'^2}{\sin^2(\gamma - 2\pi k/3)}$$

$$+\frac{(D^2_{2m}-D^2_{-2m})}{2\sqrt{3}}\mathbf{J}_3'\left[-\frac{\cot\gamma}{\beta}\frac{\partial}{\partial\beta}+\frac{1}{\beta^2}\frac{\partial}{\partial\gamma}+\frac{\sqrt{2}}{\beta\sin\gamma}\frac{\partial}{\partial\overline{\alpha}}+\frac{\sqrt{3}\sin(\gamma-2\pi/3)}{\beta^2\sin 3\gamma}\right]$$

$$+\frac{i(D^2_{1m}-D^2_{-1m})}{2\sqrt{3}}\mathbf{J}_2'\left[-\frac{\cot(\gamma-4\pi/3)}{\beta}\frac{\partial}{\partial\beta}+\frac{1}{\beta^2}\frac{\partial}{\partial\gamma}+\frac{\sqrt{2}}{\beta\sin(\gamma-4\pi/3)}\frac{\partial}{\partial\overline{\alpha}}+\frac{\sqrt{3}\sin\gamma}{\beta^2\sin 3\gamma}\right]$$

$$+\frac{(D^2_{1m}+D^2_{-1m})}{2\sqrt{3}}\mathbf{J}_1'\left[\frac{\cot(\gamma-2\pi/3)}{\beta}\frac{\partial}{\partial\beta}-\frac{1}{\beta^2}\frac{\partial}{\partial\gamma}-\frac{\sqrt{2}}{\beta\sin(\gamma-2\pi/3)}\frac{\partial}{\partial\overline{\alpha}}-\frac{\sqrt{3}\sin(\gamma-4\pi/3)}{\beta^2\sin 3\gamma}\right]$$

$$+\frac{(D^2_{2m}+D^2_{-2m})}{16\beta^2}\left[\frac{\mathbf{J}_2'^2}{\sin^2(\gamma-4\pi/3)}-\frac{\mathbf{J}_1'^2}{\sin^2(\gamma-2\pi/3)}\right]$$

$$+\frac{\sin(\gamma-4\pi/3)}{2\beta^2\sin 3\gamma}(D^2_{1m}-D^2_{-1m})\mathbf{J}_1'\mathbf{J}_3'+\frac{\sin\gamma}{2i\beta^2\sin 3\gamma}(D^2_{2m}-D^2_{-2m})\mathbf{J}_2'\mathbf{J}_1'$$

$$-\frac{\sin(\gamma-2\pi/3)}{2i\beta^2\sin 3\gamma}(D^2_{1m}+D^2_{-1m})\mathbf{J}_3'\mathbf{J}_2'+\mathbb{S}_{2m}+\mathcal{O}(1/\sigma), \tag{4.10c}$$

where $F_{2m}(\gamma,\vartheta_i)$ is given by Eq. (4.9c) and the operator $\mathbb{S}_{2m}$ is given by

$$\mathbb{S}_{2m}=\lim_{\sigma\to\infty}\sum_{i,j}(11ij|2m)\,\mathbf{Q}_{ij}$$
$$=-i(D^2_{1m}+D^2_{-1m})(\mathbb{C}_{23}+\mathbb{C}_{32})+(D^2_{1m}-D^2_{-1m})(\mathbb{C}_{31}+\mathbb{C}_{13})+i(D^2_{2m}-D^2_{-2m})(\mathbb{C}_{12}+\mathbb{C}_{21}), \tag{4.11}$$

with $\mathbf{Q}_{ij}$ and $\mathbb{C}_{ij}$ defined in (2.12f) and (4.4).

These collective generators satisfy the Sp(6) commutation relations

$$[\mathbf{B}_{l'm'},\mathbf{B}^+_{lm}]=-4\sqrt{(2l+1)(2l'+1)}\sum_{\lambda}W(l'l11;\lambda 1)\langle l'm'lm|\lambda m'+m\rangle\,\mathbf{C}_{\lambda m'+m}, \tag{4.12a}$$

$$[\mathbf{C}_{l'm'},\mathbf{B}_{lm}]=\sqrt{(2l+1)(2l'+1)}\sum_{\lambda\,(\mathrm{even})}2W(l'l11;\lambda 1)\langle l'm'lm|\lambda m'+m\rangle\,\mathbf{B}_{\lambda m'+m}, \tag{4.12b}$$

$$[\mathbf{C}_{l'm'},\mathbf{B}^+_{lm}]=(-1)^{l'+1}\sqrt{(2l+1)(2l'+1)}\sum_{\lambda\,(\mathrm{even})}2W(l'l11;\lambda 1)\langle l'm'lm|\lambda m'+m\rangle\mathbf{B}^+_{\lambda m'+m}, \tag{4.12c}$$

$$[\mathbf{C}_{l'm'},\mathbf{C}_{lm}]=-\sqrt{(2l+1)(2l'+1)}\sum_{\lambda}[(-1)^{l-l'-\lambda}-1]\,W(l'l11;\lambda 1)\,\langle l'm'lm|\lambda m'+m\rangle\,\mathbf{C}_{\lambda m'+m}, \tag{4.12d}$$

where $\langle l'm'lm|\lambda\mu\rangle$ is a standard Clebsch–Gordan coefficient and $W(l'l11;\lambda 1)$ is a Racah coefficient.

If we now define the operators

$$\mathbf{b}^+_{lm}\equiv\lim_{\sigma\to\infty}\left(-\frac{\sqrt{3}}{2\sigma}\mathbf{B}^+_{lm}\right),\quad l=0,2 \tag{4.13a}$$

and

$$\mathbf{C}_{lm}\equiv\lim_{\sigma\to\infty}\left(\mathbf{C}_{lm}+\frac{\sigma^2}{\sqrt{3}}\delta_{l0}\,\mathbb{I}\right),\quad l=0,1,2, \tag{4.13b}$$

we notice that their explicit form is given by the highest-order term in formulas (4.3), (4.9), and (4.10). Again, $\mathbf{b}^{lm}\equiv\lim_{\sigma\to\infty}(-(\sqrt{3}/2\sigma)\,\mathbf{B}^{lm})$ is just the adjoint of $\mathbf{b}^+_{lm}$.

As pointed out above, the limiting procedure specified by definitions (4.12) amounts to a contraction mechanism[21] of the collective Sp(6) generators. By means of Eqs. (4.12), it is now a straightforward exercise to compute the commutator algebra associated to the generators $\mathbf{b}^+_{lm}$, $\mathbf{b}^{lm}$, $\mathbf{C}_{lm}$, which is readily identified with that of a W(6) $\wedge$ U(3) Lie algebra, where $\wedge$ indicates a semidirect product, W(6) is a Weyl group[24] in six dimensions, and U(3) is a unitary group in three dimensions.

Thus, we conclude that W(6) $\wedge$ U(3) constitutes the dynamical group for the $A\gg 1$ collective Hamiltonian (4.5). In fact, one can verify directly that

$$\mathbf{H}_{\mathrm{coll}}=2\sum_{l,m}\mathbf{b}^+_{lm}\,\mathbf{b}^{lm}+6\,\mathbb{I}. \tag{4.14}$$

What is the symmetry group of $\mathbf{H}_{\mathrm{coll}}$? It is quite obvious from the form (4.14) for $\mathbf{H}_{\mathrm{coll}}$ that the set of operators

$$\mathscr{C}^{l'm'}_{lm}=\mathbf{b}^+_{lm}\,\mathbf{b}^{l'm'};\quad l,l'=0,2 \tag{4.15}$$

commute with it and generate a U(6) group. In regards to the U(3) group generated by the $\mathbf{C}_{lm}$, $l=0,1,2$, Eq. (4.13b), we readily find

$$\mathbf{C}_{00}=-\frac{1}{\sqrt{3}}\mathbf{H}_{\mathrm{coll}}=-\frac{1}{\sqrt{3}}\left\{2\sum_{l,m}\mathscr{C}^{lm}_{lm}+6\,\mathbb{I}\right\}, \tag{4.16}$$

and thus $\mathbf{C}_{00}$ is already contained in the U(6) algebra generated by the operators (4.15). The SU(3) subalgebra generated by $\mathbf{C}_{1m}$, $\mathbf{C}_{2m}$ is not, however, contained in the U(6) algebra, as we proceed to show. We first construct the tensor operators that generate the SU(3) subgroup of U(6), which we denote by SU*(3). As is well known,[25] these are given by

$$\mathbb{T}_{1m} = \sqrt{6} \sum_{\mu\mu'} \langle 2\mu 1m | 2\mu' \rangle \, \mathbf{b}^{+}_{2\mu'} \mathbf{b}^{2\mu}, \tag{4.17}$$

$$\mathbb{Q}_{2m} = \sqrt{\tfrac{7}{3}} \, [\mathbf{b}_2^{+} \times \mathbf{b}_2]^2_m + \sqrt{\tfrac{4}{3}} (\mathbf{b}^{+}_{00} \mathbf{b}_{2m} + \mathbf{b}^{+}_{2m} \mathbf{b}_{00}). \tag{4.18}$$

Substituting the explicit expressions for $\mathbf{b}^{+}_{lm}$ and $\mathbf{b}_{lm}$ and comparing with (4.10b) and (4.10c), we find, after a straightforward but lengthy calculation, the relations

$$\begin{aligned} \mathbf{C}_{1m} = & -(1/\sqrt{2}) \{ \mathbb{T}_{1m} + D^{\prime}_{0m} \mathbb{L}^{\prime}_3 \\ & + (1/\sqrt{2})(D^{1}_{-1m} - D^{1}_{1m}) \mathbb{L}^{\prime}_1 \\ & - (i/\sqrt{2})(D^{1}_{-1m} + D^{1}_{1m}) \mathbb{L}^{\prime}_2 \} , \end{aligned} \tag{4.19}$$

$$\mathbf{C}_{2m} = -\mathbb{Q}_{2m} + \mathbb{S}_{2m}, \tag{4.20}$$

where $\mathbb{S}_{2m}$ was defined in (4.11). Thus, it is clear that the two groups SU(3) and SU*(3) are distinct. Furthermore, since $[\mathbf{H}_{\text{coll}}, \mathbf{C}_{lm}] = -\sqrt{3}\,[\mathbf{C}_{00}, \mathbf{C}_{lm}] = 0$, $l = 1,2$, SU(3) is an additional symmetry group for $\mathbf{H}_{\text{coll}}$.

We now compute the commutators between the U(6) and SU(3) generators using the results of the contraction procedure:

$$\begin{aligned} & [\mathscr{C}^{l''m''}_{l'm'}, \mathbf{C}_{lm}] \\ & = \Big\{ (-1)^{m+1} \sqrt{(2l+1)(2l''+1)} \\ & \quad \times \sum_{\lambda} [(-1)^{\lambda} + 1] \, W(ll''11; \lambda 1) \\ & \quad \times \langle lm, l''-m'' | \lambda m - m'' \rangle \, \mathscr{C}^{\lambda m''-m}_{l'm'} \\ & \quad + (-1)^{l} \sqrt{(2l+1)(2l'+1)} \sum_{\lambda} [(-1)^{\lambda} + 1] \\ & \quad \times W(ll'11; \lambda 1) \, \langle lm, l'm' | \lambda m + m' \rangle \, \mathscr{C}^{l''m''}_{\lambda m+m'} \Big\} . \end{aligned} \tag{4.21}$$

We conclude that the hidden symmetry group for $\mathbf{H}_{\text{coll}}$ is the semidirect product group

$$G = \mathrm{U}(6) \wedge \mathrm{SU}(3). \tag{4.22}$$

From (4.19) and (4.20), it is clear that for the particular case where one takes either the scalar or the closed-shell irrep of $\mathrm{O}(A-1)$, the SU(3) and SU*(3) groups are identical, since the $\mathbf{C}_{ij}$, $i \neq j$ of (4.4) vanish and, consequently, the operators $\mathbb{L}^{\prime}_k$ and $\mathbb{S}_{2m}$ also disappear. Thus, for these cases, $G = \mathrm{U}(6)$, in accordance with previous investigations.[5,21,22]

5. CONCLUSIONS

In this paper we have analyzed the group-theoretical structure for a many-body system interacting through harmonic oscillator forces and projected its collective Hamiltonian, following the ideas proposed in recent investigations.[2,3] We focused our attention on the large A limit for this system and, in particular, on the accidental degeneracy present in the collective Hamiltonian, Eq. (4.7), for this case. By carrying out a group contraction on the collective dynamical group generators, we were able to conclude that the symmetry group responsible for this degeneracy is the semidirect product group U(6) $\wedge$ SU(3).

This analysis differs from previous ones by Deenen and Quesne[5] and Kramer[22] in that we consider the collective subspace defined by an arbitrary O(n) irrep and also in that our realization for the Sp(6) generators is in terms of the geometrically appealing ZD coordinates, in contrast with the Barut–Hilbert space[5] and Bargman–Hilbert[22] space realizations, respectively, by these authors. The explicit construction of these generators, when projected to a definite O(n) irrep as given by Eqs. (4.9) and (4.10), permits, in the large A limit, a straightforward identification of the symmetry group for the collective Hamiltonian and, as a bonus, an explicit realization for the U(6) and SU(3) generators.

As a next step, it would be important to study this problem in the case of arbitrary A, or to start by considering a small number of particles in the same spirit as the work of Chacón *et al.*[7] for the O(n) scalar collective Hamiltonian.

ACKNOWLEDGMENTS

We would like to thank E. Chacón, J. Deenen, P. Hess, C. Quesne, and specially M. Moshinsky for many helpful comments on this work.

[1] See, e.g., J. M. Eisenberg and W. Greiner, *Microscopic Theory of the Nucleus*, Vols. I–III (North–Holland, Amsterdam, 1976).

[2] V. Vanagas, "The Microscopic Nuclear Theory," Lecture Notes in Physics (University of Toronto, Toronto, 1977); "The Microscopic Theory of the Collective Motion in Nuclei," in *Group Theory and Its Applications in Physics-1980*, edited by T. H. Seligman (AIP, New York, 1980), p. 220.

[3] G. F. Filippov, V. I. Ovcharenko, and Yu. F. Smirnov, *Microscopic Theory of Collective Excitations in Nuclei* (in Russian) (Navkova Dumka, Kiev, 1981).

[4] D. J. Rowe, "The Microscopic Realization of Nuclear Collective Models," in *Group Theory and Its Applications in Physics-1980*, edited by T. H. Seligman (AIP, New York, 1980), p. 177.

[5] J. Deenen and C. Quesne, J. Math. Phys. **23**, 2004 (1982); J. Deenen and C. Quesne (private communication).

[6] B. Buck, L. C. Biedenharn, and R. Cusson, Nucl. Phys. A **317**, 205 (1979); O. L. Weaver, R. Cusson, and L. C. Biedenharn, Ann. Phys. (N.Y.) **102**, 493 (1976).

[7] E. Chacón, M. Moshinsky, and V. Vanagas, J. Math. Phys. **22**, 605 (1981).

[8] O. Castaños, A. Frank, E. Chacón, P. O. Hess, and M. Moshinsky, J. Math. Phys. **23**, 2537 (1982).

[9] L. Sabaliauskas, Liet. Fiz. Rinkinys **19**, 5 (1979) (in Russian).

[10] P. Kramer and M. Moshinsky, Nucl. Phys. **82**, 241 (1966).

[11] W. Zickendraht, J. Math. Phys. **10**, 30 (1969); **12**, 1663 (1971); A. Ya Dzublik, V. I. Ovcharenko, A. I. Steshenko, and G. F. Filippov, Yad. Fiz. **15**, 869 (1972) [Sov. J. Nucl. Phys. **15**, 487 (1972)].

[12] M. Moshinsky, Nucl. Phys. A **354**, 257 (1981).

[13] M. Moshinsky, *Group Theory and the Many Body Problem* (Gordon and Breach, New York, 1967).

[14] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR **71**, 1017 (1950).

[15] S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).

[16] M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).

[17] O. Castaños, A. Frank, E. Chacón, P. O. Hess, and M. Moshinsky, Phys. Rev. C **25**, 1611 (1982).

[18] Yu. F. Smirnov and G. F. Filippov, Yad. Fiz. **27**, 73 (1978) [Sov. J. Nucl. Phys. **27**, 39 (1978)].

[19] A. Bohr, B. Mottelson, and K. Dan Vidensk, Selsk. Mat. Fys. Medd. **27**, 16 (1953).

[20] M. Moshinsky and C. Quesne, preprint.

[21] M. Moshinsky, "Proceedings of the X International Colloquium on Group Theoretical Methods in Physics," Physica A **114**, 322 (1982).

[22] P. Kramer, preprint, University of Tübingen (1982).

[23] R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).

[24] B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1973).

[25] O. Castaños, E. Chacón, A. Frank, and M. Moshinsky, J. Math. Phys. **20**, 35 (1979).

The ballooning spectrum of rotating plasmas

Eliezer Hameiri and Peter Laurence
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012

(Received 15 April 1983; accepted for publication 9 September 1983)

Ballooning modes are shown to be part of the spectrum by using a "singular sequence" of localized modes. We show that the modes arise from Alfven and slow magnetosonic waves propagating along rays confined inside the plasma. Different ballooning modes are seen, depending on the particular rotating frame of observation, indicating that there are accumulation points of eigenvalues. The effect of rigidly rotating flow is seen to be destabilizing due to an analog of the Rayleigh–Taylor instability associated with density gradients in the presence of a centrifugal force. Flow shear also modifies the stability criterion. A certain component of the flow shear will eliminate the ballooning modes.

PACS numbers: 52.30. + r, 52.35.Bj, 52.55.Gb, 02.30.Jr

I. INTRODUCTION

Ballooning instabilities,[1,2] driven by the pressure gradient of a confined plasma in the presence of convex magnetic field lines, are believed to limit the plasma beta—the ratio of plasma pressure to magnetic pressure—to just a few percent in toroidal devices. In axisymmetric toroids they are thought to provide a means of determining the behavior of the unstable spectrum corresponding to individual mode numbers m as m tends towards infinity. In deriving the beta limits, it is generally assumed that the underlying equilibrium state contains no mass flow. This, however, is not always the case. In particular, strong flows of Mach number close to unity can be generated after heating the plasma by neutral beam injection. One of the purposes of this work is to investigate the effect of equilibrium state flows on the ballooning stability. Indeed, we find in Sec. IV that flow effects will be of the same order of magnitude as magnetic curvature effects.

A second aim of this article is to clarify the physical origins and the mathematical understanding of the ballooning modes. We will show that they arise from the presence of waves in the plasma which propagate along rays which never intersect the boundary, namely, the Alfven and slow magnetosonic waves. This phenomenon gives rise to modes localized about such rays, and the question of their stability involves a system of ordinary differential equations along these rays, whose relevant spectral properties can be determined from the spectrum of just one second-order ordinary differential equation. The ballooning equations are derived using not the common eikonal representation[1–4] but rather by a mathematically and algebraically more appealing device of constructing so-called singular sequences,[5] as was done for shearless magnetic fields.[6] Moreover, as the rays we will consider are ergodic, no recourse needs to be made to the so-called ballooning representation or to the Fourier transform technique, which involve as yet unresolved issues relating to the convergence of the series present in such representations and their connection with the physical eigenfunction.[4] Our approach, however, is not as natural for the investigation of "finite m corrections" to the ballooning modes which were treated for static equilibria,[2,4] and this work is restricted to studying the infinitely localized modes only.

The main case we consider is an axisymmetric plasma equilibrium in a toroidal configuration, where the plasma rotates both in the toroidal and poloidal directions. Because of the symmetry, it is possible to get the spectrum of modes by Fourier analyzing the system and to consider modes which behave like $\exp[i(m\theta - \omega t)]$, where θ is the ignorable toroidal angle. The spectrum of the entire system, including the ballooning modes, is simply the closure of the union of the spectra with finite m-number, and, in particular, the ballooning modes in the axisymmetric case are limit points of eigenvalues $\omega(m)$ as $m \to \infty$. This observation leads to the remarkable result that the ballooning spectrum depends on the coordinate frame which an observer uses. An observer rotating with angular frequency Ω_0 in the θ direction measures a toroidal angle $\theta' = \theta - \Omega_0 t$ and sees the phase $m\theta - \omega t$ as $m\theta' - (\omega - m\Omega_0)t$, that is, sees the wave frequency as $\omega' = \omega - m\Omega_0$. Clearly, finite accumulation points of $\omega(m)$ will go to infinity under this transformation of the spectrum, but new accumulation points may appear. It follows that, in principle, ballooning mode studies should be carried out in all possible rotating frames in order to obtain a better stability criterion. We will see that the most dangerous ballooning modes are observed in a frame rotating with the flow frequency itself, if this frequency is constant.

This work is structured as follows. In the next section we collect some known results on the equilibrium state of rotating plasmas[7,8] and on the nonlinear eigenvalue equation,[9] which determines its linear stability. In Sec. III we derive the ballooning mode equation in the frame of the moving plasma, which is close in form to the static case. We also derive a sufficient condition for the stability of these modes. Section IV discusses the effect of the flow on the stability of these modes, using a large aspect ratio asymptotic expansion. In Sec. V we present a detailed derivation of the ballooning modes arising from a single Alfven wave ray, and in Sec. VI a briefer calculation is presented for the slow wave ray.

II. EQUILIBRIUM AND STABILITY EQUATIONS

Consider a plasma confined in a perfectly conducting axisymmetric torus and obeying the ideal magnetohydrodynamics (MHD) equations

 0022-2488/84/020396-10$02.50

$$\rho \mathbf{u}_t + \rho \mathbf{u}\cdot\nabla \mathbf{u} + \nabla p = \mathbf{J}\times\mathbf{B},$$
$$\rho_t + \operatorname{div}(\rho \mathbf{u}) = 0,$$
$$\mathbf{B}_t + \operatorname{curl}(\mathbf{B}\times\mathbf{u}) = 0, \tag{1}$$
$$S_t + \mathbf{u}\cdot\nabla S = 0,$$
$$\operatorname{div}\mathbf{B} = 0, \quad \mathbf{J}\equiv\operatorname{curl}\mathbf{B}, \quad p = S\rho^{\gamma},$$

where $\mathbf{B}, \mathbf{u}, p, \rho$ are the magnetic field and the plasma velocity, pressure, and density, respectively. S is a function of the specific entropy and the last equation is an equation of state, where γ is the ratio of the specific heats. Appropriate boundary conditions for this configuration are $\mathbf{u}\cdot\mathbf{n} = \mathbf{B}\cdot\mathbf{n} = 0$ at the wall, where $\mathbf{n}$ is the normal to the boundary.

Using cylindrical coordinates (r,θ,z) and denoting by $\psi(r,z)$ the poloidal magnetic flux function such that $\mathbf{B} = \nabla\psi\times\nabla\theta + B_\theta\hat{\boldsymbol{\theta}}$, it is known[7,8] that an equilibrium flow field must be within ψ-surfaces,

$$\mathbf{u} = (1/\rho)\Phi(\psi)\mathbf{B} + r\Omega(\psi)\hat{\boldsymbol{\theta}}, \tag{2}$$

where Φ and Ω are some given function of ψ. Equation (2) states that $\mathbf{u}$ is parallel to $\mathbf{B}$ up to a rigid rotation of each individual ψ-surface. We will consider the case of a sheared magnetic field, where almost every ψ-surface is covered ergodically by a single field line. In this case if $\Phi \neq 0$ then $\mathbf{B}\cdot\nabla S = 0$, or $S = S(\psi)$. Thus the presence of a poloidal flow requires the temperature T, $T\equiv p/\rho$, to vary on ψ-surfaces. Experimentally, however, one expects to see isothermal flux surfaces, $T = T(\psi)$, with the poloidal flow damped out due to the magnetic pumping effect in the torus.[10] An equilibrium state may be obtained after specifying two more arbitrary functions of ψ and then solving a second-order partial differential equation for $\psi(r,z)$, which is known to be elliptic[8,11] if $\Phi^2/\rho < \beta$, where

$$\beta = \gamma p/(\gamma p + \mathbf{B}^2). \tag{3}$$

An interesting and useful property of the equilibrium state[8] is that the vector $\mathbf{B} - \Phi\mathbf{u}$, and its curl, or instead $\mathbf{J} - \Phi\operatorname{curl}\mathbf{u}$, must be in the ψ-surface, generalizing the well-known result for the static equilibrium state. We remark that if one observes the system from a rotating coordinate frame rotating in the toroidal direction with angular frequency Ω_0, system (1) remains unchanged except that $\mathbf{u}$ has to be replaced by $\hat{\mathbf{u}} = \mathbf{u} - r\Omega_0\hat{\boldsymbol{\theta}}$ and the right-hand side of the momentum equation has to be replaced by $\mathbf{J}\times\mathbf{B} + \rho r\Omega_0^2\hat{\mathbf{r}} + 2\rho\Omega_0\hat{\mathbf{u}}\times\hat{\mathbf{z}}$, corresponding to the additional centrifugal and Coriolis forces. Equation (2) remains unchanged except that $\Omega\rightarrow\hat{\Omega}\equiv\Omega - \Omega_0$.

Linear stability of the equilibrium state is found by linearizing Eq. (1) about it. Introducing the Lagrangian displacement vector[9] $\boldsymbol{\xi}$ via

$$\mathbf{u}_1 = \boldsymbol{\xi}_t + \hat{\mathbf{u}}\cdot\nabla\boldsymbol{\xi} - \boldsymbol{\xi}\cdot\nabla\hat{\mathbf{u}}, \tag{4}$$

where the subscript 1 indicates perturbed Eulerian quantities and quantities without subscripts are equilibrium quantities, and then expressing

$$\mathbf{B}_1 = \nabla\times(\boldsymbol{\xi}\times\mathbf{B}),$$
$$\rho_1 = -\nabla\cdot(\rho\boldsymbol{\xi}),$$
$$p_1 = -\boldsymbol{\xi}\cdot\nabla p - \gamma p\nabla\cdot\boldsymbol{\xi},$$

one gets from the momentum equation the second-order equation[9]

$$\rho\boldsymbol{\xi}_{tt} + 2A\,\boldsymbol{\xi}_t + F\boldsymbol{\xi} = 0. \tag{5}$$

A is an anti-Hermitian operator,

$$A\,\boldsymbol{\xi} = \rho\hat{\mathbf{u}}\cdot\nabla\boldsymbol{\xi} + \rho\Omega_0\hat{\mathbf{z}}\times\boldsymbol{\xi}, \tag{6}$$

and F is a Hermitian operator under the boundary condition $\boldsymbol{\xi}\cdot\mathbf{n} = 0$ and the usual inner product $(\boldsymbol{\xi},\boldsymbol{\eta}) = \int\boldsymbol{\xi}\cdot\boldsymbol{\eta}^*\,d^3\mathbf{x}$, given by

$$\begin{aligned} F\boldsymbol{\xi} = {} & \nabla(\gamma p\nabla\cdot\boldsymbol{\xi} + \boldsymbol{\xi}\cdot\nabla p - \mathbf{B}\cdot\nabla\times(\boldsymbol{\xi}\times\mathbf{B})) \\ & + \mathbf{B}\cdot\nabla(\nabla\times(\boldsymbol{\xi}\times\mathbf{B})) + (\nabla\times(\boldsymbol{\xi}\times\mathbf{B}))\cdot\nabla\mathbf{B} \\ & + \nabla\cdot(\rho\boldsymbol{\xi}\hat{\mathbf{u}}\cdot\nabla\hat{\mathbf{u}} - \rho\hat{\mathbf{u}}\hat{\mathbf{u}}\cdot\nabla\boldsymbol{\xi}) - 2\rho(\hat{\mathbf{u}}\cdot\nabla\boldsymbol{\xi} - \boldsymbol{\xi}\cdot\nabla\hat{\mathbf{u}})\times\hat{\mathbf{z}} \\ & + 2\nabla\cdot(\rho\boldsymbol{\xi})\hat{\mathbf{u}}\times\hat{\mathbf{z}} - \nabla\cdot(\rho\boldsymbol{\xi})\Omega_0^2 r\hat{\mathbf{r}}. \end{aligned} \tag{7}$$

In discussing stability, we follow Ref. 8 and single out the perturbed total pressure $p_* = p_1 + \mathbf{B}\cdot\mathbf{B}_1$. Defining

$$\mathbf{b} = \mathbf{B}\cdot\nabla\boldsymbol{\xi} - \rho\boldsymbol{\xi}\cdot\nabla(\mathbf{B}/\rho), \qquad s = (p/S)\boldsymbol{\xi}\cdot\nabla S, \tag{8}$$

one gets for the perturbed quantities

$$\rho_1 = \frac{\rho}{\gamma p + \mathbf{B}^2}(s - \mathbf{B}\cdot\mathbf{b} + p_*),$$
$$\mathbf{B}_1 = \left[I - \frac{1}{\gamma p + \mathbf{B}^2}\mathbf{B}\mathbf{B}\right]:\mathbf{b} + \frac{1}{\gamma p + \mathbf{B}^2}(s + p_*)\mathbf{B}, \tag{9}$$
$$p_* = \mathbf{B}\cdot(\mathbf{B}\cdot\nabla\boldsymbol{\xi} - \boldsymbol{\xi}\cdot\nabla\mathbf{B}) - \boldsymbol{\xi}\cdot\nabla p - (\gamma p + \mathbf{B}^2)\operatorname{div}\boldsymbol{\xi},$$

where I is the identity operator. It will be seen useful to express

$$F(\boldsymbol{\xi}) = \hat{F}(\boldsymbol{\xi}) + G(p_*), \tag{10}$$

where G operates on $\boldsymbol{\xi}$ only through p_* and $\hat{F}$ involves derivatives only of the form $\mathbf{B}\cdot\nabla$ and $\hat{\mathbf{u}}\cdot\nabla$, but no derivatives across ψ-surfaces. One has[8]

$$(G\,[p_*(\boldsymbol{\xi})], \boldsymbol{\xi}) = \int\frac{1}{\gamma p + \mathbf{B}^2}|p_*|^2\,d^3\mathbf{x}, \tag{11}$$

$$\begin{aligned} (\hat{F}\boldsymbol{\xi},\boldsymbol{\xi}) = \int\Big\{ & -\frac{1}{\rho}|\Phi\mathbf{b} - (\boldsymbol{\xi}\cdot\nabla\Phi)\mathbf{B}|^2 + Q\hat{\mathbf{B}}_1\cdot\hat{\mathbf{B}}_1^* \\ & - \frac{1}{\gamma S}(\boldsymbol{\xi}\cdot\nabla S)(\boldsymbol{\xi}^*\cdot\nabla p) + \mathbf{b}\times\boldsymbol{\xi}^*\cdot(\Phi\operatorname{curl}\mathbf{u} - \mathbf{J}) \\ & - (\boldsymbol{\xi}\cdot\nabla\Phi)\mathbf{B}\times\boldsymbol{\xi}^*\cdot\operatorname{curl}\mathbf{u}\Big\}\,d^3\mathbf{x}, \end{aligned} \tag{12}$$

where (12) is correct only if $\Omega(\psi) = \Omega_0$ is constant. (The expression for this special case will suffice for our purposes). The caret over $\hat{\mathbf{B}}_1$ indicates that p_* should be set to zero in the definition of $\mathbf{B}_1$ in (9). We have defined the positive definite operators

$$Q = I + \frac{1}{\gamma p}\mathbf{B}\mathbf{B}, \quad Q^{-1} = I - \frac{1}{\gamma p + \mathbf{B}^2}\mathbf{B}\mathbf{B}. \tag{13}$$

Notice that in (12) it was more convenient to use $\mathbf{u}$ rather than the velocity in the rotating frame $\hat{\mathbf{u}}$.

Finally, we recall[9] that, for a solution of (5) behaving like $\exp(-i\omega t)$ in time, the equation becomes quadratic in ω. Taking the inner product with $\boldsymbol{\xi}$ and solving for ω, one gets that $(F\boldsymbol{\xi},\boldsymbol{\xi})\geqslant 0$ guarantees a real ω, or a stable mode. Using the positivity of G in (11), we have from (10) a sufficient condition for stability[8] $(\hat{F}\xi,\xi)\geqslant 0$ for all admissible ξ.

III. THE BALLOONING MODE EQUATION

As described in the Introduction, ballooning modes are limit points of eigenvalues (at least in the axisymmetric case), but might not be eigenvalues themselves. Fortunately, it is not necessary to follow the accumulating eigenvalues in order to find them. For this purpose we use the device of constructing "singular sequences" due to Hermann Weyl. The technique which was used previously for systems with closed magnetic lines[6,12] is based on the following observation:

Let $(L - i\omega)\xi = 0$ be an eigenvalue equation for an operator L in some Banach space with norm $\| \ \|$. The spectrum of L is the set of $i\omega$ such that $(L - i\omega)^{-1}$ does not exist as a bounded operator. If we can find a normalized sequence $\{\xi_n\}$, $\|\xi_n\| = 1$, and $(L - i\omega)\,\xi_n \to 0$ as $n\to\infty$, then $i\omega$ must be in the spectrum of L. To prove this, assume $i\omega$ is not in the spectrum and let $\mathbf{f}_n = (L - i\omega)\xi_n$, then $\mathbf{f}_n \to 0$ by assumption. Thus $\xi_n = (L - i\omega)^{-1}\mathbf{f}_n \to 0$ since $(L - i\omega)^{-1}$ is bounded, contradicting the assumption that $\|\xi_n\| = 1$. If in addition (in a Hilbert space) ξ_n can be chosen so that the projection of ξ_n on any fixed vector $\boldsymbol{\eta}$ tends to zero as $n\to\infty$ (one then says ξ_n tends weakly to zero), Weyl's criterion ensures us that $i\omega$ is in the so-called essential spectrum which means it is not an isolated eigenvalue of finite multiplicity.[5]

In the MHD case, the sequence ξ_n will be a sequence of functions localized about some ray of one of the MHD waves. We defer the general description of the rays to Sec. V and deal here with the most unstable special case, closest to the static case, where all the Alfven and slow wave rays coincide with magnetic field lines. To clarify our treatment, we return to Eq. (1), write $\nabla p - \mathbf{J}\times\mathbf{B}$ as $\nabla(p + \mathbf{B}^2/2) - \mathbf{B}\cdot\nabla\mathbf{B}$ and notice[13] that all spatial derivatives appear as $\mathbf{B}\cdot\nabla$ and $\mathbf{u}\cdot\nabla$, except terms involving div $\mathbf{u}$ and $\nabla(p + \mathbf{B}^2/2)$. This also holds for the linearized system. Moreover, if $\Omega(\psi)$ in (2) is actually a *constant*, we can pick a rotating frame with $\Omega_0 = \Omega(\psi)$ so that the $\hat{\mathbf{u}}\cdot\nabla$ derivatives will also be proportional to $\mathbf{B}\cdot\nabla$. In this case, every flux surface generated by magnetic field lines is a six times characteristic surface of the system: it is possible to find six equations containing only $\mathbf{B}\cdot\nabla$ derivatives, and only two equations will involve derivatives across the surface. Symbolically, the linearized system can be written as

$$Au + Bv = 0,$$
$$\frac{\partial}{\partial\chi}v + Cu + Dv = 0, \tag{14}$$

where χ is a label for the family of flux surfaces under consideration and A, B, C, D are differential operators, depending linearly on ω. The 2-vector v contains p_* and the component of $\mathbf{u}_1$ normal to the χ surface—these are the two quantities to be differentiated across the surface—while the 6-vector u contains all the other variables. The reader may find it helpful to consult Ref. 13 for an explicit form of Eq. (14). The operator A is special in that all derivatives in it appear as $\mathbf{B}\cdot\nabla$.

To see more clearly what the χ surfaces are, we recall that it is possible to introduce a poloidal angle ϕ, increasing by 2π as we move on a toroidal ψ-surface the short way around the torus, such that $\mathbf{B} = \nabla\psi\times\nabla f$, $f \equiv \theta - q\phi$, where $q(\psi)$ is the "safety factor" [$q(\psi_0)$ being an irrational number implies that the surface $\psi = \psi_0$ is covered by any given field line on it ergodically, while a rational q means that all field lines on that surface close on themselves]. Any function $\chi(\psi, \theta - q\phi)$ has the property $\mathbf{B}\cdot\nabla\chi = 0$, thus $\chi =$ const. generates flux surfaces, which, however, may self-intersect for large θ and ϕ.

We now construct a test function in a tube of field lines about a particular ergodic line defined by $\psi = \psi_0$, $\chi = \chi_0$. The tube will self-intersect. However, if its initial width is very small, its length before self-intersection will be very large, and will tend to infinity as the initial width tends to zero. More precise details on such a construction for the static case can be found in Ref. 12. A more detailed description will also be given for the case considered in Sec. V. The test function has the form

$$u_\epsilon = f_1\Big(\frac{\psi - \psi_0}{\epsilon^n}\Big) f_2\Big(\frac{\chi - \chi_0}{\epsilon}\Big) f_3\Big(\frac{l}{L(\epsilon)}\Big) u_0(l),$$
$$v_\epsilon = \epsilon^k g_1\Big(\frac{\psi - \psi_0}{\epsilon^n}\Big) g_2\Big(\frac{\chi - \chi_0}{\epsilon}\Big) g_3\Big(\frac{l}{L(\epsilon)}\Big) v_0(l) + O(\epsilon), \tag{15}$$

where $k = 1 - n$ and $0 < n < 1/2$, ϵ is a continuous parameter that tends to zero (we will take $\epsilon = 1/m$), l is a coordinate along the field line, and $L(\epsilon)$ is a length of the tube before its self-intersection. All the functions f_i, g_i in (15) are chosen to vanish when their argument lies outside the interval $(-1, 1)$, so that the test function is strongly localized in the χ direction and more weakly localized in the ψ direction. In substituting (15) into (14) one has to expand the coefficients about ψ_0, χ_0 (assuming they are sufficiently smooth), so that, for example,

$$A = A_0\Big(\psi_0,\chi_0,l,\frac{\partial}{\partial l}\Big) + \epsilon^n\Big(\frac{\psi - \psi_0}{\epsilon^n}\Big)A_1 + \epsilon\Big(\frac{\chi - \chi_0}{\epsilon}\Big)A_2 + \text{higher order}.$$

It is possible[12] to choose $v_0(l)$ in terms of $u_0(l)$, relations between f_i and g_i and order ϵ terms, such that, to leading order, Eq. (14) reads

$$A_0\Big(\psi_0,\chi_0,l,\frac{\partial}{\partial l}\Big)u_0 = 0. \tag{16}$$

If (16) is satisfied, the remaining terms are generally of order ϵ^α, with some $\alpha > 0$. Thus, (15) constitutes a singular sequence as $\epsilon \to 0$.

Some remarks should be made about the boundary conditions of Eq. (16) which will presently be written out more explicitly. The leftover $O(\epsilon^\alpha)$ terms depend on u_0 and the derivative $\partial(f_3u_0)/\partial l$, or $\mathbf{B}\cdot\nabla(f_3u_0)$, and in order to preserve the ordering, it must remain bounded in norm as $\epsilon\to 0$. This means that $\|\mathbf{B}\cdot\nabla(f_3u_0)\|_L/\|u_0\|_L$ remains bounded, where the subscript L [or $L(\epsilon)$] indicates that u_0 has to be set to zero for $|l| \geqslant L(\epsilon)$. (We prefer not to specify the norm used at this point.) This boundedness requirement is the "boundary condition" for solution of (16) along the ergodic field line of infinite length. Notice that if u_0 decreases exponentially as

$|l|\to\infty$, the boundary condition will be satisfied.

The previous treatment was correct for $\Omega(\psi)$ const. If Ω is not constant, we can still move to a rotating frame with $\Omega_0=\Omega(\psi_0)$, so that only the flow on the surface $\psi=\psi_0$ will be seen to be parallel to $\mathbf{B}$. In that case the operator A also contains $(\Omega-\Omega_0)\hat{\boldsymbol{\theta}}\cdot\nabla$ derivative, or $(\Omega-\Omega_0)\,\partial/\partial\theta$. If $\Omega-\Omega_0\sim(\psi-\psi_0)^a$ as $\psi\to\psi_0$, this term when applied to the test functions (15) will be of order ϵ^{na-1}. It will be asymptotically small if $a>1/n$. Since n can be any number smaller than 1/2, we find that $a>2$ leaves the previous result unchanged. To conclude, $d\Omega/d\psi=0$ at ψ_0 is not sufficient for the ballooning mode to exist for this surface, but a sufficiently flat Ω will still give rise to a ballooning mode at ψ_0. The argument above did not necessarily provide the weakest requirement on the behavior of $\Omega(\psi)$ near ψ_0 for the ballooning mode to exist, but we do not pursue this question here.

The explicit form of (16) can be obtained, according to (14), by ignoring the two equations containing the χ derivatives and then setting the two special variables in v to zero in the other equations. In particular, the component of the momentum equations along $\nabla\chi$ is ignored, and p_* as well as $\mathbf{u}_1\cdot\nabla\chi$ are set to zero. In terms of the Lagrangian variable $\boldsymbol{\xi}$, if $\boldsymbol{\xi}\cdot\nabla\chi=0$, Eq. (4) guarantees that also $\mathbf{u}_1\cdot\nabla\chi=0$. We then ignore the normal component of (5) and set $p_*\to 0$. This amounts to setting G in (10) to zero, replacing F by $\hat{F}$. The ballooning equation is then

$$P\{-\rho\omega^2-2i\omega A+\hat{F}\}P\boldsymbol{\xi}=0, \tag{17}$$

where P is a projection operator annihilating the component along $\nabla\chi$. Equation (17) is in fact one-dimensional, an ordinary differential equation, involving derivatives along a particular field line. Instead of a volume integral, one may use the relation $d^3\mathbf{x}=d\psi\,d\chi\,dl/B$ $(B=|\mathbf{B}|)$, and change the measure of integration to dl/B. Namely, the one-dimensional inner product

$$\langle\boldsymbol{\xi},\boldsymbol{\eta}\rangle=\int\boldsymbol{\xi}\cdot\boldsymbol{\eta}^*\frac{dl}{B} \tag{18}$$

can be used, and PAP and $P\hat{F}P$ have the same symmetry properties as A and $\hat{F}$.

As described in Sec. II, a sufficient condition for ballooning stability is $\langle P\hat{F}P\boldsymbol{\xi},\boldsymbol{\xi}\rangle\geqslant 0$ for all admissible $\boldsymbol{\xi}$. Further simplification of this criterion is achieved by expressing $\boldsymbol{\xi}$ as

$$\boldsymbol{\xi}=X\mathbf{N}+Z\mathbf{B}/\rho, \tag{19}$$

where $\mathbf{N}$ is the normal to $\mathbf{B}$ within the characteristic flux surface with normal $\nabla\chi$. We normalize $\mathbf{N}$ such that $\mathbf{N}\cdot\nabla\psi=1$. The normalization is possible unless $\mathbf{N}\cdot\nabla\psi=0$, or χ is a ψ-surface, a simple case leading to the so-called Alfven continuous spectrum which is generally stable and which was treated in Ref. 8. The sufficient condition for ballooning stability is the positivity of expression (12), which is correct if $(d/d\psi)\Omega(\psi_0)=0$, evaluated for $\boldsymbol{\xi}$ in (19).

Before writing down this expression explicitly, it is helpful to note that in (9)

$$\hat{\mathbf{B}}_1=Q^{-1}[\mathbf{b}+(\boldsymbol{\xi}\cdot\nabla S)\mathbf{B}/\gamma S]$$

and that $\mathbf{b}(\mathbf{N})$ is parallel to $\mathbf{B}$,

$$\mathbf{b}(\mathbf{N})=\alpha\mathbf{B},\quad \alpha\equiv-2\boldsymbol{\kappa}\cdot\mathbf{N}+(1/B^2)\mathbf{N}\cdot\mathbf{J}\times\mathbf{B}+(1/\rho)\mathbf{N}\cdot\nabla p, \tag{20}$$

where $\boldsymbol{\kappa}$ is the curvature vector of the magnetic field, $\boldsymbol{\kappa}=\hat{\mathbf{e}}\cdot\nabla\hat{\mathbf{e}}$ with $\hat{\mathbf{e}}=\mathbf{B}/|\mathbf{B}|$. Equation (20) is proved using the identity $\mathbf{b}(\mathbf{N})=\operatorname{curl}(\mathbf{N}\times\mathbf{B})+\operatorname{div}(\rho\mathbf{N})\mathbf{B}/\rho$. For every magnetic flux function f, $\mathbf{B}\cdot\nabla f=0$, we have $\nabla f\cdot\mathbf{b}(\mathbf{N})=\operatorname{div}[\nabla f\times(\mathbf{B}\times\mathbf{N})]=\operatorname{div}[(\mathbf{N}\cdot\nabla f)\mathbf{B}]$. Taking the two independent fluxes $f=\chi$ (with $\mathbf{N}\cdot\nabla\chi=0$) and $f=\psi$ (with $\mathbf{N}\cdot\nabla\psi=1$), this expression vanishes. Thus, $\mathbf{b}(\mathbf{N})$ is parallel to $\mathbf{B}$.

Denote $\mathbf{B}\cdot\nabla$ by a prime so that, say, $X'=\mathbf{B}\cdot\nabla X$. Expression (12) can be written as[8]

$$\langle\hat{F}P\boldsymbol{\xi},P\boldsymbol{\xi}\rangle=Q_{zz}+2Q_{xz}+Q_{xx}, \tag{21}$$

where for the special case of $\Phi(\psi)\equiv 0$ (purely toroidal flow),

$$Q_{zz}=\int\frac{dl}{B}\rho^{-2}\Big[\beta\mathbf{B}^2\Big(Z'+\frac{S'Z}{\gamma S}\Big)^2-\frac{Z^2S'p'}{\gamma S}\Big],$$

$$Q_{xz}=\int\frac{dl}{B}\rho^{-1}\beta X\mathbf{N}\cdot\Big[ZS'\frac{(\rho\mathbf{u}\cdot\nabla\mathbf{u}-2\mathbf{B}\cdot\nabla\mathbf{B})}{\gamma S}-Z'\Big(\frac{\mathbf{B}^2\rho\mathbf{u}\cdot\nabla\mathbf{u}}{\gamma p}+2\mathbf{B}\cdot\nabla\mathbf{B}\Big)\Big],$$

$$Q_{xx}=\int\frac{dl}{B}\Big(|\mathbf{N}|^2X'^2+X^2\Big\{\beta\mathbf{B}^2\Big[\alpha+\frac{\mathbf{N}\cdot\nabla S}{\gamma S}\Big]^2-\alpha\mathbf{J}\times\mathbf{B}\cdot\mathbf{N}-\frac{(\mathbf{N}\cdot\nabla S)(\mathbf{N}\cdot\nabla p)}{\gamma S}\Big\}\Big).$$

These expressions are essentially the same as those given in Ref. 8 for the simpler configuration of field-reversed mirrors. We write them symbolically as

$$\langle\hat{F}P\boldsymbol{\xi},P\boldsymbol{\xi}\rangle=\int\frac{dl}{B}[a_1Z'^2+2a_2ZZ'+a_3Z^2+2b_1XZ'+2b_2XZ+c_1X'^2+c_2X^2]. \tag{22}$$

Notice that $a_1>0$ and the quadratic expression Q_{zz} in Z and Z' is positive definite if $p'S'\leqslant 0$, which we assume. This condition is satisfied for the most common equilibria,[8] where a relation

$$p=p(\psi,\rho) \tag{23}$$

holds and $S\equiv p\rho^{-\gamma}$. We consider two cases.

Case I: $S=S(\psi)$

When $S=S(\psi)$, $S'=0$, and $a_2=a_3=b_2=0$. It is then possible to rewrite (22) after completing the square as

$$\langle\hat{F}\boldsymbol{\xi},\boldsymbol{\xi}\rangle=\int\frac{dl}{B}\Big[a_1\Big(Z'+\frac{b_1}{a_1}X\Big)^2+c_1X'^2+\Big(c_2-\frac{b_1^2}{a_1}\Big)X^2\Big], \tag{24}$$

which one wishes to minimize say over all admissible functions $X(l)$, $Z(l)$ on $(-\infty,\infty)$, where for admissibility we need to define a norm, like requiring that the function as well as its derivatives be square integrable on $L^2(-\infty,\infty)$.

The minimum of (24) with respect to Z can formally be accomplished by taking $Z'=-b_1X/a_1$ (assuming $|\mathbf{B}|$ and ρ bounded away from zero). However, Z may then be an inadmissible function, not vanishing at $\pm\infty$. We can overcome this difficulty, however, by introducing a sequence of admissible function $Z_n(l)$ with the property that $\|Z_n'' + (b_1/a_1)X\|$ can be made arbitrarily small as n tends to infinity. For this we let

$$Z_n(l) = -f\left(\frac{l}{n}\right)\int_0^l \frac{b_1}{a_1} X \frac{dl}{B}, \qquad (25)$$

where $f(x)$ is a smooth function such that

$$f(x) = \begin{cases} 1, & |x| < \frac{1}{2}, \\ 0, & |x| \geqslant 1. \end{cases}$$

It can easily be checked that since $X(l)$ tends to zero at infinity, Z_n yields the desired approximation. Thus, in order for $\langle \hat{F}\xi,\xi\rangle$ to be nonnegative, we must have

$$\int \frac{dl}{B}\left[c_1 X'^2 + \left(c_2 - \frac{b_1^2}{a_1}\right)X^2\right] \geqslant 0. \qquad (26)$$

After a significant amount of algebra, it may be shown that the explicit form of the above expression can be simplified to the following when $\Phi(\psi) = 0$ and $(d/d\psi)\Omega(\psi_0) = 0$:

$$\langle \hat{F}\xi,\xi\rangle = \int\Big\{|\mathbf{N}|^2 X'^2 - X^2\Big[2(\boldsymbol{\kappa}\cdot\mathbf{N})(\mathbf{N}\cdot\mathbf{J}\times\mathbf{B}) + r\Omega_0^2(\mathbf{N}\cdot\nabla r) \times\left(\frac{\partial p}{\partial \rho}\right)^{-1}\left(\frac{\partial p}{\partial \psi} - \mathbf{N}\cdot\mathbf{J}\times\mathbf{B}\right)\Big]\Big\}\frac{dl}{B}, \qquad (27)$$

where $p = p(\psi,\rho)$ as in (23).

The positiveness is the sufficient condition for ballooning stability. The dependence of this expression on the flow will be analyzed via a high aspect ratio expansion in Sec. IV.

Case II. $S \neq S(\psi)$

In this case additional terms are present in (24). It is still possible to simplify the stability criterion if we minimize with respect to Z and Z' now thought of as independent functions, as in Ref. 8. This, of course, yields an inherently somewhat more pessimistic sufficient condition for stability than the previous case, but it is possible to show that in this case one again recovers (27) so that (27) is once again sufficient for stability.

IV. LARGE ASPECT RATIO EXPANSION

In this section we try to gain some physical understanding of the effect of the flow on ballooning stability. To simplify matters, we use a common analytical device for toroidal systems, a large aspect ratio asymptotic expansion. The aspect ratio is defined as the ratio of the major radius of the torus, R, to a typical poloidal plasma radius a. We assume $\epsilon \equiv a/R \ll 1$. The usual "high beta" tokamak ordering is $B_\theta = O(1) = \rho$, $\mathbf{B}_p = O(\epsilon) = p$, where $\mathbf{B}_p$ is the poloidal projection of $\mathbf{B}$, $\mathbf{B}_p = \mathbf{B} - B_\theta\hat{\boldsymbol{\theta}}$. We take $a = O(1)$ so that $r = O(\epsilon^{-1})$ in the plasma while $\partial/\partial r$ and $\partial/\partial z$ are $O(1)$.

To order the flow velocity, we recall that ellipticity of the equilibrium equation is guaranteed if $\Phi^2/\rho < \beta$. Hence we take $\Phi = O(\epsilon^{1/2})$. Also, the fastest toroidal flow is observed experimentally to be at most at Mach one speed (in the *ISX-B* experiment[14]). This means $\rho r^2\Omega^2/(\gamma p) \leqslant 1$. We then take $\Omega = O(\epsilon^{3/2})$. The equilibrium momentum equation is affected by the flow through the term $\rho\mathbf{u}\cdot\nabla\mathbf{u}$, which, however, is of order ϵ^2, while the pressure term ∇p is $O(\epsilon)$. Thus to order ϵ the equilibrium state is *unchanged* by the presence of flow. In particular, to leading order $p = \epsilon p_1(\psi)$.

Returning to the ballooning equation (17), in order to have an instability we will see shortly that $\mathbf{B}\cdot\nabla\xi$ is $O(\epsilon)$ and also $\omega = O(\epsilon)$. The middle term in (17) is $-2i\omega[\Phi\,\mathbf{B}\cdot\nabla\xi - \rho\Omega_0\xi\times\hat{\mathbf{z}}]$ and is of order $\epsilon^{5/2}$ as compared to the $\rho\omega^2$ term which is $O(\epsilon^2)$. Thus it is negligible. To leading order for such modes, Eq. (17) reduces to

$$P\hat{F}P\,\xi = \rho\omega^2\xi, \qquad (28)$$

which is a self-adjoint eigenvalue equation. Thus, for ω of order ϵ the ballooning modes are stable if and only if $\langle P\hat{F}P\,\xi,\xi\rangle \geqslant 0$ for all ξ. Notice, however, that near marginal stability ω^2 may get so small as to be of the same size as the neglected middle term of (17) and a modification of the expansion will be required.

For simplicity we treat the purely toroidal flow case $\Phi \equiv 0$, and also assume $S = S(\psi)$. Equation (27) is the applicable criterion for stability. The coefficient of X^2, the only possible cause for instability, is of order ϵ^2. To see this note that $|\boldsymbol{\kappa}| \sim 1/r = O(\epsilon)$. Writing $\partial p/\partial\psi - \mathbf{N}\cdot\mathbf{J}\times\mathbf{B} = \mathbf{N}\cdot(\nabla p - \mathbf{J}\times\mathbf{B}) - (\partial p/\partial\rho)\mathbf{N}\cdot\nabla\rho$ so only $(\partial p/\partial\rho)\mathbf{N}\cdot\nabla\rho$ needs to be kept to leading order, we have to $O(\epsilon^2)$

$$\langle \hat{F}\xi,\xi\rangle = \int\{X'^2|\mathbf{N}|^2 - X^2[2(\boldsymbol{\kappa}\cdot\mathbf{N})(\mathbf{N}\cdot\nabla p) - (\mathbf{N}\cdot\nabla\rho)\mathbf{N}\cdot\nabla(\tfrac{1}{2}r^2\Omega^2)]\}\frac{dl}{B}. \qquad (29)$$

From here it is immediately seen that for unstable modes, with $\langle\hat{F}\xi,\xi\rangle < 0$, $X' = O(\epsilon)$ at most, for otherwise it would have dominated the negative terms. Likewise, from the previous section we get for the worst case $Z' = O(\epsilon)$, or $\mathbf{B}\cdot\nabla\xi = O(\epsilon)$ as claimed before. Notice also that $|\omega|^2 = -\langle F\xi,\xi\rangle/\langle\rho\xi,\xi\rangle$, so $\omega = O(\epsilon)$ as assumed.

The last term in (29) expresses the effect of the flow on the ballooning stability. It is seen to be simply the Rayleigh–Taylor effect of destabilization when a heavy fluid is on top of a light fluid in a gravity field. Here the gravity is replaced by the centrifugal force $\nabla(r^2\Omega^2/2)$, always pointing out radially in the $\hat{\mathbf{r}}$ direction. Since to leading order $p = p(\psi)$, ρ is also a function of ψ. For a confined plasma $\rho(\psi)$ decreases towards the boundary; thus $\nabla\rho$ points into the plasma. The flow term thus tends to destabilize on the outer side of the torus, and, to stabilize in the inner side, similar to the effect of the curvature $\boldsymbol{\kappa}$. In fact, if S is a global constant, then defining $\hat{\boldsymbol{\kappa}} = \boldsymbol{\kappa} - \rho\nabla(r^2\Omega^2)/(4\gamma p)$, expression (29) with $\hat{\boldsymbol{\kappa}}$ appears as if there was no flow. The result indicates further *destabilization* of the plasma in the presence of rigidly rotating flow. Note, however, that if $\Omega \neq$ const. the ballooning modes may disappear all together so the effect of the flow on stability is not clear cut. We would like to stress that the effect of the flow appears not as a modification of the pressure due to the centrifugal force—this is a higher order effect—but as a modification of the curvature.

We did not discuss the case of $\Phi \neq 0$. One expects, however, to see in (29) additional terms expressing the effect of flow shear—the Kelvin–Helmholtz instability.

V. LOCALIZATION FOR THE ALFVEN RAY

As mentioned in the Introduction, the essential spectrum of the Lundquist equations when linearized around static equilibria in toroidal systems is intimately connected with the propagation of waves along rays of the equilibrium's characteristic surfaces. Here we wish to make these statements more precise and then show that a stable part of the essential spectrum, when there is flow in the equilibrium, still can be characterized in this way.

As is well known,[15] the Lundquist equations (1) constitute a symmetric hyperbolic system. This means that they can be expressed in the form

$$LU = U_t + \sum_i A_i \frac{\partial}{\partial x_i} U + BU = 0,$$

where the A_i are 8×8 symmetric, real matrices, and U is an 8-vector. See Ref. 15 for a precise description of the equations. Thus the characteristic determinant[15] (we will use $\|\ \|$ to denote determinant) is

$$\left|\left|\phi_t I - \sum_i A_i \phi_{x_i}\right|\right| = 0, \tag{30}$$

where $\phi=$ const. is a characteristic surface and subscripts of ϕ denote its derivatives. I is the identity matrix and the equation has eight real but not necessarily distinct roots ϕ_t, for any choice of ϕ_{x_i}.

If, however, we consider only the spatial part of L, the characteristic determinant

$$\left|\left|\sum_i A_i \phi_{x_i}\right|\right| \tag{31}$$

is a polynomial of degree 8 in $\xi_i = \phi_{x_i}$, $i=1,2,3$ in $\Omega \subset R^3$, of the form

$$Q(\xi_1,\xi_2,\xi_3) = 0.$$

The latter may not have only real roots. If such real roots remain, however, they correspond to a hyperbolic part of the spatial part of L and to characteristic surfaces which are generated by rays.

Specifically in the case of the Lundquist equations, it can be shown that[15]

$$\left|\left|\sum_i A_i \phi_{x_i}\right|\right| = (\mathbf{u}\cdot\nabla\phi)^2((\mathbf{u}\cdot\nabla\phi)^2 - (\mathbf{A}\cdot\nabla\phi)^2)((\mathbf{u}\cdot\nabla\phi)^2 - c_s^2|\nabla\phi|^2)((\mathbf{u}\cdot\nabla\phi)^2 - c_f^2|\nabla\phi|^2) = 0, \tag{32}$$

with

$$\mathbf{A} = \frac{\mathbf{B}}{\sqrt{\rho}},\quad a^2 = \frac{\partial p(\rho,S)}{\partial\rho},$$

$$c_{s,f} = \{a^2 + A^2 \mp [(a^2+A^2)^2 - a^2(\mathbf{A}\cdot\nabla\phi)^2]^{1/2}\}^{1/2},$$

where c_s (resp. c_f) corresponds to taking the minus sign (resp., the plus sign) in the above expression.

The families of characteristic surfaces associated with this determinant are given by

$$\mathbf{u}\cdot\nabla\phi = 0,\quad \mathbf{u}\cdot\nabla\phi \pm \mathbf{A}\cdot\nabla\phi = 0,\quad \mathbf{u}\cdot\nabla\phi \pm c_s|\nabla\phi| = 0,$$
$$\mathbf{u}\cdot\nabla\phi \pm c_f|\nabla\phi| = 0.$$

Denoting a particular family by $Q_j(\mathbf{x},\nabla\phi)=0$, we recall that the surfaces $\phi=$ const. are generated by the rays solving the bicharacteristic equations

$$\frac{dx_i}{dt} = \frac{\partial}{\partial\phi_{x_i}} Q_j(\mathbf{x},\nabla\phi),$$
$$\frac{d\phi_{x_i}}{dt} = -\frac{\partial}{\partial x_i} Q_j(\mathbf{x},\nabla\phi), \tag{33}$$

where t is a parameter along the ray.

In the following we will consider in particular the Alfven family generated by

$$\mathbf{u}\cdot\nabla\phi \pm \mathbf{A}\cdot\nabla\phi = 0. \tag{34}$$

Moreover, without loss of generality, we will restrict ourselves to the family

$$\mathbf{u}\cdot\nabla\phi + \mathbf{A}\cdot\nabla\phi = 0. \tag{35}$$

We will see that an ordinary first-order differential equation along this ray determines points in the essential spectrum of L.

We now digress to introduce the relevant concepts about the equilibrium configuration which we will make use of subsequently.

It can be shown, as in the static case, that there exist "irrational" surfaces such that the trajectories of the field $\mathbf{u}+\mathbf{B}/\sqrt{\rho}$ on these surfaces never close and are in fact ergodic on these surfaces. We will indicate the proof of this fact later. We will need to construct a coordinate ϕ such that

$$(\mathbf{u}+\mathbf{B}/\sqrt{\rho})\cdot\nabla\phi = 0, \tag{36}$$

in a neighborhood of the ergodic field line. Actually we define ϕ in a sequence of tubular neighborhoods T_n of the field line. Where T_n has the properties that the radius of T_n, $r(T_n)$, tends to zero as n tends towards infinity and the length of T_n, $L(T_n)$, tends to infinity as n tends towards infinity. Also the tube T_n does not self-intersect. Thus T_n is collapsing as it is growing onto the particular field line. That such a sequence of tubes T_n indeed exists can be shown by contradiction if one uses the continuous dependence of ordinary differential equations on their initial data, making use of the continuous differentiability of $\mathbf{u}+\mathbf{B}/\sqrt{\rho}$. We note that if one uses Hamada coordinates[16] (ψ,θ,α) to represent the magnetic field in the form

$$\mathbf{B} = \nabla\psi\times\nabla[\theta - q(\psi)\alpha], \tag{37}$$

where

$$\nabla\psi\times\nabla\theta\cdot\nabla\alpha = 1$$

and makes use of (2), one obtains the most general local solution of (36) in the form

$$\phi\left(\psi, -q(\psi)\alpha - \int_{\alpha_0}^{\alpha}\Omega(\psi)\left[\frac{\Phi(\psi)}{\rho} + \frac{1}{\sqrt{\rho}}\right]^{-1} d\tilde{\alpha} + \theta\right), \tag{38}$$

where $\phi(\psi,\chi)$ is determined by the distribution of $\nabla\phi$ on an initial surface containing one point on the ergodic field line.

The vector field

$$(\Phi/\rho \pm 1/\sqrt{\rho})^{-1}(\mathbf{u} \pm \mathbf{B}/\sqrt{\rho}) \tag{39}$$

is divergence-free, so that since we know that

$$(\mathbf{u} \pm \mathbf{B}/\sqrt{\rho})\cdot\nabla\psi = 0, \tag{40}$$

with closed surfaces ψ, we can apply known methods[16] to deduce the existence of a multivalued function χ and a function $\tilde{q}(\psi)$ such that

$$(\Phi(\psi)/\rho \pm 1/\rho)^{-1}(\mathbf{u} \pm \mathbf{B}/\sqrt{\rho}) = \nabla\psi\times\nabla\chi = \nabla\psi\times(\nabla\theta - \tilde{q}(\psi)\nabla\alpha), \tag{41}$$

where if $2\pi\tilde{q}(\psi_0)$ is irrational the trajectory of $\mathbf{u} \pm \mathbf{B}/\sqrt{\rho}$ is ergodic on $\psi = \psi_0$. We assume throughout then that the localization takes place on such an ergodic trajectory. With these preliminaries out of the way, we now return to the spectral problem.

In the following we give an argument to demonstrate that a part of the essential spectrum, namely, the so-called ballooning spectrum of static equilibria in ideal MHD, is also present when there is flow in the equilibrium. We will show this as mentioned above by using singular sequences due to Weyl. For the sake of definiteness, but not because this is the only possible set up, we will work in a Hilbert space $L^2(\Omega)$ and let

$$D(L) = \Big\{U = (u_1,u_2,u_3,B_1,B_2,B_3,S,\rho)\in[L^2(\Omega)]^8, \sum_i A_i \frac{\partial}{\partial x_i} U\in\{L^2(\Omega)\}^8,\ \mathbf{u}\cdot\mathbf{n} = 0 \text{ on } \partial\Omega\Big\}, \tag{42}$$

where L is the operator associated with the spatial part of the Lundquist equations.

It can be shown[17] that $L + \lambda I$ for sufficiently large λ is a closed operator. We now show how to construct the singular sequences mentioned above by localizing around an Alfven surface on a particular ray lying in this surface. Unlike the case discussed in Sec. III this localization will lead only to stable spectrum.

The key fact that we will use in the course of the localization process is that if we go to the coordinate (ϕ,χ,s) [where s is determined by $d\mathbf{x}/ds = \mathbf{u} + \mathbf{B}/\sqrt{\rho}$] the symmetric characteristic matrix $\Sigma_i A_i \phi_{x_i}$, which is the coefficient of the $\partial/\partial\phi$ derivatives in LU, has precisely one left null vector corresponding to the Alfven wave root (34).

This left-right eigenvector can be shown to be

$$\begin{pmatrix}\delta\mathbf{u}\\ \delta\mathbf{B}\\ \delta p\\ \delta S\end{pmatrix} = \begin{pmatrix}-\alpha\,\rho c\mathbf{B}\times\mathbf{n}\\ (\mathbf{B}\cdot\mathbf{n})\mathbf{B}\times\mathbf{n}\\ 0\\ 0\end{pmatrix}. \tag{43}$$

We now factor our system $(L - i\omega)U$ in the following way (denoting left eigenvectors by $l_j, j = 1,\dots,8$):

$$(L - i\omega)U = \begin{pmatrix} l_1(L - i\omega)U\\ \cdot\\ \cdot\\ \cdot\\ l_8(L - i\omega)U\end{pmatrix}. \tag{44}$$

We introduce the notation $\tilde{U} = U - P_{r_1}(U)$, where P_{r_1} is the projector onto the space spanned by r_1.

We now take note of the fact that in the first equation in (44), $l_1(L - \lambda)U$ may be reexpressed in the form

$$l_1(L - \lambda)U = \eta\Big(\frac{\partial}{\partial s}, \phi;\chi,s\Big)P_{r_1}(U) + \gamma\Big(\frac{\partial}{\partial s}, \frac{\partial}{\partial\chi};\phi,\chi,s\Big)\tilde{U}, \tag{45}$$

where we introduce here the notation $F(\partial/\partial s, \partial/\partial\chi, \phi, \chi, s)$ or, ocassionally, just $F(\partial/\partial s, \partial/\partial\chi)$ to indicate that F is a linear operator involving differentiation with respect to the variables s and χ, with coefficients depending on the variables ϕ, χ, s; and η and γ are such operators as well.

That $P_{r_1}(V)$ is preceded by an operator η involving only differentiation with respect to s can be checked by direct calculation (we will give the end result here) and is also related to general theorems arising in the propagation of singularities for hyperbolic equations.[15] A involves no differentiation with respect to ϕ, as l_1 is a left null vector.

Explicitly it is possible to show that (with scalar function σ_1)

$$\begin{aligned} l_1(L - i\omega)(P_{r_1}(U)) &= l_1(L - i\omega)(\sigma_1 r_1)\\ &= (\mathbf{B}\cdot\mathbf{n})(\mathbf{B}\times\mathbf{n})\cdot[\lambda\,\rho\sigma_1(\mathbf{B}\cdot\mathbf{n})\mathbf{B}\times\mathbf{n} + \rho\mathbf{u}\cdot\nabla(\sigma_1(\mathbf{B}\cdot\mathbf{n})\mathbf{B}\times\mathbf{n})\\ &\quad + \mathbf{B}\cdot\nabla(\sigma_1\,\rho c(\mathbf{B}\times\mathbf{n})) + \rho\sigma_1(\mathbf{B}\cdot\mathbf{n})(\mathbf{B}\times\mathbf{n})\cdot\nabla\mathbf{u}\\ &\quad + \sigma_1\,\rho c(\mathbf{B}\times\mathbf{n})\cdot\nabla\mathbf{B}] - \rho c(\mathbf{B}\times\mathbf{n})\\ &\quad \cdot[-\lambda\sigma_1\,\rho c\mathbf{B}\times\mathbf{n} - \mathbf{u}\cdot\nabla(\sigma\,\rho c(\mathbf{B}\times\mathbf{n}))\\ &\quad + \sigma_1(\mathbf{B}\cdot\mathbf{n})(\mathbf{B}\times\mathbf{n})\cdot\nabla\mathbf{B}] - \mathbf{B}\cdot\nabla(\sigma_1(\mathbf{B}\cdot\mathbf{n})(\mathbf{B}\times\mathbf{n}))\\ &\quad + \sigma_1\,\rho c(\mathbf{B}\times\mathbf{n})\cdot\nabla\mathbf{u} - \mathbf{B}\cdot\nabla(\sigma_1(\mathbf{B}\cdot\mathbf{n})(\mathbf{B}\times\mathbf{n}))\\ &\quad - \rho c\sigma_1(\mathbf{B}\times\mathbf{n})\cdot\nabla\mathbf{u} - \mathbf{u}\cdot\nabla(-\rho c\sigma_1(\mathbf{B}\times\mathbf{n})), \end{aligned} \tag{46}$$

where $c = \mathbf{u}\cdot\mathbf{n}$.

Using vector identities, it is possible to show, after some manipulation, that the above expression reduces to

$$\sigma_1' + \sigma_1[\log(\sqrt{\rho}|\mathbf{B}\times\mathbf{n}|^2/(\Phi + \sqrt{\rho}))]' = i\omega\sigma_1, \tag{47}$$

where

$$g' = (\mathbf{u} + \mathbf{B}/\sqrt{\rho})\cdot\nabla g.$$

This can be integrated to yield

$$\sigma_1(s) = e^{i\omega s}(\Phi + \sqrt{\rho})/\sqrt{\rho}|\mathbf{B}\times\mathbf{n}|^2. \tag{48}$$

We will see that if ω is real, ω is in the essential spectrum of L and that this fact in turn is related to the above form of σ_1. We note that there exists a constant $d > 0$ such that

$$|\mathbf{B}\times\mathbf{n}| > d. \tag{49}$$

This is so because

$$(\mathbf{u} + \mathbf{B}/\sqrt{\rho})\cdot\mathbf{n} = 0,$$

$$|(\mathbf{u} + \mathbf{B}/\sqrt{\rho})\times\mathbf{n}| = |\mathbf{u} + \mathbf{B}/\sqrt{\rho}|.$$

Therefore,

$$\begin{aligned} |(\Phi/\sqrt{\rho} + 1)\mathbf{B}\times\mathbf{n}| &= |(\sqrt{\rho}\mathbf{u} + \mathbf{B})\times\mathbf{n} - \sqrt{\rho}r\Omega(\psi)\nabla\theta\times\mathbf{n}|\\ &\geqslant |(\sqrt{\rho}\mathbf{u} + \mathbf{B})\times\mathbf{n}| - |\sqrt{\rho}r\Omega(\psi)\nabla\theta\times\mathbf{n}|\\ &= |\sqrt{\rho}\mathbf{u} + \mathbf{B}| - |\sqrt{\rho}r\Omega(\psi)\nabla\theta\times\mathbf{n}|\\ &\geqslant \{(\Phi/\sqrt{\rho} + 1)^2B_p^2 + [B_\theta(1 + \Phi/\sqrt{\rho})\\ &\quad + \sqrt{\rho}r\Omega(\psi)]^2\}^{1/2} - \sqrt{\rho}|\Omega(\psi)|. \end{aligned}$$

The above expression is then clearly bounded away from zero provided that either the toroidal component of the flow and that of the magnetic field are in the same direction or that the poloidal magnetic field is large compared with $\sqrt{\rho}\Omega(\psi)$.

For the remaining equations,

$$\begin{matrix} l_2(L-\lambda)U \\ \cdot \\ \cdot \\ \cdot \\ l_8(L-\lambda)U \end{matrix} \tag{50}$$

of (44), we will make use of the fact that they may be written as

$$D\frac{\partial \tilde U}{\partial\phi} + B\Big(\frac{\partial}{\partial\chi},\frac{\partial}{\partial s}\Big)\tilde U + C\Big(\frac{\partial}{\partial\chi},\frac{\partial}{\partial s}\Big)P_{r_1}(U), \tag{51}$$

where D is a nonsingular diagonal 7×7 matrix

$$D=\begin{pmatrix}\lambda_2 & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \lambda_8\end{pmatrix},$$

the λ_i are the remaining nonvanishing roots of (34), and $\tilde U = U - P_{r_1}(U)$.

We now expand the coefficients of all the operators around the field line, for instance,

$$\gamma\Big(\frac{\partial}{\partial\chi},\frac{\partial}{\partial s}\Big) = \gamma^s(\phi_0,\chi_0,s)\frac{\partial}{\partial s} + \gamma^\chi(\phi_0,\chi_0,s)\frac{\partial}{\partial\chi} + \tilde\gamma\Big(\frac{\partial}{\partial\chi},\frac{\partial}{\partial s}\Big),$$

where $\tilde\gamma(\partial/\partial\chi,\partial/\partial s)$ is $O(\phi-\phi_0,\chi-\chi_0)$. Such an expression is valid given the assumed smoothness of the equilibrium quantities.

We now consider the action of $L - i\omega$ on the following sequence of test functions (Weyl sequence). Here ϵ_m indicates a discrete parameter that tends towards zero as $m\uparrow\infty$:

$$U_m = \Phi_1\Big(\frac{\phi-\phi_0}{\epsilon_m}\Big)\Omega_1\Big(\frac{\chi-\chi_0}{\epsilon_m^n}\Big)f\Big(\frac{s}{L(m)}\Big)\sigma_1(s)r_1(s) + \epsilon_m^k\Phi_2\Big(\frac{\phi-\phi_0}{\epsilon_m}\Big)\Omega_2\Big(\frac{\chi-\chi_0}{\epsilon_m^n}\Big)f\Big(\frac{s}{L(m)}\Big)\Big(\sum_{i=2}^{i=8}\sigma_i(s)r_i(s)\Big)$$
$$+\epsilon_m\Big[\tilde\Phi_2\Big(\frac{\phi-\phi_0}{\epsilon_m}\Big)\tilde\Omega_2\Big(\frac{\chi-\chi_0}{\epsilon_m^n}\Big)f\Big(\frac{s}{L(m)}\Big)\Big(\sum_{i=2}^{i=8}\sigma_i(s)r_i(s)\Big) + \chi\Big(\frac{\phi-\phi_0}{\epsilon_m}\Big)\beta\Big(\frac{\phi-\phi_0}{\epsilon_m^n}\Big)f\Big(\frac{s}{L(m)}\Big)\sum_{i=2}^{i=8}\tilde\sigma_i(s)r_i(s)\Big].$$

Here the Φ_i, Ω_i, $\tilde\Phi_i$, $\tilde\Omega_i$, χ, β are smooth functions whose support is contained in $[-1,1]$. Also $\sigma_1(s)$ is chosen to satisfy (48). We will let

$$\Phi_1\Big(\frac{\phi-\phi_0}{\epsilon_m}\Big)\Omega_1\Big(\frac{\chi-\chi_0}{\epsilon_m^n}\Big)f\Big(\frac{s}{L(m)}\Big)\sigma_1(s)r_1(s) = \hat u_m \tag{52}$$

and

$$\Phi_2\Big(\frac{\phi-\phi_0}{\epsilon_m}\Big)\Omega_2\Big(\frac{\chi-\chi_0}{\epsilon_m^n}\Big)\sum_{i=2}^{i=8}\sigma_i(s)r_i(s) = \hat{\hat U}_m.$$

Our object for the remainder will be to show that

$$\|(L-i\omega)U_m\| \leqslant N(m)\|P_{r_1}(U_m)\| \leqslant N(m)\|U_m\|, \tag{53}$$

where $N(m)\downarrow 0$ as $m\uparrow\infty$.

We now will also use the notation

$$\sigma_1(s)r_1(s) = \hat u(s),$$
$$\hat U(s) = \sum_{i=2}^{i=8}\sigma_i(s)r_i(s), \quad \hat V(s) = \sum_{i=2}^{i=8}\tilde\sigma_i(s)r_i(s).$$

We recall as indicated above that the first equation may be written, in view of (45),

$$l_1(L-i\omega)U_m = \frac{1}{L(m)}f'\Big(\frac{s}{L(m)}\Big)\hat u_m + A_0^\chi(s,\psi_0,\chi_0)\frac{\partial}{\partial\chi}\hat{\hat U}_m + A_0^s(s,\psi_0,\chi_0)\frac{\partial}{\partial s}\hat{\hat U}_m + O(\psi-\psi_0,\chi-\chi_0)$$
$$= \epsilon_m^{k-m}\Phi_2\Big(\frac{\phi-\phi_0}{\epsilon_m}\Big)\Omega_2'\Big(\frac{\chi-\chi_0}{\epsilon_m^n}\Big)f\Big(\frac{s}{L(\epsilon_m)}\Big)\hat U_m(s)$$
$$+\hat A_0^\chi\epsilon^{1-m}\Big[\tilde\Phi_2\Big(\frac{\phi-\phi_0}{\epsilon_m}\Big)\tilde\Omega_2'\Big(\frac{\chi-\chi_0}{\epsilon_m^n}\Big)f\Big(\frac{s}{L(m)}\Big)\hat U_m(s) + \chi\Big(\frac{\phi-\phi_0}{\epsilon_m}\Big)\beta'\Big(\frac{\chi-\chi_0}{\epsilon_m^n}\Big)f\Big(\frac{s}{L(m)}\Big)\hat V_m(s)\Big]$$
$$+\hat A_0^s\epsilon_m^k\Phi_2\Big(\frac{\phi-\phi_0}{\epsilon_m}\Big)\Omega_2\Big(\frac{\chi-\chi_0}{\epsilon_m^n}\Big)f\Big(\frac{s}{L(m)}\Big)\hat U_m'(s) + \frac{f'}{L(m)}V_m + O(\epsilon^\alpha). \tag{54}$$

Here

$$\hat A_0^\chi = l_1\cdot A_0^\chi, \quad \hat A_0^s = l_1\cdot A_0^s, \quad \text{and} \quad \alpha>0.$$

Similarly, the remaining equations in (50) can be summarized as

$$\epsilon_m^{k-1} D_0 \phi_2'\left(\frac{\phi-\phi_0}{\epsilon_m}\right)\Omega_2\left(\frac{\chi-\chi_0}{\epsilon_m^n}\right)$$
$$\times f\left(\frac{s}{L(m)}\right)\hat{U}(s) + D_0 \tilde{\Phi}_2'\left(\frac{\phi-\phi_0}{\epsilon_m}\right)$$
$$\times \tilde{\Omega}_2\left(\frac{\chi-\chi_0}{\epsilon_m^n}\right) f\left(\frac{s}{L(m)}\right)\hat{V}(s)$$
$$+ D_0 \chi'\left(\frac{\phi-\phi_0}{\epsilon_m}\right)\beta\left(\frac{\phi-\phi_0}{\epsilon_m^n}\right) f\left(\frac{s}{L(m)}\right)\hat{U}(s)$$
$$+ C_0^s\left[\Phi_1\left(\frac{\phi-\phi_0}{\epsilon_m}\right)\Omega_1\left(\frac{\chi-\chi_0}{\epsilon_m^n}\right) f\left(\frac{s}{L(m)}\right)\hat{u}'(s)\right]$$
$$+ \epsilon_m^{-n} C_0^\chi\left[\Phi_1\left(\frac{\phi-\phi_0}{\epsilon_m}\right)\Omega_1'\left(\frac{\chi-\chi_0}{\epsilon_m^n}\right)\right.$$
$$\left.\times f\left(\frac{s}{L(m)}\right)\hat{u}(s)\right] + O(\epsilon^r), \tag{55}$$

where $r > 0$.

Also n and k must be chosen so $n + k = 1$. To eliminate all quantities which do not tend to zero as ϵ_m tends to zero, we chose for given $\hat{u}(s)$ [determined from (48)]

$$D_0 \hat{U}(s) = C_0^\chi \hat{u}(s),$$
$$D_0 \hat{V}(s) = D_0 \hat{U}(s) - C_0^s \hat{U}'(s), \tag{56}$$

where I is the 8×8 identity matrix, and

$$\Phi_2' = \Phi_1 = \chi', \quad \tilde{\Omega}_2' = \beta = \Omega_1,$$
$$\Omega_1' = \Omega_2.$$

This being done it remains to show that the order of magnitude of $\|\hat{U}\|$ is bounded by that of $\|u\|$, where $\|\ \|$ designates the L_2 norm in Ω.

With

$$J = |\nabla\phi \times \nabla\chi \cdot \nabla s|,$$

we have

$$\|\hat{U}_m\|^2 = \int_\Omega |\hat{U}_m|^2\, dV = \int_\Omega |\hat{U}_m|^2 |J|\, ds\, d\chi\, d\phi$$
$$= \epsilon^{1+n}\int_Q d\phi\, d\chi \int_{-L(m)}^{L(m)} |\hat{U}_m|^2 |J|\, ds,$$

where

$$Q = [-1, 1] \times [-1, 1].$$

The only dangerous contribution is the one due to σ_1' in $\hat{U}$, given by

$$\epsilon^{1+n}\int d\phi\, d\chi \int_{-L(m)}^{L(m)} |\sigma_1'(s)|\, |J|\, ds.$$

Now from (48) $\sigma_1(s) = \exp(i\omega s)\,(\Phi + \sqrt{\rho})/\sqrt{\rho}\,|\mathbf{B}\times\mathbf{n}|^2$, so the only nontrivial part to estimate is the contribution in $|\mathbf{B}\times d\mathbf{n}/ds|$.

However, using the bicharacteristic equation, mentioned above (41),

$$\frac{d\phi_{x_i}}{ds} = -\frac{\partial}{\partial x_i}\left[(\mathbf{u} + \mathbf{B}/\sqrt{\rho})\cdot\nabla\phi\right] \tag{57}$$

for the normal, and, splitting off the dependence on the normal and dividing by $|\nabla\phi|$, we obtain

$$|\nabla\phi|\frac{d\mathbf{n}}{ds} + \frac{d|\nabla\phi|}{ds}\mathbf{n} = -\nabla\left[\left(\mathbf{u} + \frac{\mathbf{B}}{\sqrt{\rho}}\right)\cdot\nabla\phi\right]. \tag{58}$$

Taking the cross product of this equation with $\mathbf{n}$ and again with $\mathbf{n}$, using $d\mathbf{n}/ds\cdot\mathbf{n} = 0$, and then, taking the cross product with $\mathbf{B}$, we see that

$$\left|\mathbf{B}\times\frac{d\mathbf{n}}{ds}\right| \leqslant C, \tag{59}$$

provided that $\mathbf{u} + \mathbf{B}/\sqrt{\rho}$ has bounded derivatives as we have assumed throughout. So,

$$|\sigma_1'(s)| \leqslant C \leqslant C' \inf|\sigma_1(s)|.$$

Thus

$$\epsilon^{1+n}\int d\phi\, d\chi \int_{-L(m)}^{L(m)} |\sigma_1'(s)|^2 |J|\, ds$$
$$\leqslant \epsilon^{1+n}\int d\phi\, d\chi \int_{-L(m)}^{L(m)} |\sigma_1(s)|^2 |J|\, ds = \|\hat{u}\|_2^2.$$

This concludes the estimates needed to show that U_m consititute a Weyl sequence provided that σ_1 satisfies (53), as was to be shown.

VI. SLOW WAVE

Here we would like to address the question as to whether the "slow" branch of the dispersion relation (38) contributes to the ballooning equations as it does in the static case. A $(\pm)$ slow ray is generated by the solution of the bicharacteristic systems

$$\frac{dx_i}{dt} = \frac{\partial}{\partial\phi_{x_i}}\{\mathbf{u}\cdot\nabla\phi \pm c_s|\nabla\phi|\}, \tag{60}$$

$$\frac{d\phi_{x_i}}{dt} = -\frac{\partial}{\partial x_i}\{\mathbf{u}\cdot\nabla\phi \pm c_s|\nabla\phi|\}, \tag{61}$$

where

$$c_s = \{\tfrac{1}{2}(a^2 + A^2 \pm [(a^2 + A^2)^2$$
$$- 4a^2((\mathbf{A}\cdot\nabla\phi)/|\nabla\phi|)^2]^{1/2}\}^{1/2}.$$

In the limit $\mathbf{A}\cdot\nabla\phi \downarrow 0$, the ray trajectory described by (60) will become parallel to $\mathbf{u} + \sqrt{\beta}\,\mathbf{A}$ since the bracket in (60) tends in this limit to $\mathbf{u}\cdot\nabla\phi \pm \sqrt{\beta}\,\mathbf{A}\cdot\nabla\phi$.

In the general case $\mathbf{A}\cdot\nabla\phi \neq 0$, (60) reduces to

$$\frac{dx_i}{dt} = \mathbf{u} + c_s\mathbf{n} \pm a^2\frac{\mathbf{A}\cdot\mathbf{n}}{c_s C}(\mathbf{A} - (\mathbf{A}\cdot\mathbf{n})\mathbf{n}), \tag{62}$$

where

$$C = [(A^2 + a^2)^2 - 4a^2(\mathbf{A}\cdot\nabla\phi/|\nabla\phi|)^2]^{1/2}.$$

These remarks allow us to deduce that for the discussion of the slow wave there are essentially two major cases according as to whether $\mathbf{A}\cdot\nabla\phi = 0$ or $\mathbf{A}\cdot\nabla\phi \neq 0$ at any given point on the ray.

In the first case it is possible to show that $\mathbf{A}\cdot\nabla\phi$ remains zero along the entire trajectory of the ray. Thus since as mentioned above $d\mathbf{x}/dt = \mathbf{u} + \sqrt{\beta}\,\mathbf{A}$, and $\mathbf{u}\cdot\nabla\psi = \mathbf{A}\cdot\nabla\psi = 0$, where we recall that $\psi =$ const. is the family of single-valued flux surfaces introduced in Sec. II, it follows that $d\mathbf{x}/dt\cdot\nabla\psi = 0$ so that a ray that at any point intersects a given flux surface remains on it forever. Moreover, as long as the part of $\mathbf{u}$ not parallel to $\mathbf{B}$, $r\Omega(\psi)\hat{\boldsymbol{\theta}}$, is not zero, $\mathbf{u}\times\mathbf{B}$ "is parallel to" $\nabla\phi$ so that $\nabla\phi$ "is parallel to" $\nabla\psi$. Thus in this case the equation for the $\pm$ slow ray couples with the Alfven rays as in the case studied in Ref. 13.

When $\Omega(\psi_0)=0$, one must distinguish the case $\Omega(\psi)\sim(\psi-\psi_0)^n$ for $n>2$ from the case $n\leqslant 2$. In the former case we are in the situation studied in Sec. III so that again the slow wave couples with the Alfven to yield points in the spectrum. Note that here $\nabla\phi$ need not be parallel to $\nabla\psi$.

In the latter case, it is possible that the normal $\nabla\phi$ assumes the direction $\nabla\psi$ at certain points on trajectory of the ray. If this happens, it is not presently known what implications this has for the spectrum nor whether the cut off indeed occurs for $n=2$.

In case $\mathbf{A}\cdot\nabla\phi\neq 0$, little can be said. However, by taking the inner product of (62) with $\nabla\psi$, we observe that

$$\frac{d\mathbf{x}}{dt}\cdot\nabla\psi = \pm c_s\mathbf{n}\cdot\nabla\psi \pm \frac{(\mathbf{A}\cdot\mathbf{n})^2\mathbf{n}\cdot\nabla\psi a^2}{c_s C}$$
$$= \mathbf{n}\cdot\nabla\psi\,\frac{\pm c_s^2 C - (\mathbf{A}\cdot\mathbf{n})^2 a^2}{c_s C}.$$

From (34), $\mathbf{n}\cdot\nabla\psi$ never vanishes, and by inspection, nor does the square bracket; therefore, the ray trajectories leave the ψ surfaces and can intersect the boundary. In this case it is not possible to get an eigenvalue problem in a neat way as one must take into account the reflection of the ray from the boundary, causing complications which are not tractable within the framework which we identify with ballooning modes.

VII. CONCLUSION

We have given a mathematical derivation of points in the essential spectrum of ideal MHD, the so-called ballooning modes. We showed them to correspond to Alfven and slow magnetosonic waves propagating one-dimensionally along their rays. (We hope to discuss in the future the possible implications of these spectra on heating mechanism.) In axisymmetric configurations the ballooning modes are finite limit points of eigenvalues with increasing azimuthal Fourier mode number, and, therefore, an observer in a rotating frame will see different ballooning modes. This statement should not be misinterpreted. The fact that a rotating observer does not see, say, an unstable ballooning mode does not mean that he sees a stable plasma. He will still see a sequence of discrete unstable eigenvalues, but they will diverge to infinity.

We showed that the most unstable ballooning modes are seen in a frame rotating with the plasma toroidal frequency, while in any other frame only stable ballooning modes are observed. Nonconstancy of the plasma toroidal frequency $\Omega(\psi)$ will stabilize the ballooning modes except at magnetic surfaces ψ_0 where all derivatives of Ω vanish up to $(d^n/d\psi^n)\,\Omega(\psi_0)=0$, with n sufficiently high ($n>2$ will certainly do). Although we expect that this shear stabilization of the ballooning modes is genuine, it is still possible that it only means that the accumulation point, instead of being finite, is shifted to infinity. Even if the flow shear does eliminate the high m unstable modes, and their accumulation points, it may not substantially affect modes with $m\Omega'(\psi_0)\leqslant 1$. Thus stabilization by flow shear cannot be firmly concluded and needs further analysis.

The effect of the plasma rotation on ballooning stability was investigated by an asymptotic expansion in the torus aspect ratio. Our analysis shows that, to leading order, stability requires a minimum energy state, and the perturbed energy consists of the same curvature term as in the static case, plus terms familiar from fluid dynamical flow stability. Regretably, the expected case of confined plasma with only toroidal rotation is potentially destabilized by a rigidly rotating flow.

ACKNOWLEDGMENT

This work was carried out under Air Force Contract No. AFOSR-81-0020 and the U. S. Department of Energy Contract No. DE-AC02-76ER03077.

[1]D. Dobrott, D. B. Nelson, J. M. Greene, A. H. Glasser, M. S. Chance, and E. A. Frieman, Phys. Rev. Lett. **39**, 943 (1977).
[2]J. W. Connor, R. J. Hastie, and J. B. Taylor, Proc. R. Soc. London A **365**, 1 (1979).
[3]D. Correa Restrepo, S. Naturforsch. **33A**, 789 (1978).
[4]R. L. Dewar and A. H. Glasser, Princeton Plasma Physics Laboratory Report PPPL-1890 UC20-G, 1982.
[5]T. Kato, *Perturbation Theory of Linear Operators* (Springer-Verlag, Berlin, 1980), 2nd ed.
[6]E. Hameiri, presented at the Sherwood Meeting on Theoretical Aspects of Controlled Thermonuclear Research, Mount Pocono, PA, 18–20 April 1979.
[7]A. I. Morozov and L. S. Soloviev, Sov. Phys. Dokl. **8**, 243 (1963).
[8]E. Hameiri, Phys. Fluids **26**, 230 (1983).
[9]E. A. Frieman and M. Rotenberg, Rev. Mod. Phys. **32**, 898 (1960).
[10]A. B. Hassam and R. M. Kulsrud, Phys. Fluids **21**, 2271 (1978).
[11]H. P. Zehrfeld and B. J. Green, Nucl. Fusion **12**, 569 (1972).
[12]E. Hameiri, Comm. Pure Appl. Math. (to be published).
[13]E. Hameiri and J. H. Hammer, Phys. Fluids **22**, 1700 (1979).
[14]D. Sigmar, private communication.
[15]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. II.
[16]C. Mercier, *The Magnetohydrodynamics Approach to the Problem of Plasma Confinement in Closed Magnetic Field Configurations*, EUR 5127e (European Communities Directorate General Scientific and Technical Information and Information Management, Luxembourg, 1974).
[17]P. Laurence, Comm. Math. Phys. (to be published).

Global attractors and global stability for closed chemical systems

Ariel Fernández and Oktay Sinanoğlu
Department of Chemistry, Yale University, New Haven, Connecticut 06511

(Received 22 March 1983; accepted for publication 22 September 1983)

The structural stability in the sense of Adronov and Pontriaguin for kinetic models of a closed reacting mixture is investigated. Necessary and sufficient conditions for the mass-action kinetics model to be structurally stable are derived.

PACS numbers: 82.20.Fd

1. INTRODUCTION

A useful model for a physical system must be structurally stable[1] in the sense defined by Adronov and Pontriaguin[2] so that when the control parameters, often not known precisely, are slightly changed or any small perturbation is introduced, the model remains qualitatively the same. The most frequent model for describing the rate of change of chemical concentrations is the "mass-action kinetics" model. By this kinetics, in the theoretical literature[3] is meant a rate law describing a reaction made up of a set of elementary steps where the stoichiometry of each step reflects its molecularity. In the present paper we ask: When is the mass-action kinetics model a structurally stable model for a closed system? In closed systems the concentrations of the chemicals are confined to the so-called reaction simplex R [4] that is, starting from an initial vector of concentrations, the set of all possible concentrations that satisfy the law of conservation of atoms and of conservation of mass.

The boundaries of R are given by the conditions

$$\{c_i = 0,\ i = \alpha_1, \alpha_2, \ldots, \alpha_r\}$$

(c_i = molar concentration of the ith chemical) for certain families of indices $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ such that the laws of conservation of mass and atoms are fulfilled. The simplex R is a segment for a two-component system, a triangle for a three-component system, a tetrahedron for a four-component one, etc.

By a stationary state in the boundary of R we will then mean a state for which all the concentrations remain fixed in time and at least one concentration remains zero. The existence of a thermodynamic potential defined on R does not rule out the possibility of such steady states. In fact they are frequently encountered in autocalytic models which are compatible with thermodynamics (c.f. Ref. 5).

In order to make certain that the equilibrium state is the only global attractor, Krambeck[5] and, implicitly, Gavalas[6] have introduced the additional requirement that no steady state with a vanishing concentration of one of the chemicals should occur. As we shall conclude from this paper, the structural stability restriction is stronger than that of Krambeck and Gavalas.

The weakest condition that ensures the structural stability can be formulated as follows: Consider any face of the boundary of the simplex R given by

$$\{c_i = 0,\ i = \alpha_1, \alpha_2, \ldots, \alpha_n\};\quad n \leqslant r$$

then, at each point of the face there should be at least one α_j, $j = 1, 2, \ldots, n$ such that $dc_{\alpha_j}/dt > 0$. This condition will be referred to as the transversality condition.

It obviously implies that no stationary state with at least one concentration equal to zero (that is, in the boundary of R) can occur.

Actually, it implies more than that. The restriction of Krambeck and Gavalas stated above is derived from the w-stability condition of Wallwork and Perelson[7] which is a necessary but not sufficient condition for the structural stability (c.f. Sec. 3).

The transversality condition can be translated for the mass-action case into restrictions on the kind of molecularity that the elementary steps can have. This is in the sense that (c.f. also Sec. 4) first-order kinetics models fit these restrictions.

For higher-order kinetics further complications arise: Even if a system in which a control chemical acts exclusively as a catalyst (or autocatalyst) in one elementary step and it does not present any steady state in the boundary of R (as defined above), such a model will not be structurally stable since the transversality condition is not fulfilled.

2. A NECESSARY CONDITION FOR STRUCTURAL STABILITY

We shall consider a closed system of S independent elementary reaction steps and N reacting species.

The system is assumed to consist of a single phase and the reactions to occur in a well-stirred vessel under isothermal and isobaric conditions. The changes in composition due to elementary reaction steps obey

$$\mathbf{c}(t) = \mathbf{c}(0) + \underset{\sim}{\nu}\xi \tag{1}$$

in the case of the mass-action kinetics model.

$$\mathbf{c}(t) = (c_1(t), c_2(t), \ldots, c_N(t))$$

is the concentration vector in moles per liter. $\mathbf{c}(0)$ represents the initial concentration vector. $\underset{\sim}{\nu}$ is the stoichiometric matrix $\mathbb{R}^{N\times S} \ni \underset{\sim}{\nu}$.

$$\xi = (\xi_1, \xi_2, \ldots, \xi_S) \tag{2}$$

represents the molar extents of the reactions.[8]

For each $\mathbf{c}(0)$, $\underset{\sim}{\nu}$ maps $\mathbb{R}^S$ into $\mathbb{R}^N$. The reaction simplex R [4] is then the restriction to positive concentrations of the image of $\underset{\sim}{\nu}$. R is a compact set since it is also convex and its dimension is the rank of $\underset{\sim}{\nu}$.

 0022-2488/84/020406-04$02.50

The mass-action kinetics are defined as

$$\frac{d}{dt}\mathbf{c} = \dot{\mathbf{c}} = \mathbf{X}(\mathbf{c}). \tag{3}$$

The vector field $\mathbf{X} = (x_1, x_2,..., x_N)$ is C^1 and each component is a polynomial in c_i's with fixed real coefficients: the mass-action rate constants.

The kinetics are said to be thermodynamically compatible if the Gibbs free energy G defined on R is a Liapunov function, that is

(a) There exists a constant $\kappa > 0$ for which $G(\mathbf{c}) + \kappa > 0$ for all $\mathbf{c} \in R$;

(b) G is strictly convex:

$$\partial^2 G / \partial\xi_i \partial\xi_j > 0 \quad \text{in } R; \tag{4}$$

$$\text{(c)}\ \frac{dG}{dt} = \sum_i^N \frac{\partial G}{\partial c_i} \times \frac{dc_i}{dt} = \sum_i^N \mu_i X_i \leqslant 0, \tag{5}$$

where the equality only holds at the equilibrium, that is at the zero of $\mathbf{X}$. The w-limit set of a point $\mathbf{c} \in R$ is defined: $w(\mathbf{c}) = \lim_{t\to\infty} F_t \mathbf{c}$, where the flow $\{F_t, t \geqslant 0\}$ is the monoparametric family of differomorphisms associated with $\mathbf{X}$ representing the evolution of the system.

$L(R)$ represents the set of C^1-vector fields defined on R and $L_\#(R)$ is the set of C^1-vector fields whose flows leave R invariant.[9] These spaces are normed spaces with the C^1-norm ($\| \ \|_1$) defined as follows:

$$\|\mathbf{X}\|_1 = \underset{\substack{\mathbf{c}\in R \\ i,j = 1,2,...,N}}{\text{Maximum}} \left(|X_j(\mathbf{c})|, \left| \frac{\partial X_j}{\partial c_i}(\mathbf{c}) \right| \right). \tag{6}$$

As we shall prove later, the only valid models $\mathbf{X}$ for the kinetics must be in $L_\#(R)$.

A model $\mathbf{X}$ is said to be w-stable [in $L(R)$] if there exist a neighborhood $U(\mathbf{X})$ of $\mathbf{X}$ in $L(R)$: $U(\mathbf{X}) \subset L(R)$ such that for every $\mathbf{Y} \in U(X)$ there exists a homeomorphism h (bijective and bicontinuous map) of R *onto* itself such that $hw_{\mathbf{X}}(R) = w_{\mathbf{Y}}(R)$. Wallwork and Perelson[7] have shown that a model $\mathbf{X}$ which is w-stable and thermodynamically compatible has exactly one zero in R which is the only global attractor and this is the equilibrium state. We shall define the R-stability of a kinetic model $\mathbf{X}$ as the structural stability of $\mathbf{X}$ regarded as a point in $L_\#(R)$. Rigorously, if there exists a neighborhood $V(\mathbf{X})$ with $V(\mathbf{X}) \subset L_\#(R)$ such that for any $\mathbf{Y} \in V(\mathbf{X})$, there exists a homeomorphism h of R onto itself mapping trajectories of $\mathbf{Y}$ onto trajectories of $\mathbf{X}$.

The concept of R-stability leads to the following definition: An R-perturbation $\boldsymbol{\theta}$ is a perturbation of a vector field $\mathbf{X} \in \mathrm{L}_\#(R)$ such that $\boldsymbol{\theta} + \mathbf{X} \in \mathrm{L}_\#(R)$. As shown in Sec. 4, if $\mathbf{Y}$ and $\mathbf{X}$ are two mass-action kinetics models for the same system, then $\mathbf{Y} - \mathbf{X}$ is an R-perturbation. The physical significance of the R-stability can be given by the following laws: Let $\mathbf{Y}$ be obtained from $\mathbf{X}$ by an R-perturbation. Let E_t be the flow of $\mathbf{Y}$ and γ_{ij} ($i = 1,..., M$) the number of atoms of the class i in the species j and M_j the molecular weight of species j. Then

$$\sum_{j=1}^N \gamma_{ij} c_j = \sum_{j=1}^N \gamma_{ij} E^j_t c_j, \quad i = 1,2,..., M, \tag{7}$$

$$\sum_{j=1}^N M_j c_j = \sum_{j=1}^N M_j E^j_t c_j, \tag{8}$$

where

$$E_t \mathbf{c} = (E^1_t c_1,..., E^N_t c_N), \quad t \geqslant 0. \tag{9}$$

The following theorem relates the w-stability with the R-stability defined above.

Theorem I: Assume that the system is compatible with thermodynamics; that is, G is a convex function defined on the reaction simplex R, and $dG/dt = \nabla G \cdot \mathbf{X} < 0$, except at the point where G attains its minimum; at that point $dG/dt = 0$. Then the w-stability of the vector field $\mathbf{X}$ is equivalent to the R-stability of $\mathbf{X}$.

Proof: The w-stability implies (see Ref. 7) that $\mathbf{X}$ has exactly one global attractor in R. This global attractor is the only zero of $\mathbf{X}$ in R and lies in the interior of R. This is the thermodynamic equilibrium point ϕ where G takes its minimum value on R. This point ϕ retains part of its basin of attraction under C^1-perturbations (hence, under C^1-R-perturbations), that is (for details, see Hirsch and Smale, Ref. 10, p. 316), there exists a neighborhood $U_\beta(\mathbf{X}) \subset L(R)$ and a number $r > 0$ such that the set $U_r(\phi)$ contains the thermodynamic equilibrium point $s = s(\mathbf{Y})$ for each $\mathbf{Y}$ belonging to $U_\beta(\mathbf{X})$ and $U_r(\phi)$ is contained in the basin of attraction of s and it is positively invariant under the flow of $\mathbf{Y}$. Hence $U_r(\phi)$ is retained as a part of the basin of attraction of the thermodynamic equilibrium point of $\mathbf{X}$ under C^1-R-perturbations. We can, therefore assume that $U_\beta(\mathbf{X})$ is contained in $L_\natural(U_r(\phi))$. As the border of the reaction simplex is compact and, from our hypotheses, all trajectories tend asymptotically to a point in the interior of R, there exists $t_0 > 0$ such that $F_t(x)$ belongs to $U_r(\phi)$ for any $t > t_0$ and any x belonging to R.

To see this it suffices to give a covering of ∂R in the following way: For every x belonging to ∂R, there is a neighborhood $U_x \subset \partial R$ of x and $t_x > 0$ such that $F_t(y)$ belongs to $U_r(\phi)$ for every y in $U_x \subset \partial R$ and $t > t_x$. From this covering we can extract a finite subcovering.

There exists a neighborhood $V(\mathbf{X}) \subset L_\#(R)$, such that if $\mathbf{Y}$ belongs to $V(\mathbf{X})$ and E_t is the flow of $\mathbf{Y}$, $E_t(x)$ belongs to $U_r(\phi)$ for $t > t_1 > t_0$ and x in R.

Let $W(\mathbf{X})$ be $U_\beta(\mathbf{X}) \cap V(\mathbf{X})$; we shall show that every $\mathbf{Y}$ belonging to $W(\mathbf{X})$ is equivalent to $\mathbf{X}$.

Any trajectory of $\mathbf{Y}$ belonging to $W(\mathbf{X})$ tends asymptotically to $s(\mathbf{Y})$. We define an homeomorphism $H: R \to R$ as follows: For every x belonging to R, $x \neq \phi$, there is a unique x_0 belonging to ∂R and $t_x > 0$ such that $F_{t_x}(x_0) = x$. We define

$$H(x) = E_{t_x}(F_{-t_x}(x)) = E_{t_x}(x_0),$$

$$H(\phi) = s(\mathbf{Y}).$$

This is a continuous 1-1 map that makes the following diagram commutative:

$$\begin{tikzcd} R \arrow[r, "F_t"] \arrow[d, "H"'] & R \arrow[d, "H"] \\ R \arrow[r, "E_t"] & R \end{tikzcd} \tag{10}$$

Hence we have shown that w-stable vector fields on R which are thermodynamically compatible are also structurally stable with respect to vector fields in $L_{\#}(R)$.

3. THE WEAKEST ADDITIONAL CONDITION TO ENSURE STRUCTURAL STABILITY IN THE GENERAL CASE

The existence of the homeomorphism H is obviously a stronger condition than the existence of a homeomorphism $h: R \to R$ mapping the w-limit set of $\mathbf{x}$ onto the w-limit set of $\mathbf{Y}$. Since this is true for every $\mathbf{Y}$ belonging to $W(\mathbf{X})$, the converse part of the theorem is also proven.

Now assume the following condition is imposed on the boundary of R: If $c_j = 0, j = a, b,..., L$, then there exists $i \in \{a, b,..., L\}$ such that

$$\dot{c}_i > 0. \tag{11}$$

If the system is thermodynamically compatible, then, again, there exists a unique global attractor: the point of thermodynamic equilibrium, hence, as it follows from our proof, the system regarded as a point of $L_{\#}(R)$ is structurally stable. Let $Z(\mathbf{X}) \subset L(R)$ be a sufficiently small neighborhood of $\mathbf{X}$ so that if $\mathbf{Y}$ belongs to $Z(\mathbf{X})$, $\mathbf{Y}$ points inwards along ∂R. Then $Z(\mathbf{X})$ is also contained in $L_{\#}(R)$.

We take now $Z(\mathbf{X}) \cap W(\mathbf{X})$; this is a neighborhood in $L(R)$ [and also in $L_{\#}(R)$] such that every $\mathbf{Y}$ belonging to it is equivalent to $\mathbf{X}$, that is, there exists H as in the theorem above. But then the system $\mathbf{X}$ is structurally stable.

We obtain the following implications for a thermodynamically compatible $\mathbf{X}$.

$$\begin{tikzcd}
\text{w-stability} \arrow[r] \arrow[d, leftrightarrow] & R \text{ repellent} \\
R\text{-stability} & \\
\text{structural stability} \arrow[u] \arrow[r] & \begin{array}{l}\text{At any face } \{c_i = 0, \\ i = \alpha_1, \alpha_2,..., \alpha_r\} \\ \text{there exists } j \in \{1,2,... r\} \\ \text{with } \dot{c}_{\alpha_j} > 0.\end{array} \arrow[uu]
\end{tikzcd}$$

The condition of repellent boundary obtained by Wallwork and Perelson[7] from the postulated w-stability is a necessary but not sufficient condition to ensure the structural stability of the kinetic model.

R could be positively invariant with respect to the flow F_t of $\mathbf{X}$, its boundary could be repellent, and still there could be a point in the boundary of the reaction simplex at which $\mathbf{X}$ is tangent to the boundary, that is, at that point condition (11) does not hold. Under these circumstances it cannot be ensured that under arbitrary small perturbations of $\mathbf{X}$, the resulting flow will map R into itself for all times. This is because for any given $\epsilon > 0$, there exists a perturbation of $\mathbf{X}$ of C^1-norm less than ϵ such that the perturbed vector field points outwards at the point where $\mathbf{X}$ was tangent to the boundary. Hence, the condition of w-stability does not ensure for arbitrarily small perturbations that the resulting vector fields will lie in $L_{\#}(R)$.

4. THE WEAKEST CONDITION FOR STRUCTURAL STABILITY AS APPLIED TO THE MASS-ACTION MODEL

Let us first consider the case of first-order kinetics. The following result readily follows:

Theorem II: If a system $\mathbf{X}$ corresponds to first-order elementary steps, there exists a neighborhood of $\mathbf{X}$ in $L(R)$: $U = U(\mathbf{X})$ such that for every $\mathbf{Y}$ belonging to $U(\mathbf{X})$, $\mathbf{Y}$ obeys the laws given by Eqs. (7)–(9).

Proof: If some species A has zero concentration, then the forward velocity of any reaction step in which it acts as reactant vanishes while the reverse velocity is zero unless the concentration of the product corresponding to that step is zero. By an inductive argument, since the number of steps is finite, there exists a species whose concentration is zero and the forward velocity of the reaction step is zero but the reverse velocity is not. Therefore, we have proven that the vector field points inwards along the boundary ∂R, that is for any point of ∂R there exists at least one i with $c_i = 0$ but $\dot{c}_i > 0$. Then for sufficiently small $\epsilon > 0$, there exists $U = U_\epsilon(\mathbf{X})$ such that for every $\mathbf{Y}$ belonging to U, $\mathbf{Y}$ points inwards along ∂R. Then the flow of $\mathbf{Y}$ leaves R invariant, therefore, Eqs. (7)–(9) hold. Q.E.D.

The same argument does not apply for second- (or higher-) order kinetics; for example in the case

$$\begin{aligned} &A + B \rightleftharpoons C, \\ &A \rightleftharpoons D, \end{aligned} \tag{12}$$

the vector field $\mathbf{X}$ is tangent to the face $\{B = 0, C = 0\}$ of ∂R, hence, for every $\delta > 0$. There exist a C^1-perturbation $\boldsymbol{\theta}$, with $\|\boldsymbol{\theta}\|_1 < \delta$ such that $\boldsymbol{\theta} + \mathbf{X}$ points outwards at the face $\{B = 0, C = 0\}$. Also, if the system contains a catalytic step $A + B \rightleftharpoons B + C$ or an autocatalytic step $A + B \rightleftharpoons 2B$ and species B does not participate in other steps, there the transversality condition for $\mathbf{X}$ again does not hold. The vector field is tangent to the boundary ∂R at the $(N-1)$-dimensional face $\{B = 0\}$: $\dot{B} = 0$. For first-order kinetics, Theorem II above leads to:

Theorem III: A thermodynamically-compatible first-order kinetic system $\mathbf{X}$ is structurally stable.

Proof: There exists (Theorem II) a neighborhood $U \subset L(R)$ of $\mathbf{X}$ which is also contained in $L_{\#}(R)$; hence, from Theorem I, the system is structurally stable.

As for higher-order kinetics, we derive the following.

Theorem IV: If the system $\mathbf{X}$ is subject to the following restriction: For any face of the boundary of the reaction simplex R the following condition does not hold:

> Let $\{A_j = 0\}_{j = a_1, a_2,..., a_L}$ be a generic face of the boundary of the reaction simplex of dimension $(N-L)$. [A $(N-L)$-simplex by itself.] For every $j \in \{a_1,..., \alpha_L\}$, let $\{S_i^{A_j}\}_{i = \alpha, \beta,..., A_j}$ be the set of all elementary steps that include species A_j.
> For each $i \in \{\alpha, \beta,..., A_j\}$ there exists $\kappa \in \{a_1, a_2,..., a_L\}$ such that if A_j is a reactant in step $S_i^{A_j}$ then A_κ is a product in that step and vice versa: if A_j is a product, then A_κ is a reactant in $S_i^{A_j}$.

Then, the thermodynamic compatibility implies the structural stability.

Proof: This condition is clearly the necessary and sufficient condition for $\mathbf{X}$ to be tangent to the $(N-L)$-simplex $\{A_j = 0\}_{j=a_1,a_2,\ldots,a_L}$. Hence, if the condition is not fulfilled, $\mathbf{X}$ points inwards along ∂R, therefore, any sufficiently small C^1-perturbation obeys the laws given by Eqs. (7)–(9) and Theorem I applies. In particular, the condition means that if a species acts exclusively as a catalyst or autocatalyst in a single reaction step, the mass-action kinetics model is structurally unstable.

5. CONCLUSION

The existence of a unique global attractor in the form of a steady state is a feature clearly suggested from the experimental fact that every closed reacting system decays to the thermodynamic equilibrium. In order to ensure this uniqueness in a model, several authors (Krambeck,[5] Galvalas,[6] Wallwork and Perelson[7]) have imposed conditions in addition to the existence of a thermodynamic potential. These conditions are:

(a) the w-stability (Wallwork and Perelson[7]); or the weaker,

(b) no stationary state with at least one vanishing concentration can occur. (In the terminology of this paper, there is no stationary state in the boundary of the reaction simplex R) (Krambeck,[5] Gavalas[6]).

We have demonstrated in this paper that the decay towards a unique equilibrium state is not the only feature that a model should reflect but in addition it should be structurally stable. The properties (a) and (b) are consequences of this inherent global stability of the model.

6. EXAMPLE

Consider the isomerization

$$A_1 \underset{\kappa_{21}}{\overset{\kappa_{12}}{\rightleftharpoons}} A_2 \underset{\kappa_{32}}{\overset{\kappa_{23}}{\rightleftharpoons}} A_3$$

with initial concentrations

$$A_{1_0} = 1, \quad A_{2_0} = A_{3_0} = 0. \tag{13}$$

We concentrate on the following problem: To what extent does the mass-action kinetic model

$$\mathbf{X} = \begin{cases} \dot{A}_1 = -\kappa_{12}A_1 + \kappa_{21}A_2, \\ \dot{A}_2 = -(\kappa_{21}+\kappa_{23})A_2 + \kappa_{12}A_1 + \kappa_{32}A_3, \\ \dot{A}_3 = +\kappa_{23}A_2 - \kappa_{32}A_3 \end{cases} \tag{14}$$

represent the "actual" phase portrait of the system?

Since the model (14) is structurally stable (Theorem II) there exists $\epsilon > 0$ such that the phase portrait of $\mathbf{X}$ is preserved[11] for every $\mathbf{Y}$ such that $\|\mathbf{Y}-\mathbf{X}\|_1 < \epsilon$. Any C^1-perturbation $\boldsymbol{\theta}$ must be tangent to the reaction simplex

$$\{A_1 + A_2 + A_3 = 1,\ A_i > 0,\ i = 1,2,3\},$$

that is, it is of the form

$$\theta = (\theta_1,\theta_2,\theta_3) \quad \text{with } \theta_1+\theta_2+\theta_3 = 0. \tag{15}$$

The preservation of the phase portrait implies that $\mathbf{Y}$ will satisfy the laws given by Eqs. (7)–(9) and it will have strictly one global attractor: a sink which is the image under H of the equilibrium. The vector field $\mathbf{X}$ might defer from the "actual" vector field because

(a) the rate constant κ_{ij} $(i,j = 1,2,3)$ cannot be measured accurately; and

(b) the system obeys mass-action kinetics only approximately.

The following perturbations account for inaccuracies in the measurement of the rate constants in the model given by $\mathbf{X}$:

$$\begin{aligned} \boldsymbol{\theta}^{(1)} &= (-\delta A, \delta A, 0), \\ \boldsymbol{\theta}^{(2)} &= (\delta B, -\delta B, 0), \\ \boldsymbol{\theta}^{(3)} &= (0, -\delta B, \delta B), \\ \boldsymbol{\theta}^{(4)} &= (0; \delta C, \delta C). \end{aligned} \tag{16}$$

If $\delta < \epsilon$, the inaccuracies will not affect the topology of the phase portrait represented by $\mathbf{X}$. Consider the space $L_a(R)$ generated by linear combinations of these perturbations of $\mathbf{X}$.

Then, $L_a(R) \subset L_\#(R)$ since, for any $M > 0$, the vector fields $\mathbf{X} + \boldsymbol{\theta}^{(j)}$ with $\|\boldsymbol{\theta}^{(j)}\|_1 \geqslant M, j = 1,2,3,4$ obey $[A_i = 0 \Rightarrow \dot{A}_i > 0, i = 1,2,3]$. This result implies that these perturbations can only affect the topology of the phase portrait by changing the basin of attraction of the equilibrium of $\mathbf{X}$. Or more rigorously:

Theorem V: If the model (14) for our system (13) lies in the category (a) as shown above, then ϵ can be chosen equal to β. (See Theorem I.)

Proof: (Following the notation of Theorem I). Since we restrict the space $L(R)$ to $L_a(R)$, $V(\mathbf{X}) = L_a(R)$, and $W(\mathbf{X}) = U_\beta(\mathbf{X}) \cap L_a(R)$; $U_\beta(\mathbf{X}) \subset L(R)$.

[1]R. Thom, *Structural Stability and Morphogenesis* (Benjamin, New York, 1975).

[2]A. Adronov and L. Pontriaguin, "Systemes Grossiers," Dok. Akad. Nauk, SSSR **14**, 247–251 (1937).

[3]J. Wei, J. Chem. Phys. **36**, 1578–1584 (1962).

[4]F. Horn and R. Jackson, Arch. Rational Mech. Anal. **47**, 81–116 (1972).

[5]F. Krambeck, Arch. Rational Mech. Anal. **38**, 317–347 (1970).

[6]G. Gavalas, *Nonlinear Differential Equations of Chemically Reacting Systems* (Springer, New York, 1968).

[7]D. Wallwork and A. Perelson, J. Chem. Phys. **65**(1), 284–292 (1976).

[8]M. Boudart, *Kinetics of Chemical Processes* (Prentice-Hall, Englewood Cliffs, NJ, 1968).

[9]That is, for every $\mathbf{Y} \in L(R)$, if $\{E_t, t \geqslant 0\}$ is the flow of $\mathbf{Y}$, $E_t(R) \subseteq R$ for $t \geqslant 0$.

[10]M. Hirsch and S. Smale, *Differential Equations, Dynamical Systems and Linear Algebra* (Academic, New York, 1974).

[11]In the sense that the homeomorphism H of Theorem I exists.