A unified theory of the point groups. II. The general projective representations and their application to space groups

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

(Received 20 July 1982; accepted for publication 1 October 1982)

The general expressions of all inequivalent irreducible projective representations of crystallographic and noncrystallographic point groups (except for the icosahedral group) are given in terms of the vector representations of the proper double point groups. Its application to the representations of the space group of wave vector is discussed. Construction of the basis sets is discussed.

PACS numbers: 02.20. + b

In the previous work[1] (referred to as I) we have presented the general irreps (irreducible representations) of the double group G' of a point group G. Based on these, we shall construct the general projective irreps[2–4] of the double group G'. As is well known, the projective irreps of a group are given by the ordinary vector irreps of its representation group.[4,5] Here, we have the definite advantage that the representation group G'' of G' is much simpler in structure than that of the single group. On account of the special simple structure of G'', which will be given explicitly, one can construct their general irreps by means of a modified theory of induced representation.[6–7] The general projective irreps of the double group G' thus obtained are explicit in terms of the ordinary irreps of the double groups given in I; there is no need of further tabulation. The projective irreps previously available for point groups are limited to crystallographic point groups and are given for each individual group in terms of actual matrices.[3,4] The present result applies for all point groups with a single exception of the icosahedral group; it is more than sufficient to find all the irreps of the space groups of wave vector through simple gauge transformations. In the following, the representative group G'' of the double groups G' will be constructed first, then its vector irreps will be constructed with use of a modified theory of induced representations. Two typical examples of the wave vector groups will be discussed to show how to determine their irreps through gauge transformations.

Since half-integral representations of a point group become vector representations of its double group, many double groups have only one class of the factor systems, i.e., all their representations are p-equivalent to the vector representations. In fact, all proper double groups, the double group T_h' of the group T_h and hence the double groups isomorphic to these have only one class of the factor systems.[2–4] Thus, all double groups with more than one class of the factor systems have the following special form of a direct product:

$$G' = H' \times C_i, \quad C_i = \{e, \hat{i}\}, \tag{1}$$

where H' is limited to the proper double groups C_{2p}', D_n', and O' with integers p and n, and C_i is the group of inversion with e being the identity operation and $\hat{i}$ the inversion. Since H' has only one class, the representation group[4,5] G'' of G' given by (1) can easily be constructed. We may express G'' for the double groups $C_{2p,h}'$, D_{ni}' ($D_{2p,h}'$ or $D_{2p+1,d}'$), O_h' by the defining relations with the group generators as follows (cf. Ref. 4):

$$C_{2p,h}''; \quad a^{2p} = \bar{e}, \hat{i}a = \beta a\hat{i}, \hat{i}^2 = e, \tag{2a}$$

$$D_{ni}''; \quad a^n = b^2 = (ab)^2 = \bar{e}, \hat{i}a = \beta a\hat{i}, \hat{i}b = \gamma b\hat{i}, \hat{i}^2 = e, \tag{2b}$$

$$O_h''; \quad a^4 = b^3 = (ab)^2 = \bar{e}, \hat{i}a = \beta a\hat{i}, \hat{i}b = b\hat{i}, \hat{i}^2 = e, \tag{2c}$$

where $\bar{e}(2\pi$ rotation), β, and γ are all second-order elements in the center of the respective group G''. Their matrix representations are ± 1 with appropriate dimensions except that $\beta = 1$ for $D_{2p+1,d}''$. The representation $\bar{e} = 1$ or -1 identifies the integral or half-integral representations of G''. A set of the vector irreps of G'' characterized by a representation of β or $\{\beta,\gamma\}$ identifies a set of the projective irreps of G' belonging to a class of the factor systems. Since different sets of the possible values of β or $\{\beta,\gamma\}$ correspond to different classes of the factor systems, there exist only two classes of the factor systems for $C_{2p,h}'$, $D_{2p+1,d}'$, and O_h' while there exist only four classes of the factor systems for $D_{2p,h}'$.

In order to construct the irreps of the representation groups G'' given in (2) we observe that G'' has an invariant subgroup H'' of index 2 which has the general structure $H' \times G_\beta$ or $H' \times G_\beta \times G_\gamma$ with $G_\beta = \{e, \beta\}$ and $G_\gamma = \{e,\gamma\}$. Let us decompose G'' into left cosets with respect to H'',

$$G'' = H'' + AH'', \tag{3}$$

where A is the coset representative.

Then, one can construct all irreps of G'' by the induced representations from the irreps of H'' which are easily determined from those of H' given in I. We shall approach this problem in a more direct manner than the ordinary approach such that one can retain all the irreps of H'' and their bases without modification for the direct sums which may occur in the induced irreps. The irreducibility and the completenss of the irreps thus obtained follow naturally without involved arguments on the small (allowable) representations of the little group method.[6]

Let $\{\Gamma^\nu(B)\}$ be a complete set of the irreps of $B \in H$, and let Ψ^ν be a basis row vector of dimension d_ν belonging to the νth irrep $\Gamma^\nu(B)$ such that

$$B\psi^\nu = \psi^\nu \Gamma^\nu(B). \tag{4}$$

 0022-2488/83/030411-03$02.50

Then $A\Psi^{\nu}$ belongs to the conjugate irrep $\Gamma^{\nu}(A^{-1}BA)$ which must be unitarily equivalent to one of the given irreps, say, the μth irrep $\Gamma^{\mu}(B)$ of H''. Thus, there exists a unitary matrix $N(A)$ such that

$$\Gamma^{\nu}(A^{-1}BA) = N(A)^{-1}\Gamma^{\mu}(B)N(A),$$
$$A\psi^{\nu} = \phi^{\mu}N(A), \tag{5}$$

where ϕ^{μ} is a basis belonging to $\Gamma^{\mu}(B)$. Accordingly, the set $[\psi^{\nu}, \phi^{\mu}]$ provides a basis row vector which defines a representation $D^{(\nu,\mu)}$ of G'' given by

$$D^{(\nu,\mu)}(B) = \begin{pmatrix} \Gamma^{\nu}(B) & 0 \\ 0 & \Gamma^{\mu}(B) \end{pmatrix},$$
$$D^{(\nu,\mu)}(A) = \begin{pmatrix} 0 & \Gamma^{\nu}(A^2)N(A)^{-1} \\ N(A) & 0 \end{pmatrix}. \tag{6}$$

This representation is irreducible if and only if $\nu \neq \mu$. The proof is easily established by calculating the sum of the absolute squares of the characters of $D^{(\nu,\mu)}$. In the case $\nu = \mu$, we can take $N(A)$ such that $N(A)^2 = \Gamma^{\nu}(A^2)$ adjusting the phase factor of $N(A)$. Then one sees immediately that the representation (6) can be reduced by a unitary transformation into two inequivalent irreps $D^{(\nu+)}$ and $D^{(\nu-)}$ given by

$$D^{(\nu\pm)}(B) = \Gamma^{\nu}(B), \quad D^{(\nu\pm)}(A) = \pm N(A), \tag{7}$$

with the respective basis

$$\psi^{(\nu\pm)} = \psi^{\nu} \pm A\psi^{\nu}N(A)^{-1} \tag{8}$$

excluding the null case. The completeness of the induced irreps thus obtained is easily established by calculating the sum of the squares of the dimensions of the induced irreps, provided that one uses every irrep of H'' once and only once in constructing the induced irreps.

The representations given by (6) and (7) are completely explicit in terms of the vector irreps of H' except for the matrix $N(A)$. Obviously, $N(A) = 1$ when Γ^{ν} is one dimensional. In general, if one uses the general irreps of C'_{2p}, D'_n, and O' given in I, one can easily determine $N(A)$ explicitly for each class of the factor systems with use of the commutation relations of the group generators given in (2). It is simplest to take $A = \hat{i}$. Then $\Gamma^{\nu}(\hat{i}^2) = 1$ for all Γ^{ν}, and for given values of β and γ we have

$$\beta\Gamma^{\nu}(a) = N(\hat{i})^{-1}\Gamma^{\mu}(a)N(\hat{i}),$$
$$\gamma\Gamma^{\nu}(b) = N(\hat{i})^{-1}\Gamma^{\mu}(b)B(\hat{i}) \tag{9}$$

from which follows $N(\hat{i})$. The explicit representations thus obtained are denoted by $D(\Gamma^{\nu},\Gamma^{\mu};N(\hat{i}))$ for $D^{(\nu,\mu)}$ of (6) and by $D(\Gamma^{\nu};\pm N(\hat{i}))$ for $D^{(\nu\pm)}$ of (7). In the following, we shall give all projective irreps of $C'_{2p,h}$, D'_{ni}, and O'_h except the vector irreps given by $D(\Gamma^{\nu};\pm 1)$ corresponding to the class characterized by $\beta = 1$ or $\{\beta = 1, \gamma = 1\}$.

Group $C'_{2p,h}$: According to I, the full set of the vector irreps of C'_{2p} is given by a set of $4p$ one-dimensional irreps $\{M_m;\ -p < m \leqslant p;\ m$ an integer or half-integer$\}$. The projective irreps corresponding to $\beta = -1$ are given by a total of $2p$ two-dimensional irreps,

$$D(M_m, M_{m-p};1), \quad m = \tfrac{1}{2},1,\ldots,p. \tag{10}$$

Note that in the limiting case of $C'_{\infty h}$, we have only one-dimensional vector representations as in the case of S_{4p+2}.

Group D'_{ni} ($D'_{2p,h}$ *or* $D'_{2p+1,d}$): The irreps of D'_n are given in I by four one-dimensional irreps A_1, A_2, B_1, and B_2 and $(N-1)$ two-dimensional irreps E_m, $m = \frac{1}{2}, 1, \ldots, \frac{1}{2}(n-1)$ defined by 2×2 matrices $M^j{}_m$, where $j \geqslant m$. It turns out that it is simplest to take the integral part of j even to determine the matrix $N(\hat{i})$. Then, the projective irreps belonging to each class $K(\beta,\gamma)$ of the factor systems are given by

$$K(1,-1): D(A_1,A_2;1),\ D(B_1,B_2;1),\ D(E_m;\pm\sigma_y), \quad m = \tfrac{1}{2}, 1, \ldots, \tfrac{1}{2}(n-1), \tag{11a}$$

$$K(-1,1): D(A_1,B_2;1),\ D(A_2,B_1;1),\ D(E_{p/2};\pm\sigma_z),\ D(E_{m'},E_{p-m'};\sigma_z),\ m' = \tfrac{1}{2}, 1, \ldots, \tfrac{1}{2}(p-1);\ n = 2p, \tag{11b}$$

$$K(-1,-1)\ D(A_1,B_1;1),\ D(A_2,B_2;1)(,\ D(E_{p/2};\pm\sigma_x)\ D(E_{m'},E_{p-m'};\sigma_x),\ m_1 = \tfrac{1}{2}, 1, \ldots, \tfrac{1}{2}(p-1);\ n = 2p, \tag{11c}$$

where $\sigma_x, \sigma_y, \sigma_z$ are the Pauli spin matrices representing $N(\hat{i})$ and the last two classes occur only for even n. There are $2n$ two-dimensional irreps for the first class $K(1,-1)$, while there exists four two-dimensional irreps and $(p-1)$ four-dimensional irreps for the last two classes. In the limiting case of $D'_{\infty h}$, there exist only two classes of the factor systems $K(1,1)$ and $K(1,-1)$ analogous to $D'_{2p+1,d}$. The projective irreps of $D'_{\infty h}$ belonging to $K(1,-1)$ are given by

$$D(A_1,A_2;1), \quad D(E_m;\pm\sigma_y); \quad m = \tfrac{1}{2},1,\ldots,\infty \tag{12}$$

since there exist only two one-dimensional irreps A_1 and A_2 for D'_{∞} according to I.

Group O'_h: The irreps of O' has been given in I by $\Gamma_1 \sim \Gamma_5$ and Γ_6, $\Gamma_7 = \Gamma_6\times\Gamma_2$, $\Gamma_8 = \Gamma_6\times\Gamma_3$ with dimensions (1, 1, 2, 3, 3; 2, 2, 4) where the last three irreps are half-integral representations. The projective irreps belonging to $K(\beta = -1)$ are given by

$$D(\Gamma_1\Gamma_2;1),\ D(\Gamma_3;\pm\sigma_y),\ D(\Gamma_4,\Gamma_5;1_3);\ D(\Gamma_6,\Gamma_7;1_2),$$
$$D(\Gamma_8;\pm\Sigma_y) \tag{13}$$

with dimensions (2, 2, 2, 6; 4, 4, 4) where 1_3 and 1_2 are the 3×3 and 2×2 unit matrices and Σ_y is the 4×4 Dirac spin matrix given by the direct sum $\sigma_y \oplus \sigma_y$.

We have thus constructed all projective irreps of all point groups in terms of the vector irreps of the proper double groups except for the icosahedral group. These results (10)–(13) are more than sufficient to find the representations[7–9] of any space group of wave vector $G(k)$. It is only necessary to find the appropriate gauge transformations[3–4] which make the generators of $G(k)$ satisfy the defining relations of the corresponding representation groups given in (2). Let us illustrate the procedure through two typical examples of $G(k)$ in the following:

Example 1: 0_h^2 (No. 222) at $R = \{\pi/a,\pi/a,\pi/a\}$ of the Brillouin zone following the notations given by Zak *et al.*[8] The wave vector group $G(k)$ has the following symmetry,

$$0 + \left(I \,\middle|\, \frac{a}{2}, \frac{a}{2}, \frac{a}{2}\right) 0, \tag{14}$$

where 0 stands for the symmorphic space group, with the point group symmetry 0. If we identify the generating opera-

tors by

$$a\leftrightarrow(C_4^z|000),\quad b\leftrightarrow(C_3^{xyz}|000),\quad \hat{i}\leftrightarrow\left(I\left|\frac{a}{2}\frac{a}{2}\frac{a}{2}\right.\right),\tag{15}$$

we obtain $\beta=-1$ with use of (2c). Accordingly, all irreps of $G(k)$ are given by (13).

Example 2: D_{6h}^4 (No. 194) at $A=\{0,0,\pi/c\}$ of the Brillouine zone. The symmetry of $G(k)$ is described by

$$D_{3d}+\left(C_2\left|00\frac{c}{2}\right.\right)D_{3d}=\left[D_3+\left(C_2\left|00\frac{c}{2}\right.\right)D_3\right]\times C_i,\tag{16}$$

where C_i is the group of inversion. We make the following correspondence for the generating operators of D_{6h}:

$$a\leftrightarrow\left(C_6\left|00\frac{c}{2}\right.\right),\quad b\leftrightarrow i(U^x|000),\quad \hat{i}\leftrightarrow(I|000),\tag{17}$$

where $i=\sqrt{-1}$ and U^x is a binary rotation perpendicular to the sixfold rotation axis C_6. Then, we obtain the following defining relations:

$$a^6=b^2=(ab)^2=-\bar{e},\quad \hat{i}a=-a\hat{i},\quad \hat{i}b=b\hat{i},$$
$$\hat{i}^2=e.\tag{18}$$

Comparison with (2b) yields $\beta=-1$ and $\gamma=1$. The sign difference in front of $\bar{e}$ between (18) and (2b) simply exchanges the integral and the half-integral irreps. Thus, we obtain six irreps of $G(k)$ from (17) and (11b), which are explicitly given by

$$D(A_1,B_2;1_2),\ D(A_2,B_1;1_2),\ D(E_1,E_2;\sigma_z);$$
$$D(E_{1/2},E_{5/2};\sigma_z);\ D(E_{3/2};\pm\sigma_z),\tag{19}$$

of which the first three describe the half-integral representations of $G(k)$ while the last three describe the integral representations of $G(k)$. This kind of exchange is not uncommon (see Ref. 8). It is interesting to note that if we take $A=(C_2|00\frac{c}{2})$ for the augmenting operator A of (3), this kind of exchange does not occur. This requires, however, recalculation of the induced representations given by (6) and (7) in terms of the irreps of D_{3d}.

CONCLUDING REMARKS

We have presented the general expressions of all inequivalent irreducible projective representatios of the double groups $C'_{2p,h}$. $D'_{2p,h}$, $D'_{2p+1,d}$, O'_h and their limiting double groups in (10)–(13). These are explicitly given in terms of the vector irreps of the corresponding proper double point groups. The irreps of the remaining double point groups are p-equivalent to their vector irreps. Here, we have excluded the icosahedral groups. The present results and the general irreps of the proper double point groups given in I are more than sufficient to find all the representations of any space group of wave vector $G(k)$. This is exemplified by the wave vector groups of O_h^2 and D_{6h}^4 as given in Examples 1 and 2.

One of the most important applications of the projective irreps is to construct the symmetry adapted functions belonging to the irreps.[3] The present formalism is particularly suitable to this purpose since according to (5) and (8), the basis functions of the induced irreps of G' are given in terms of the basis functions of the subgroup H'. Thus there is no need to use the laborious projection operator method to construct the basis functions of G'. The construction of the basis of H' is also very simple since all projective irreps of H' are p-equivalent to vector irreps of H'. Thus, one can easily incorporate the theorems[10] on the symmetry adapted functions developed for the vector irreps into those belonging to the projective irreps.

As one can see from the flow of logic in the present treatment of induced representation, if one replaces the augmenting operator A in (3) by an antilinear operator, the representations given by (6) and (7) hold for the corepresentations[11] for a magnetic group with minor modifications. It is also mentioned that the present treatment of the induced representations can easily be extended to a finite group with a normal subgroup of prime index. These extensions of the present work as well as further applications of the present theory in constructing the symmetry adapted linear combinations of the equivalent basis functions will, however, be discussed somewhere else.[12]

[1] S. K. Kim, J. Math. Phys. **22**, 2101 (1981); Mol. Phys. **43**, 1447 (1981).
[2] A. Kitz, Phys. Status Solidi **8**, 813 (1965).
[3] A. C. Hurley, Philos. Trans. R. Soc. London Ser. A **260**, 1108 (1966).
[4] G. L. Bir and G. E. Pikus, *Symmetry and Strain-induced Effects in Semiconductors*, translated from Russian by P. Shelnitz (Wiley, New York, 1974).
[5] S. L. Altmann, *Induced Representations in Crystals and Molecules* (Academic, New York, 1977).
[6] V. V. Raghavacharyulu, Can. J. Phys. **39**, 830 (1961).
[7] 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).
[8] J. Zak, A. Casher, M. Glück, and Y. Gur, *The Irreducible Representations of Space Groups* (Benjamin, New York, 1969).
[9] C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972) and references therein.
[10] S. K. Kim, Physica A **105**, 577 (1981); **106**, 521 (1981).
[11] E. P. Wigner, *Group Theory and Its Applications to Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
[12] S. K. Kim (to be published).

A unified theory of the point groups. III. Classification and basis functions of improper point groups

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

(Received 20 July 1982; accepted for publication 1 October 1982)

This paper introduces a new system of classification of improper point groups which is most effective for describing their general irreducible representations. The complete set of the general angular momentum eigenfunctions is classified according to the general irreducible representations of the seven sets of the improper point groups corresponding to $C_n(C_\infty)$ and $D_n(D_\infty)$ for an aribtrary n.

PACS numbers: 02.20. + b

1. INTRODUCTION

In the previous work[1] (referred to as I), the double group of a proper point group has been constructed as a subgroup of SU(2) by imposing a set of polynomial equations on the matrix elements of a 2×2 unitary matrix belonging to SU(2). The formalism is proven to be very effective in constructing the general expressions of the vector and spinor irreps (irreducible representations) of the point groups in a unified manner. Moreover, it has enabled us to classify the complete set of the pure spinors $\{\phi(j,m); j, m$ integers or half integers; $j \geqslant m\}$ by the general irreps of the uniaxial group $C_n(C_\infty)$ and the dihedral group $D_n(D_\infty)$ for an arbitrary n.

In the present paper, we shall classify the complete set of the general angular momentum eigenfunctions $\{\psi(j,m)\}$ according to the general irreps of the improper point groups corresponding to C_n and D_n for a general n. This will be achieved by using the correspondence[2-4] between $\{\phi(j,m)\}$ and $\{\psi(j,m)\}$ under the point operations. It turns out, however, that the existing classifications of the improper point groups $\bar{G}$ by the Schönflies[5] or Shubnikov[6] or the international symbols[7] are not really suitable for describing the general irreps of $\bar{G}$ corresponding to C_n and D_n for a general n. This can be demonstrated by the following simple examples of the isomorphisms written in Schönflies symbols:

$$D_{2n_0} \simeq D_{n_0 h}, \quad D_{2n_e} \simeq D_{n_e d}, \tag{1.1}$$

where $\simeq$ denotes the isomorphism and $n_0(n_e)$ is odd (even) n. Thus, it is highly desirable to have a modified system of classification of $\bar{G}$ which does not require such distinction on n when describing the isomorphisms [see Eq. (2.5)] or the general irreps of $\bar{G}$. In Sec. 2, we shall show that such a convenient classification of $\bar{G}$ follows very naturally from the method of constructing $\bar{G}$ from its halving subgroup. With the modified classification, we shall, then, describe the basis functions belonging to the irreps of $\bar{G}$ corresponding to C_n and D_n in terms of the complete set of angular momentum eigenfunctions $\psi(j,m)$ in Sec. 3. We shall leave out, however, the discussion on the basis functions of the improper point groups related to tetrahedral (T), cubic (O), and icosahedral (Y) groups partly because we are as yet unable to classify the complete set of $\phi(j,m)$ according to the irreps of T, O, and Y, and partly because comprehensive tables of the basis functions for the cubic groups are available in the existing tables.[8]

TABLE I. Improper point groups, $\bar{G} = H + AH$.

Present	H	A	Isomorphism	Schönflies	Shubnikov
$SO(3)_i$	SO(3)	$\hat{i}$	$SO(3)\times C_i$	$SO(3)_h$	$\infty/\infty \cdot m$
$C_{\infty i}$	C_∞	$\hat{i}$	$C_\infty \times C_i$	$C_{\infty h}$	$\infty:m$
$C_{\infty v}$	C_∞	$\bar{c}_2'$	D_∞	$C_{\infty v}$	$\infty \cdot m$
$D_{\infty i}$	D_∞	$\hat{i}$	$D_\infty \times C_i$	$D_{\infty h}$	$m \cdot \infty:m$
C_{ni}	C_n	$\hat{i}$	$C_n \times C_i$	$C_{n_e h}, S_{2n_0}$	$n_e:m,\ \overline{2n}_0$
C_{np}	C_n	$\bar{c}_{2n}$	C_{2n}	$C_{n_0 h}, S_{2n_e}$	$n_0:m,\ \overline{2n}_e$
$C_{nv}(n>1)$	C_n	$\bar{c}_2'$	D_n	C_{nv}	$n \cdot m$
$D_{ni}(n>1)$	D_n	$\hat{i}$	$D_n \times C_i$	$D_{n_e h}, D_{n_0 d}$	$m \cdot n_e:m,\ \overline{2n}_0 \cdot m$
$D_{np}(n>1)$	D_n	$\bar{c}_{2n}$	D_{2n}	$D_{n_0 h}, D_{n_e d}$	$m \cdot n_0:m,\ \overline{2n}_e \cdot m$
T_i	T	$\hat{i}$	$T \times C_i$	T_h	$\bar{6}/2$
T_p	T	$\bar{c}_4$	O	T_d	$\overline{3/4}$
O_i	O	$\hat{i}$	$O \times C_i$	O_h	$\bar{6}/4$
Y_i	Y	$\hat{i}$	$Y \times C_i$	Y_h	$3/\overline{10}$

(1) $\bar{c}_{2n} = \hat{i}c_{2n}, \bar{c}_2' = \hat{i}c'_2; \hat{i}$ = inversion.
(2) n_e (n_0) is even (odd) n.
(3) The different notations on the fifth line mean that $C_{n_e i} \leftrightarrow C_{n_e h} \leftrightarrow n_e:m$ and $C_{n_0 i} \leftrightarrow S_{2n_0} \leftrightarrow \overline{2n}_0$.

 0022-2488/83/030414-05$02.50

TABLE II. The irreps and the bases of $C_{\infty i}$ ($C_{\infty h}$).

$C_{\infty i}$	$C_z(\alpha)$	$\hat{i}$	Bases
M_{mg}	$\exp(-im\alpha)$	1	$\psi(j,m)_g$
M_{mu}	$\exp(-im\alpha)$	-1	$\psi(j,m)_u$

(1) Cf. Eq. (5.2) of I.
(2) $-\pi<\alpha\leqslant\pi$, $m=0,\ \pm\frac{1}{2},\ \pm 1,\cdots$.

2. CLASSIFICATION OF THE IMPROPER POINT GROUPS

In order to arrive at a new modified system of classification which is best suitable for describing the general irreps of improper point groups $\overline{G}$, we shall first reexamine the construction of $\overline{G}$. In general,[5] an improper point group $\overline{G}$ has an invariant subgroup H of index 2 which is a proper point group. Thus, one can decompose $\overline{G}$ into the left cosets with respect to H,

$$\overline{G}=H+AH, \tag{2.1a}$$

where A is an improper point operation which satisfies

$$ABA^{-1},\quad A^2\in H \quad \text{for any } B\in H. \tag{2.1b}$$

On account of condition (2.1b), only a very limited number ($\leqslant 3$) of improper operators are compatible with any given H, obviously within a multiplicative operator $B\in H$. For example, when $H=C_n$, the allowed set of A may be chosen to be

$$A=\hat{i},\quad \bar{c}_{2n}(=\hat{i}c_{2n}),\quad \bar{c}_2'(=\hat{i}c_2'), \tag{2.2}$$

where $\hat{i}$ is the inversion, $\bar{c}_{2n}$ is the $2n$-fold rotatory-inversion axis parallel to $c_n\in C_n$, and $\bar{c}_2'$ is a mirror reflection parallel to c_n. These correspond to $\bar{1}$, $\overline{2n}$, and $\bar{2}$ in the international notations. For $H=D_n$, we have only two augmenting operators,

$$A=\hat{i},\quad \bar{c}_{2n}, \tag{2.3}$$

where $\bar{c}_2'$ does not come in since $c_2'\in D_n$. Analogously, for $H=T$, we have $A=\hat{i},\bar{c}_4$ and for $H=0$ or Y we have only $A=\hat{i}$. In this way, we have constructed the complete set of improper point groups as given by Table I. It shows that only three types of operations $\hat{i}$, $\bar{c}_{2n}$, and $\bar{c}_2'$ are sufficient for A in order to construct all the improper point groups from the proper point groups.

Since a given set $\{H,A\}$ uniquely defines a $\overline{G}$, it seems most natural to classify $\overline{G}$ by the set $\{H,A\}$ and denote the set by H_A with A being expressed by symbols;

$$H_i,\ H_p,\ H_v, \tag{2.4}$$

where suffixes i,p, and v correspond to the augmenting operators $\hat{i}$, $\bar{c}_{2n}$, and $\bar{c}_2'$, respectively. Thus, for example, when $H=C_n$ or D_n we have C_{ni}, C_{np}, C_{nv} or D_{ni}, D_{np}. In Table I, these symbols are compared with the Schönflies[5] and the Shubnikov[6] symbols. In particular, note that $T_i=T_h$, $T_p=T_d$, and $O_i=O_h$. The symbol H_v agrees with the Schönflies while H_i agrees with it only when $\overline{G}=C_{n_0i}$. It is also noted that the present symbols and the international symbols use rotatory-inversion axis instead of rotation-reflection axis.

Now we are ready to construct all the irreps of $\overline{G}$ from those of H. By definition, all irreps of H_i follow from

$$H_i\simeq H\times C_i,\quad C_i=\{e,\hat{i}\}, \tag{2.5a}$$

while those of H_p and H_v follow from

$$C_{np}\simeq C_{2n},\quad D_{np}\simeq D_{2n},\quad T_p\simeq 0,\quad C_{nv}\simeq D_n,\ n>1 \tag{2.5b}$$

with the correspondence $\bar{c}_{2n}\leftrightarrow c_{2n}$ and $\bar{c}_2'\leftrightarrow c_2'$. These isomorphisms will hold for their double groups as well. There exist further isomorphisms given by

$$C_{n_0i}\simeq C_{2n_0},\quad D_{n_0i}\simeq D_{2n_0},\quad C_{2i}\simeq D_2 \tag{2.6}$$

with the correspondence $\hat{i}\leftrightarrow c_2$, which holds, however, only for the single groups since $\hat{i}^2=e$, $c_2^2=\bar{e}$ (2π rotation) for a double group. Thus, only the former set of isomorphisms is necessary and sufficient to construct all the irreps (vector and spinor) of H_i, H_p, and H_v from those of proper point

TABLE III. The irreps and the bases of $C_{\infty v}$ ($\simeq D_\infty$).

$C_{\infty v}$	$C_z(\alpha)$	$\bar{C}_{2,\beta}'$	Bases
A_1	1	1	$\psi(j_e,0)_g$, $\psi(j_0,0)_u$
A_2	1	-1	$\psi(j_e,0)_u$, $\psi(j_0,0)_g$
E_m	$M_m(\alpha)$	$M^j{}_m(\beta)$	$[\psi_+(j,m),\psi_-(j,m)]_g$
			$[\psi_-(j,m),-\psi_+(j,m)]_u$

(1) Cf. Table I of I.
(2) $-\pi<\alpha,\beta\leqslant\pi$.
(3) j_e (j_0) is a j with an even (odd) integral part.
(4) $M_m(\alpha)=\begin{bmatrix}\cos m\alpha & -\sin m\alpha\\ \sin m\alpha & \cos m\alpha\end{bmatrix}$,

$$M^j{}_m(\beta)=(-i)^{2j}\begin{bmatrix}\cos m\beta & \sin m\beta\\ \sin m\beta & -\cos m\beta\end{bmatrix}.$$

TABLE IV. The irreps and the bases of $D_{\infty i}$ ($D_{\infty h}$).

$D_{\infty i}$	$C_z(\alpha)$	$C_{2,\beta}'$	$\hat{i}$	Bases
A_{1g}	1	1	1	$\psi(j_e,0)_g$
A_{1u}	1	1	-1	$\psi(j_e,0)_u$
A_{2g}	1	-1	1	$\psi(j_0,0)_g$
A_{2u}	1	-1	-1	$\psi(j_0,0)_u$
E_{mg}	$M_m(\alpha)$	$M^j{}_m(\beta)$	1	$[\psi_+(j,m),\psi_-(j,m)]_g$
E_{mu}	$M_m(\alpha)$	$M^j{}_m(\beta)$	-1	$[\psi_+(j,m),\psi_-(j,m)]_u$

(1) Cf. Table III.

TABLE V. The irreps and bases of C_{ni} ($C_{n_e h}$, S_{2n_o}).

C_{ni}	C_n^k	$\hat{i}$	Bases
M_{mg}	$\exp(-i2\pi mk/n)$	1	$\psi(j,m)_g$
M_{mu}	$\exp(-i2\pi mk/n)$	-1	$\psi(j,m)_u$

(1) Cf. (5.4) of I, where k and m are defined.
(2) $\psi(j,m)$ and $\psi(j,m\pm n)$ belong to the same irrep.

groups. The results of construction will be presented in the next section together with their basis sets.

Before concluding this section, it is worthwhile to mention that the present system of the symbols is also very effective in describing the complete set of the projective irreps of the point groups[9] and that it can easily be extended to classify the Shubnikov black and white point groups.[10] The present symbols also simplify the notations for the space groups. One can see this by comparing the present symbols for the seven crystal systems (the point symmetry of the Bravais lattice),

$$C_i,\ C_{2i},\ D_{2i},\ D_{4i},\ D_{3i},\ D_{6i},\ O_i, \tag{2.7}$$

with those written in Schönflies notation,

$$S_2,\ C_{2h},\ D_{2h},\ D_{4h},\ D_{3d},\ D_{6h},\ O_h. \tag{2.8}$$

One also sees that (2.8) follows easily from (2.7) since $\bar{c}_2'$ is always a reflection through a plane perpendicular to c_2'. The symmetry hierarchy and the solvability of improper point groups are also easier to understand with the present symbols.

3. THE BASIS FUNCTIONS OF THE IMPROPER POINT GROUPS

We shall now construct the basis functions belonging to the irreps of the improper point groups. For this purpose, we shall first discuss the correspondence between the pure spinor bases $\{\phi(j,m)\}$ and the general angular momentum eigenfunctions $\{\psi(j,m)\}$. The former[1] is defined by a set of $(2j+1)$ monomials of two elementary spinors ξ_1 and ξ_2,

$$\phi(j,m)=\xi_1^{j-m}\xi_2^{j-m}/[(j+m)!(j-m)!]^{1/2}, \tag{3.1}$$
$$m=j,,j-1,\ldots,-j,$$

where j is an integer or a half integer. It is stressed here that we have assumed the positive sign convention for the square root in (3.1) without any other phase factor which may depend on m. The correspondence between $\{\phi(j,m)\}$ and $\{\psi(j,m)\}$ has been established first by Weyl,[2] then by van der Waerden[3] and by others[4] with varying degrees of sophistication. We find it most convenient to restate it in a form of a theorem as follows:

Theorem: The representation $D^j(U)$ of $U\in$SU(2) based on the set of the $(2j+1)$ pure spinors $\{\phi(j,m)\}$ defined by (3.1) is *identical* with the representation of the double group of SO(3) based on a set of $(2j+1)$ general angular momentum eigenfunctions $\{\psi(j,m)\}$ for a system of any number of the electrons, provided that the set satisfies the phase convention due to Condon and Shortley.[4]

It should be noted that $\psi(j,m)$ may depend on the spatial variable or the spin variable or both and j, m are the resultant angular momentum quantum numbers. In spite of the existing proofs of the theorem, it may be worthwhile to give the following simple proof which manifests the pahse convention. Let $\{\Phi(j,m)\}$ be a set of spin angular momentum eigenfunctions of a system of $2j$ electrons given by a set of symmetrized spinors,

$$\Phi(j,m)=[(2j)!(j+m)!(j-m)!]^{-1/2}\sum_P P\xi_1(1)\cdots\xi_1(j+m)\xi_2(j+m+1)\cdots\xi_2(2j), \tag{3.2}$$

where the summation is over $(2j)!$ permutations P's of the $2j$ electrons. Here, Condon–Shortley's convention amounts to the positive sign convention for the square root in (3.2). In view of (3.1) and (3.2) it is evident that the representation $D^j(U)$ of $U\in$SU(2) based on $\{\phi(j,m)\}$ is the same as those based on $\{\Phi(j,m)\}$ under the direct product operation U(1)U(2)$\cdots$U($2j$) of $2j$ particles. Now, the theorem is proven since both sets $\{\Phi(j,m)\}$ and $\{\psi(j,m)\}$ transform exactly in the same manner under the infinitesimal rotations expressed by the total angular momentum operator of the system.

Now, to account for the inversion symmetry, let us denote an even and odd $\psi(j,m)$ by $\psi(j,m)_g$ and $\psi(j,m)_u$, respectively. Then in comparison with the pure spinor bases of the proper point groups given in I, one can express all the basis functions belonging to the vector or spinor representations of the improper point groups by $\psi(j,m)_{g/n}$ or by their linear combinations

$$\psi_+(j,m)_{g/u}=2^{-1/2}[\psi(j,m)+\psi(j,-m)]_{g/u},$$
$$\psi_-(j,m)_{g/u}=-i2^{-1/2}[\psi(j,m)-\psi(j,-m)]_{g/u}. \tag{3.3}$$

As is well known, when j equals an integer l one can express

TABLE VI. The irreps and bases of C_{np} ($C_{n_o h}$, S_{2n_e})$\simeq C_{2n}$.

C_{np}	C_{2n}^{ke}	$\bar{C}_{2n}^{k_o}$	Bases
A	1	1	$\psi(j,0)_g$, $\psi_\pm(j,2n)_g$, $\psi_\pm(j,n)_u$
B	1	-1	$\psi(j,0)_u$, $\psi_\pm(j,2n)_u$, $\psi_\pm(j,n)_g$
M_m	$\exp(-i\pi mk_e/n)$	$\exp(-i\pi mk_o/n)$	$\psi(j,m)_g$, $\psi(j,m-n)_u$

(1) k_e (k_o) is even (odd) k with $k=0,\pm1,\ldots,\pm(n-1),n$.
(2) $-n<m\neq 0<n$.
(3) $\psi(j,m)$ and $\psi(j,m\pm 2n)$ belong to the same irrep.

TABLE VII. The irreps and the bases of $C_{nv} \simeq D_n$.

C_{nv}	C_n^k	$\bar{C}'_{2,q}$	Bases
A_1	1	1	$\psi(j_{e/o}, 0)_{g/u}, \psi_\pm(j_e, n)_{g/u}, \psi_\pm(j_0, n)_{u/g}$
A_2	1	-1	$\psi(j_{e/o}, 0)_{u/g}, \psi_\pm(j_e, n)_{u/g}, \psi_\pm(j_0, n)_{g/u}$
B_1	$(-1)^k$	$(-1)^{q+1} i^{-\epsilon(n)}$	$\psi_\pm(j_0, n/2)_{g/u}, \psi_\pm(j_e, n/2)_{u/g}$
B_2	$(-1)^k$	$(-1)^q i^{-\epsilon(n)}$	$\psi_\pm(j_0, n/2)_{u/g}, \psi_\pm(j_e, n/2)_{g/u}$
E_m	$M_m(k)$	$M'^j_m(q)$	$[\psi_+(j, m), \psi_-(j, m)]_g$
			$[\psi_-(j, m), -\psi_+(j, m)]_u$

(1) Cf. Table II of I where k, q, and m are defined.
(2) $\epsilon(n) = 0$ for even n and $\epsilon(n) = 1$ for odd n.
(3) $\psi_\pm(j_e, n)_{g/u}$ means $\psi_+(j_e, n)_g$ or $\psi_-(j_e, n)_u$.
(4) $M_m(k) = M_m(\alpha = 2\pi k/n)$, $M'^j_m(q) = M'^j_m(\beta = 2p/n)$ where $M_m(\alpha)$ and $M'^j_m(\beta)$ are defined in Table III.

TABLE VIII. The irreps and the bases of D_{ni} ($D_{n_e h}$, $D_{n_o d}$).

D_{ni}	C_n^k	$C'_{2,q}$	$\hat{i}$	Bases
A_{1g}	1	1	1	$\psi(j_e, 0)_g, \psi_\pm(j_{e/o}, n)_g$
A_{1u}	1	-1	-1	$\psi(j_e, 0)_u, \psi_\pm(j_{e/o}, n)_u$
A_{2g}	1	-1	1	$\psi(j_o, 0)_g, \psi_\pm(j_{o/e}, n)_g$
A_{2u}	1	-1	-1	$\psi(j_o, 0)_u, \psi_\pm(j_{o/e}, n)_u$
B_{1g}	$(-1)^k$	$(-1)^{q+1} i^{-\epsilon(n)}$	1	$\psi_\pm(j_{o/e}, n/2)_g$
B_{1u}	$(-1)^k$	$(-1)^{q+1} i^{-\epsilon(n)}$	-1	$\psi_\pm(j_{o/e}, n/2)_u$
B_{2g}	$(-1)^k$	$(-1)^q i^{-\epsilon(n)}$	1	$\psi_\pm(j_{e/o}, n/2)_g$
B_{2u}	$(-1)^k$	$(-1)^q i^{-\epsilon(n)}$	-1	$\psi_\pm(j_{e/o}, n/2)_u$
E_{mg}	$M_m(k)$	$M'^j_m(q)$	1	$[\psi_\pm(j, m), \psi_-(j, m)]_g$
E_{mu}	$M_m(k)$	$M'^j_m(q)$	-1	$[\psi_\pm(j, m), \psi_-(j, m)]_u$

(1) Cf. Table VII.

TABLE IX. The irreps and the bases of D_{np} ($D_{n_o h}$, $D_{n_e d}$)$\simeq D_{2n}$.

D_{np}	C_{2n}^{ke}	C'_{2,q_e}	$\bar{C}_{2n}^{k0}$	$\bar{C}'_{2,q_0}$	Bases
A_1	1	1	1	1	$\psi(j_e, 0)_g, \psi_\pm(j_{e/o}, 2n)_g, \psi_\pm(j_{e/o}, n)_u$
A_2	1	-1	1	-1	$\psi(j_o, 0)_g, \psi_\pm(j_{e/o}, 2n)_g, \psi_\pm(j_{o/e}, n)_u$
B_1	1	-1	-1	1	$\psi_\pm(j_{o/e}, n)_g, \psi(j_o, 0)_u, \psi_\pm(j_{o/e}, 2n)_u$
B_2	1	1	-1	-1	$\psi_\pm(j_{e/o}, n)_g, \psi(j_e, 0)_u, \psi_\pm(j_{e/o}, 2n)_u$
E_m	$M_m(k_e)$	$M'^j_m(q_e)$	$M_m(k_0)$	$M'^j_m(q_0)$	$[\psi_+(j, m), \psi_-(j, m)]_g$
					$[\psi_+(j, m-n), \psi_-(j, m-n)]_u$

(1) Cf. of Table VII.
(2) $m = \frac{1}{2}, 1, \ldots, n - \frac{1}{2}$.
(3) Note that $M_m(k) = M_m(\alpha = \pi k/n)$, $M'^j_m(q) = M'^j_m(\beta = \pi q/n)$.

the bases by harmonic polynomials $Y(l, m)$ which satisfy Condon–Shortley's convention. In this important special case, $\psi_\pm(j, m)$ reduce to

$$Y_+(l, m) = N_{l,m} P_{l-m}(z) \mathrm{Re}(y - ix)^m,$$
$$Y_-(l, m) = N_{l,m} P_{l-m}(z) \mathrm{Im}(y - ix)^m, \tag{3.4}$$

where N_{lm} is the normalization constant, $P_{l-m}(z)$ is a real polynomial of z with degree $l - m$ and Re (Im) denotes the real (imaginary) part.

In Tables II–IX we have classified the complete set of $\psi(j, m)$ according to the irreps of the improper point groups $\bar{G}$ corresponding to C_n and D_n. Here, we have used the similar notations for the symmetry operations and for the irreps as in I as much as possible. It is also noted that the spinor representations given in these tables are the corresponding projective representations whose characters are the class functions of $\bar{G}$. This has been discussed in detail in Sec. 8 of I.

An additional remark is helpful in utilizing these comprehensive sets of the basis functions given in the tables. Frequently, these basis functions can be simplified by further linear combinations of $\psi(j, m)$ with different j or simply using the fact that the variables z and (x, y) do not mix under the operations of the uniaxial or dihedral groups. For example,

according to Table VI, a basis of the irrep $E_{1/2,u}$ of the group D_{ni} is given by a row vector $[\psi_+(\frac{1}{2},\frac{1}{2}), \psi_-(\frac{1}{2},\frac{1}{2})]_u$. Using the vector coupling coefficients[11] one obtains

$$\psi(\tfrac{1}{2}, \tfrac{1}{2})_u \sim z\phi\,(\tfrac{1}{2}, \tfrac{1}{2}) + (x + iy)\phi\,(\tfrac{1}{2}, -\tfrac{1}{2}),$$
$$\psi(\tfrac{1}{2}, \tfrac{1}{2})_u \sim (x - iy)\phi\,(\tfrac{1}{2}, \tfrac{1}{2}) - z\phi\,(\tfrac{1}{2}, -\tfrac{1}{2}). \tag{3.5}$$

Since z and (x, y) do not mix under the operations of D_{ni}, we obtain two simplified basis sets belonging to $E_{1/2,u}$ given as follows:

$$[z\phi_-, -z\phi_+], \quad [x\phi_+ + y\phi_-, -x\phi_- + y\phi_+], \tag{3.6}$$

where $\phi_+ = \phi_+(\frac{1}{2}, \frac{1}{2})$ and $\phi_- = \phi_-(\frac{1}{2}, \frac{1}{2})$. It is noted that the same results could be obtained by the linear combinations of two sets $\psi_\pm(\frac{1}{2}, \frac{1}{2})_u$ and $\psi_\pm(\frac{3}{2}, \frac{1}{2})_u$ belonging to the same irrep.

These basis sets given in Tables II–IX expressed by the elementary basis sets $Y(l, m)$ can be used to construct the SALC's (symmetry adapted linear combinations) of any given set of equivalent basis functions by means of the correspondence theorem developed recently by the author.[12]

[1] S. K. Kim, J. Math. Phys. **22**, 2101 (1981).
[2] H. Weyl, *Group Theory and Quantum Mechanics*, 2nd ed. (Dutton, New York, 1931), p. 146.
[3] B. L. van der Waerden, *Group Theory and Quantum Mechanics* (Springer-Verlag, New York, 1974), p. 97.
[4] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, NJ, 1957), p. 27.
[5] L. Jansen and M. Boon, *Theory of Finite Groups. Applications in Physics* (North-Holland, Amsterdam, 1967), Appendix, p. 341.
[6] A. V. Shubnikov, Sov. Phys. Crystallogr. **4**, 267 (1959).
[7] *International Tables for X-Ray Crystallography*, Vol. 1, *Symmetry Groups* edited by N. F. M. Henry and K. Lonsdale (Kynoch, Birmingham, England, 1969).
[8] C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972).
[9] S. K. Kim, J. Math. Phys. **24**, 411 (1983).
[10] S. K. Kim, J. Math. Phys. **24**, 419 (1983).
[11] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
[12] S. K. Kim, Physica A **105**, 577 (1981); **106**, 521 (1981).

A unified theory of the point groups. IV. The general corepresentations of the crystallographic and noncrystallographic Shubnikov point groups

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

(Received 20 July 1982; accepted for publication 1 October 1982)

A new system of classification for all the Shubnikov point groups G^s is presented, which is best suited for describing their isomorphisms as well as their construction. By the new classification and the modified form of the corepresentations, there are presented the simple general expressions of the irreducible corepresentations for the 10 minimum sets of proper G^s, from which all coirreps of the remaining G^s follow. (Icosahedral groups are excluded.)

PACS numbers: 02.20. + b

1. INTRODUCTION

In a series of papers[1–3] (referred to as I, II, and III) we have developed the representation theory of point groups by regarding their double groups as subgroups of SU(2). In these papers, we have derived the general irreps[1] and the general projective irreps[2] of the double point groups and introduced a new system of classifications of the improper point groups which is best suitable for describing their isomophisms.[3] In the present work we shall construct the general coirreps (irreducible corepresentations[4]) of all the Shubnikov[5–7] (or antiunitary) point groups G^s with an exception of icosahedral groups.

The basic theory of corepresentations has been developed by Wigner.[4] Based on this theory, the coirreps of the crystallographic Shubnikov groups[5–7] (point and/or space groups) are worked out and tabulated by Dimmock and Wheeler,[8] Bradley and Cracknell,[7] and Miller and Love[9] in varying degrees of sophistication. In spite of this, certain simple features of the theory are not well recognized; one is the isomorphism between the proper and improper Shubnikov point groups[10] and the other is a modified form of corepresentation which is more explicit than the usual form[4,7] originally due to Wigner. In order to utilize these features, we shall reconstruct the Shubnikov point groups G^s and introduce a new system of classification which is best suitable for describing the isomorphisms and hence for describing the general coirreps of the infinite set of G^s. By means of this system and the modified form of the corepresentations we shall present the simple general expressions of the irreducible corepresentations for the 10 minimum sets of proper G^s, from which all the coirreps of the remaining G^s will follow.

In Sec. 2, we shall introduce a new construction and classification of G^s and establish the isomorphism between the proper and improper G^s along the line developed in III. In Sec. 3, we shall develop the modified theory of the corepresentations of antiunitary groups following the approach developed in II. Then in Sec. 4, we shall present actual expressions of the coirreps of the minimum 10 Shubnikov point groups mentioned above.

2. A NEW CLASSIFICATION OF THE FULL SHUBNIKOV POINT GROUPS

In order to arrive at a new modified system of classification for the Shubnikov point groups G^s, which is best suited for describing their general coirreps, we shall first discuss the construction. Since we are interested in both the integral and half-integral coirreps, we shall base our argument on the double point groups from the outset as in I. Thus, hereafter, we mean by a point group its double group unless otherwise specified. Let us denote SU(2) by G_s and the improper G_s by G_{si}. Then, a point group is defined by a subgroup of G_{si}.

A Shubnikov point group G^s may be defined by its halving subgroup $H(\leqslant G_{si})$ in the form of a left coset decomposition,

$$G^s = H + a_0 H, \quad a_0 = \theta x; \quad x \in G_{si}, \tag{2.1}$$

where a_0 is an antiunitary operator defined by time inversion θ and a unitary operator x. Following Wigner,[4] we set $\theta = \sigma_y K$ (σ_y the Pauli spin, K the complex conjugation). By definition, θ commutes with all elements of G_{si} and $\theta^2 = \bar{e}$ (2π rotation).

Now, according to Wigner, the corepresentations of G^s are constructed by augmentation of the representations of H with a_0. Thus, it seems best to construct G^s through (2.1) for the present purpose. We first note that, H being a normal subgroup of G^s, x has to satisfy

$$xhx^{-1}, \quad x^2 \in H \quad \text{for all } h \in H. \tag{2.2}$$

On account of this condition, only a very limited number ($\leqslant 6$) of x is allowed for a given H, obviously within a multiplicative factor of $h \in H$. As in the case of improper point groups,[2] the maximum set of x for a given H occurs when $H = C_n$ (uniaxial group). The set may be chosen to be

$$\{x\}_1 = e, \hat{i}, \quad c_{2n}, \bar{c}_{2n} (= \hat{i}c_{2n}), \quad c_2', \bar{c}_2' (= \hat{i}c_2'), \tag{2.3}$$

where e is the identity operator, $\hat{i}$ the inversion, c_{2n} a $2n$-fold rotation axis parallel to $c_n \in C_n$, and c_2' a binary rotation perpendicular to $c_n \in C_n$, and the rest are the rotatory inversion axes. For the remaining point groups with higher symmetries than C_n, only subsets of the maximum set $\{x\}$ of (2.3) are sufficient to provide x. Thus, for example, when $H = D_n$ we have $x = e, \hat{i}, c_{2n}, \bar{c}_{2n}$, and when $H = T, x = e, \hat{i}, c_4, \bar{c}_4$. In this way we can easily construct all G^s through (2.1).

Since the set of compatible pair $\{H,x\}$ uniquely determines G^s, it is most convenient to express G^s by H^x using symbols for x such that

$$H^e, \ H^i, \ H^q, \ H^p, \ H^u, \ H^v, \tag{2.4}$$

where the superscripts correspond to the operations given in

 0022-2488/83/030419-05$02.50

TABLE I. The Shubnikov point groups.

No.	H^X	Shubnikov	International	Isomorphisms	G
1	G_s^i	$\infty/\infty\cdot\underline{m}$	$\infty\,\infty m'$	G_s^e	G_{si}
2	C_∞^i	$\infty:\underline{m}$	∞/m'	C_∞^e	$C_{\infty i}$
3	C_∞^u	$\infty:\underline{2}$	$\infty 2'$	C_∞^u	D_∞
4	C_∞^v	$\infty\cdot\underline{m}$	$\infty m'$	C_∞^u	$C_{\infty v}$
5	$C_{\infty i}^u$	$\underline{m}\cdot\infty:m$	∞/mm'	$C_\infty^u\times C_i$	$D_{\infty i}$
6	$C_{\infty v}^i$	$m\cdot\infty:\underline{m}$	$\infty/m'm$	D_∞^e	$D_{\infty i}$
7	D_∞^i	$\underline{m}\cdot\infty:\underline{m}$	$\infty/m'm'$	D_∞^e	$D_{\infty i}$
8	C_n^i	$\overline{2n_0}, n_e:\underline{m}$	$\bar{n}_0', n_e/m'$	C_n^e	C_{ni}
9	C_n^q	$\underline{2n}$	$(2n)'$	C_n^q	C_{2n}
10	C_n^p	$\overline{2n_e}, \overline{n_0}$	$\overline{2n}'$	C_n^q	C_{np}
11	C_n^u	$n:\underline{2}$	$n_e2'2', n_02'$	C_n^u	D_n
12	C_n^v	$n\cdot\underline{m}$	$n_em'm', n_0m'$	C_n^u	C_{nv}
13	C_{ni}^q	$\underline{2n_e}:m, \underline{2n_0}:m$	$(2n_e)'/m, (2n_0)'/m'$	$C_n^q\times C_i$	C_{2ni}
14	C_{ni}^u	$\underline{m}\cdot n_e:m, \overline{2n_0}\cdot\underline{m}$	$n_e/mm'm', \overline{n_0}m'$	$C_n^u\times C_i$	D_{ni}
15	C_{np}^i	$\underline{2n_e}:\underline{m}, \underline{2n_0}:m$	$(2n_e)'/m', (2n_0)'/m$	C_{2n}^e	C_{2ni}
16	C_{np}^u	$\overline{2n_e}\cdot\underline{m}, \underline{m}\cdot n_0:m$	$\overline{2n_e}2'm', \overline{2n_0}m'2'$	C_{2n}^u	D_{np}
17	C_{nv}^i	$m\cdot n_e:\underline{m}, \overline{2n_0}\cdot m$	$n_e/m'mm, \bar{n}_0'm$	D_n^e	D_{ni}
18	C_{nv}^q	$\underline{2n}\cdot m$	$(2n)'m'm$	D_n^q	C_{2nv}
19	C_{nv}^p	$\overline{2n_e}\cdot m, m\cdot n_0:\underline{m}$	$\overline{2n_e'}2'm, \overline{2n_0'}m2'$	D_n^q	D_{np}
20	D_n^i	$\underline{m}\cdot n_e:\underline{m}, \overline{2n_0}\cdot\underline{m}$	$n_e/m'm'm', \bar{n}_0'm'$	D_n^e	D_{ni}
21	D_n^q	$\underline{2n}:2$	$(2n)'2'2$	D_n^q	D_{2n}
22	D_n^p	$\overline{2n_e}\cdot m, \underline{m}\cdot n_0:\underline{m}$	$\overline{2n_e'}2m', \overline{2n_0'}m'2$	D_n^q	D_{np}
23	D_{ni}^q	$m\cdot\underline{2n_e}:m, m\cdot 2n_0:\underline{m}$	$(2n_e)'/mm'm, (2n_0)'/m'm'm$	$D_n^q\times C_i$	D_{2ni}
24	D_{np}^i	$m\cdot\underline{2n_e}:\underline{m}, m\cdot\underline{2n_0}:m$	$(2n_e)'/m'm'm, (2n_0)'/mm'm$	D_{2n}^e	D_{2ni}
25	T^i	$\bar{6}/2$	$m'3$	T^e	T_i
26	T^q	$3/\underline{4}$	$4'32'$	T^q	O
27	T^p	$3/\underline{\bar{4}}$	$\bar{4}'3m'$	T^q	T_p
28	T_i^q	$\underline{\bar{6}}/4$	$m3m'$	$T^q\times C_i$	O_i
29	T_p^i	$\underline{\bar{6}}/4$	$m'3m$	O^e	O_i
30	O^i	$\underline{\bar{6}}/4$	$m'3m'$	O^e	O_i
31	Y^i	$3/\underline{\overline{10}}$	$5'3m'$	Y^e	Y_i

(1) The grey groups H^e are not listed.
(2) $n_e(n_0)$ denotes even (odd) n; thus, equivalence of the notations mean that, e.g., $C_{n_0}^i \leftrightarrow \overline{2n_0}$, $C_{n_e}^i \leftrightarrow n_e:\underline{m}$ for No. 8.
(3) $n>1$ for Nos. 17–24.

(2.3) in that order, respectively. We also express the unitary point group H by its symbol introduced in III; e.g., the improper point groups corresponding to a proper point group P augmented by $\hat{i}$, $\bar{c}_{2n}$, and $\bar{c}_2'$ are expressed by P_i, P_p, and P_v, respectively. Using these symbols we have presented all the Shubnikov point groups in Table I (except for the grey group[7] H^e which occurs for every H) and compared with the corresponding Shubnikov and international symbols (see Ref. 6, p. 142). It is noted that the present system of symbols classifies whole G^s into 44 sets, of which 31 sets belong to the Shubnikov magnetic (or black and white) point groups and 13 sets belong to the grey groups.

Now, we shall describe the isomorphism between the improper and proper Shubnikov point groups. The present system of symbols is introduced for this purpose. Making the one-to-one correspondence by

$$\theta\hat{i}\leftrightarrow\theta,\quad \bar{c}_{2n}\leftrightarrow c_{2n},\quad \bar{c}_2'\leftrightarrow c_2', \tag{2.5}$$

we obtain the following isomorphisms,

$$H^i\simeq H^e,\quad H^p\simeq H^q,\quad H^v\simeq H^u, \tag{2.6a}$$

$$C_{np}^x\simeq C_{2n}^x,\quad D_{np}^x\simeq D_{2n}^x,\quad T_p^x\simeq O^x,\quad C_{nv}^x\simeq D_n^x, \tag{2.6b}$$

where x stands for an operator compatible with correspond-

ing H and the symbols for improper point groups are defined in III. Moreover, when $H = P_i$, i.e., H is a direct product of a proper point group P and the group of inversion C_i, we have

$$P_i^x = P^x \times C_i. \tag{2.6c}$$

Thus, from (2.6) we can obtain the coirreps of the improper G^s from those of the corresponding proper G^s, which can be expressed by one of the following general forms,

$$P^e, \quad P^q, \quad P^u. \tag{2.7}$$

These are explicitly given in the fifth column of Table I. It should be noted that out of the total 44 sets of G^s, only 13 sets belong to the proper G^s given in (2.7). Thus, the isomorphisms reduce the labor of constructing the coirreps roughly to its one fourth. As it will be seen in the next section, all the coirreps of P^u are trivially written down being one dimensional. Thus, we will literally be left with only P^e and P^q. There exist further isomorphisms[3] which hold for the single Shubnikov point groups. These, naturally, do not help to construct their spinor representations.

We shall make an additional remark on the ordinary method[7,8] of constructing G^s. Let G be a unitary group with a halving subgroup H. Then, G^s is constructed through $G^s = H + \theta(G - H)$ except for its grey group. Obviously, this method is equivalent to the present one since the left coset decomposition yields

$$G = H + xH. \tag{2.8}$$

For the sake of comparison, we have also listed G in the last column of Table I. As one can see from the table, the classification by the Schönflies[8] notation $G(H)$ or by the Shubnikov or the international symbols is not really effective in describing the isomorphisms (see also III). Precisely, this situation made us introduce the present system of symbols for G^s.

3. GENERAL CONSTRUCTION OF THE COIRREPS OF G^s

As is mentioned in the Introduction, the basic theory of corepresentations of the antiunitary groups has been worked out by Wigner.[4] We shall present the basic coirreps in a form which is more explicit than the original form of Wigner following the approach developed in II for constructing the projective irreps. This approach makes the theory more transparent and also simplifies the explicit presentation of the final results.

Let us assume that there is given a complete set of unitary irreps $\{\Delta^\nu(h)\}$ of h belonging to the halving subgroup H of G^s, and let ψ^ν be a basis row vector belonging to Δ^ν. Then we have

$$h(a_0\psi^\nu) = (a_0\psi^\nu)\Delta^\nu(a_0^{-1}ha_0)^*. \tag{3.1}$$

Since $\Delta^\nu(a_0^{-1}ha_0)^*$ is an irrep of H, it must be equivalent to one of the irreps of H, say, the μth irrep Δ^μ. Then there exists a unitary matrix $N(a_0)$ such that

$$\Delta^\nu(a_0^{-1}ha_0)^* = N(a_0)^{-1}\Delta^\mu(h)N(a_0), \tag{3.2a}$$

$$a_0\psi^\nu = \phi^\mu N(a_0), \tag{3.2b}$$

where ϕ^μ is a basis belonging to Δ^μ. Accordingly, the set of row vectors $[\psi^\nu, \phi^\mu]$ defines a corepresentation $D^{(\nu,\mu)}$ of G^s as follows:

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

$$D^{(\nu,\mu)}(a_0) = \begin{bmatrix} 0 & \Delta^\nu(a_0^2)N(a_0)^{-1*} \\ N(a_0) & 0 \end{bmatrix}. \tag{3.3}$$

There exist three cases for $D^{(\nu,\mu)}$. It is customary to discuss the third case first[4]:

Case (*c*): $\nu \neq \mu$. The corepresentation $D^{(\nu,\mu)}$ is irreducible.

Case (*b*): $\nu = \mu$, $N(a_0)N(a_0)^* = -\Delta^\nu(a_0^2)$.
is simplified to

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

$$D^{(\nu,\nu)}(a_0) = \begin{bmatrix} 0 & -N(a_0) \\ N(a_0) & 0 \end{bmatrix} \tag{3.4}$$

which is again irreducible. Hence, ψ^ν and ϕ^ν are linearly independent unless they are null.

Case (*a*): $\nu = \mu$, $N(a_0)N(a_0)^* = \Delta^\nu(a_0^2)$. The corepresentation $D^{(\nu,\mu)}$ is reduced to two coirreps $D^{(\nu+)}$ and $D^{(\nu-)}$ given by

$$D^{(\nu\pm)}(h) = \Delta^\nu(h), \quad D^{(\nu\pm)}(a_0) = \pm N(a_0). \tag{3.5}$$

Either one of them provides the required coirrep since $D(a_0)$ has an arbitrary phase factor for a given $D(h)$. The corresponding basis sets are given by

$$\psi^{(\nu\pm)} = \psi^\nu \pm (a_0\psi^\nu)N(a_0)^{-1} \tag{3.6a}$$

which satisfy, as they should,

$$a_0\psi^{(\nu\pm)} = \psi^{(\nu\pm)}D^{(\nu\pm)}(a_0). \tag{3.6b}$$

At least, one of $\psi^{(\nu+)}$ and $\psi^{(\nu-)}$ is not null unless ψ^ν is null.

The corepresentations given by (3.3)–(3.5) are basic to all the coirreps of G^s. These are completely explicit in terms of the irreps of the halving subgroup H if the transformation matrix $N(a_0)$ is known. Wigner's form of the coirrep which corresponds to $D^{(\nu,\mu)}$ of (3.3) is different from the present form since it has $\Delta^\nu(a_0^{-1}ha_0)^*$ in the place of $\Delta^\mu(h)$ and $N(a_0) = 1$ for $\nu \neq \mu$. Thus, it does not identify Δ^μ and requires further calculation of $\Delta^\nu(a_0^{-1}ha_0)^*$ for all h, while the present form requires calculation of single matrix $N(a_0)$ for most cases of practical applications; this may be seen from that the new basis ϕ^μ is given by $(a_0\psi^\nu)N(a_0)^{-1}$.

Now, when Δ^ν is one dimensional, one may take $N(a_0) = 1$ adjusting the phase factor. In the case of higher dimensions, it is determined through (3.2a) or (3.2b). We shall calculate $N(a_0)$ through (3.2b) with use of the spinor bases of the point groups given in I. On account of the isomorphisms of G^s discussed in Sec. 2, one needs $N(a_0)$ only for $a_0 = \theta$, θc_{2n}, and $\theta c_2'$. There is no need to introduce many particle operators for θ since we base our calculation on the one-particle spinor basis: The representations ± 1 of $\theta^2 (= \bar{e})$ simply follow from the dimensions of the representations.

Following the notation given in I, let $\{\phi(j,m)\}$ be a set of $2j$ monomials of two elementary spinors ξ_1 and ξ_2 belonging to the irrep $D^{(j)}$ of G_s. If we adapt the phase convention

$$\theta\phi(j,m) = (-1)^{j-m}\phi(j,-m), \tag{3.7}$$

following Kramers and Wigner,[4] and Condon–Shortley,[11] then we have

$$\theta\hat{c}_{2n}\phi(j,m) = (-1)^{j-m}\exp(i\pi m/n)\phi(j,-m), \tag{3.8}$$

$$\theta\hat{c}_2'\phi(j,m) = i^{-2m}\phi(j,m), \tag{3.9}$$

where use have been made of the spinor representations of c_{2n} and c_2' assuming c_{2n} and c_2' are parallel to the z and x axes, respectively. From these relations (3.7)–(3.9), one can calculate $N(a_0)$ for the proper Shubnikov point groups P^e, P^q, and P^u.

Before proceeding further, we shall exemplify the present approach by calculating the coirreps of the grey group G_s^e and those of P^u from (3.7) and (3.9), respectively. If one compares (3.7) with (3.6b) one immediately concludes that all the coirreps of G_s^e belong to case (a) and are given by

$$D^{(j)}(\theta)_{nm} = N(\theta)_{nm} = (-1)^{j-m}\delta(n,-m)$$
$$n,m = j, j-1,\ldots,-j. \tag{3.10}$$

This was first derived by Wigner[4] through matrix transformation (3.2a) with ingeneous argument.

Next, to calculate the coirreps of P^u, it is noted that the only groups contained in P^u are C_n^u and C_∞^u. Since the irreps of a uniaxial group are all one dimensional and their spinor bases are given by $\phi(j,m)$ where m identifies the irrep; comparison of (3.9) and (3.6b) yields that all coirreps of P^u are one dimensional belonging to case (a). Thus, one can take $N(\theta c_2') = 1$ for all P^u.

To construct the coirreps of P^e and P^q, it is necessary to discuss the transformation properties of spinor bases $\phi_\pm(j,m)$ introduced in I. For the present purpose, however, it is most convenient to redefine $\phi_\pm(j,m)$ with an extra phase factor

$$\phi_+(j,m) = 2^{-1/2}[\phi(j,m) + \phi(j,-m)]i^{j+m},$$
$$\phi_-(j,m) = 2^{-1/2}i^{-1}[\phi(j,m) - \phi(j,-m)]i^{j+m}. \tag{3.11}$$

Then, we have

$$\theta\phi_\pm(j,m) = \phi_\pm(j,m) \quad \text{for integral } m,$$
$$\theta\phi_\pm(j,m) = \pm i\phi_\mp(j,m) \quad \text{for half-integral } m. \tag{3.12}$$

Thus, for a two-dimensional basis $\Phi(j,m) = [\phi_+(j,m),\phi_-(j,m)]$ belonging to the irrep E_m of dihedral groups given in I, we have

$$\theta\Phi(j,m) = \Phi(j,m)Y_m,$$
$$\theta\hat{c}_{2n}\Phi(j,m) = \Phi(j,m)Z_m, \tag{3.13a}$$

where $Y_m = 1_2$ or σ_y depending on m integer or half-integer, 1_2, the 2×2 unit matrix, σ_y the Pauli spin, and

$$Z_m = Y_m\begin{bmatrix}\cos(\pi m/n) & -\sin(\pi m/n)\\ \sin(\pi m/n) & \cos(\pi m/n)\end{bmatrix}. \tag{3.13b}$$

Analogously, one can calculate $N(a_0)$ for the higher dimensional representations.

With the preparations given above we can calculate all the coirreps of P^e and P^q. We shall, however, present them in the next section together with all the other results after introducing a suitable set of notations for the coirreps.

4. THE COIRREPS OF THE SHUBNIKOV POINT GROUPS

We shall now present all coirreps of the proper Shubnikov point groups in terms of the irreps Δ^ν of the halving subgroups and $N(a_0)$. We shall use the following notations

$$D(\Delta^\nu,\Delta^\mu;N(a_0)) \quad \text{for } D^{(\nu,\mu)} \text{ of (3.3)}\in\text{Case (c)},$$
$$D(\Delta^\nu,\Delta^\nu;N(a_0)) \quad \text{for } D^{(\nu,\nu)} \text{ of (3.4)}\in\text{Case (b)}, \tag{4.1}$$
$$D(\Delta^\nu;N(a_0)) \quad \text{for } D^{(\nu+)} \text{ of (3.5)}\in\text{Case (a)}.$$

For simplicity, however, we shall not list the trivial cases of one-dimensional coirreps which can be expressed by $D(\Delta^\nu;1)$. Thus, the coirreps of P^u will not be given since all of them belong to the trivial one-dimensional case. Then in view of Table I, it is only necessary to give the coirreps of the 10 sets of the proper G^s,

$$G_s^e;\quad C_\infty^e, C_n^e, C_n^q;\quad D_\infty^e, D_n^e, D_n^q;\quad T^e, T^q;\quad O^e, \tag{4.2}$$

of which the coirreps of G_s^e are already given by (3.10). It is also noted that Y^u is simply excluded.

In the following, the coirreps of these nine sets of the proper G^s with the dimensions higher than one are given together with the irreps of the respective halving subgroups given in I:

1. C_∞; $M_0, M_{\pm m}$; $m = \frac{1}{2},1,\ldots,\infty$, (4.3a)

   C_∞^e; $D(M_m, M_{-m};1)$. (4.3b)

2. C_n; $M_0, M_{n/2}, M_{\pm m}$; $m = \frac{1}{2},1,\ldots,\frac{1}{2}(n-1)$, (4.4a)

   C_n^e; $D(M_{n/2}, M_{n/2};1)_{n=\text{odd}}, D(M_m, M_{-m};1)$, (4.4b)

   C_n^q; $D(M_{n/2}, M_{n/2};1)_{n=\text{even}}, D(M_m, M_{-m};1)$. (4.4c)

3. D_∞; A_1A_2, E_m; $m = \frac{1}{2},1,\ldots,\infty$, (4.5a)

   D_∞^e; $D(E_m;Y_m)$; for Y_m see Eq. (3.13). (4.5b)

4. D_n; A_1, A_2, B_1, B_2, E_m; $m = \frac{1}{2},1,\ldots,\frac{1}{2}(n-1)$, (4.6a)

   D_n^e; $D(B_1, B_2;1)_{n=\text{odd}}, D(E_m;Y_m)$, (4.6b)

   D_n^q; $D(B_1, B_2;1)_{n=\text{even}}, D(E_m;Z_m)$, (4.6c)

   where Y_m and Z_m are given in Eq. (3.13).

5. T; $A, B_1, B_2, T, E_{1/2}, E_{1/2}' = B_1\times E_{1/2}$,
   $E_{1/2}'' = B_2\times E_{1/2}$, (4.7a)

   T^e; $D(B_1, B_2;1), D(E_{1/2};\sigma_y), D(E_{1/2}', E_{1/2}'';\sigma_y)$, (4.7b)

   T^q; $D(T;\hat{c}_4^z), D(E_{1/2};Z), D(E_{1/2}';Z), D(E_{1/2}'';Z)$, (4.7c)

   where[12]

$$\hat{c}_4^z = \begin{bmatrix}0 & -1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix},\quad Z = 2^{-1/2}\begin{bmatrix}-i & -i\\ i & -i\end{bmatrix}. \tag{4.7d}$$

6. O; $A_1, A_2, E, T_1, T_2, E_{1/2}$,
   $E_{1/2}' = A_2\times E_{1/2}, Q = E\times E_{1/2}$ (4.8a)

   O^e; $D(E;1_2), D(T_1;1_3), D(T_2;1_3)$,
   $D(E_{1/2};\sigma_y), D(E_{1/2}';\sigma_y), D(Q;\Sigma_y)$, (4.8b)

where Σ_y is the 4×4 Dirac spin given by $\sigma_y \oplus \sigma_y$.

It is interesting to note that the first coirreps of $C^e_{n_0}$ and $C^q_{n_e}$ given in (4.4b) and (4.4c), respectively, are the rare ones belonging to case (b). It is also noted that all the coirreps of C^q_1, G^e_s, C^u_∞, C^u_n, $D^e_{n_e}$, $D^q_{n_0}$, D^e_∞, T^q, and O^e belong solely to case (a). One can immediately draw similar conclusions for improper G^s through the isomorphisms given by (2.6).

5. CONCLUDING REMARKS

It has been shown that the new classification of the full Shubnikov point groups G^s based on the halving subgroups H and the augmenting operators θx are very effective in constructing G^s as well as describing their coirreps. By establishing the isomorphisms between the proper and improper G^s we are able to describe the coirreps of all G^s by those of 10 minimum sets of the proper G^s given in (4.2). Since the basic form of the coirreps introduced in (3.3) is more explicit than the usual form originally due to Wigner, we are able to describe their coirreps by simple expressions given in (4.3)–(4.8) and (3.10). These are completely explicit in terms of the irreps of the halving subgroups. As is well known there exist 90 Shubnikov crystallographic point groups. In the usual presentations,[7,8] each G^s is separately tabulated. These are easily reproduced with use of the simple general expressions given in (4.3)–(4.8) based on the present general classification.

The form of the coirreps given in the present work is particularly suitable for constructing the symmetry adapted functions belonging to the coirreps; these functions are calculated from the symmetry adapted functions belonging to the irreps of the halving subgroups by (3.2b) and (3.6a). There is no need of introducing the cumbersome projective operators of the coirreps of G^s to construct the symmetry adapted functions.

The present work can be extended to calculate the projective coirreps of G^s, based on the projective irreps of the point groups introduced in II. This then enables us to write down the coirreps of the Shubnikov space groups. This extension, however, will be discussed somewhere else.

[1] S. K. Kim, J. Math. Phys. **22**, 2101 (1981); see also Mol. Phys. **43**, 1447 (1981).
[2] S. K. Kim, J. Math. Phys. **24**, 411 (1983).
[3] S. K. Kim, J. Math. Phys. **24**, 414 (1983).
[4] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
[5] A. V. Shubnikov, Sov. Phys. Crystall. **6**, 394 (1962).
[6] A. V. Shubnikov, N. V. Belov *et al.*, *Colored Symmetry*, translated from Russian by J. Itzkoff and J. Gollob (Pergamon, New York, 1964).
[7] C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972).
[8] J. O. Dimmock and R. G. Wheeler, J. Phys. Chem. Solids **23**, 729 (1962).
[9] S. C. Miller and W. F. Love, *Tables of Irreducible Representations of Space Groups and Corepresentations of Magnetic Groups* (Pruett, Boulder, Colorado, 1967).
[10] These may be distinguished by whether or not their operations involve improper point operations.
[11] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1957).
[12] It is noted here that the basis of $E_{1/2}$ is taken to be $\Phi(\frac{1}{2},\frac{1}{2})$ of Eq. (3.13a) which is different from the one given in Table IV of I. If one uses the latter basis, we have $Z = i2^{-1/2}(\sigma_x - \sigma_y)$. The results given in (4.8) hold for both bases of $E_{1/2}$.

Phase conventions consistent with duality between SUn and S_L

John J. Sullivan
Department of Physics, University of New Orleans, New Orleans, Louisiana 70148

(Received 8 February 1980; accepted for publication 4 June 1982)

The Schur–Weyl construction of irreducible tensors by symmetrized L th rank exterior products of a defining n-dimensional vector space establishes a duality between the coupling algebra of the unitary group Un and the symmetric group S_L. The coupling coefficients shared by these two groups are real and possess symmetries due to complex conjugation in Un and association in S_L. The unitary group is factored by its unitary unimodular subgroup SU$n \simeq$ Un/U1 so it is usual to identify the coupling algebras of Un and SUn by choosing the trivial (phaseless) coupling for U1. This establishes a set of equivalence relations since all pseudoscalar irreducible representations (irreps) of Un subduce onto the scalar irrep of SUn. The question remains can this trivial identification with the coupling algebra of SUn be made consistent with the symmetries which follow from duality? Specifically the symmetries which follow from duality must be consistent with the equivalence relations. We argue such consistency may be obtained. The resulting symmetries place significant restrictions on the resolution of multiplicities and the choice of phase conventions for transformations which in the absence of duality are normally considered independently. We review the usual formulation of a coupling algebra and the identifications which follow from duality using a double coset decomposition of S_L. Particular attention is given to the phase transformation introduced by a transposition in order of the two component irreps being coupled. Phase transformations needed for association and complex conjugation are examined to determine the symmetries they must exhibit to be consistent with duality. Two phase transformations needed for complex conjugation, the Derome–Sharp matrix and the 1jm factor, are shown to be related by association. A transposition phase convention and an association phase convention are proposed and shown to be consistent with all the symmetries required by duality.

PACS numbers: 02.20. + b, 03.65.Fd, 11.30.Jw

I. DUALITY AND THE COUPLING ALGEBRA

Schur and Weyl[1] established the fundamental duality between the integral irreps of the general linear group Gln with its unitary Un and unitary unimodular SUn subgroups and the representation theory of the symmetric group S_L. The generation of a complete vector space for the representation of any compact or finite group by taking symmetrized products of a finite-dimensional vector defining space (Stone's version of the Peter–Weyl theorem[2]) expresses this same duality. Inner products of Un and SUn are dual to outer products of S_L and vice versa. In a series of papers,[3] using double coset (DC) decomposition of S_L, we have developed the consequences of this duality for the generalized Racah coupling algebra of Un. The most important consequences have been (a) the identification of recoupling coefficients of Un with double coset matrix elements (DCME) of S_L, (b) the identification of isoscalar factors of Un with weighted double coset matrix elements (WDCME) of S_L, and (c) the identification of DCME of S_L with DCME of Un. The theory of S_L implies these elements in suitable bases may be chosen real and possess the symmetry of association with respect to (w.r.t.) the alternating group A_L. The theory of Un implies these elements possess symmetry under the operation of complex conjugation in this group. Considering either group alone the symmetries which follow from the other group are surprising not only with regard to magnitudes but also with regard to phase. Iteration procedures via generalized back coupling relations have been developed for evaluating a generalized DCME. However, these relations require definite phases be assigned at each step.

The unitary group Un possesses a denumerably infinite set of pseudoscalar irreps because the unit matrix times a phase generates the normal abelian subgroup U1. One therefore considers the unimodular subgroup or more exactly SU$n/Cn \simeq$ Un/U1, where Cn is the cyclic group on n items represented by the nth roots of unity times the identity matrix. The irreps of Un remain irreducible on restriction to SUn but one obtains the symmetry of equivalence due to the fact all pseudoscalar irreps of Un map to the scalar irrep of SUn. Since pseudoscalar irreps are one-dimensional these equivalences require further relations must hold for the magnitudes of the DCME and WDCME. These relations although surprising from the point of view of S_L are readily verified in any specific evaluation of DCME or WDCME. One naturally wishes to choose the trivial coupling algebra (i.e., + 1) for any abelian group and thus identify the coupling algebra of Un and SUn. The question remains can one resolve multiplicities and establish a phase convention that is consistent with the consequences of duality (i.e., the choice of real coupling coefficients evidencing symmetries due to association, complex conjugation, *and* equivalence)? We answer in the affirmative, first arguing that such a choice is possible and then proposing a specific phase convention which is compatible with duality. Our argument proceeds by examining all possibilities. Phase and multiplicity resolution

 0022-2488/83/030424-17$02.50

are established within a triad class giving the outer product prescription for coupling two component irreps to form a third. Equivalence relations fix equivalent triads into a class within which multiplicity is resolved in the same way and phases are fixed in definite relations. Association and complex conjugation relate in general distinct triad classes but with the same range of multiplicities. We argue the only cases in which there is a question about the possibility of arranging a phase convention consistent with duality is when these classes are identical. The triad classes under these identities separate into three cases. A detailed examination of these cases is made using the proposed phase convention of Sec. VII showing compatibility with the symmetries of duality.

Phases for coefficients involving several triads, such as the DCME, are simple products of phases due to each constituent triad. The back-coupling relations are consistent with any phase convention which is essentially a product of phases from the constituent triads. They also preserve parity and are consistent with the symmetries of equivalence. Therefore any DCME or WDCME evaluated with their use will also be consistent with the symmetries of duality. While the phase convention we propose is used in our argument to exhaust the possibilities in the special cases and there are substantial restrictions placed on a phase convention by duality, we make no claim that the convention proposed here is unique. The proposed phase convention is consistent with the usual choice of positive phase for maximum coupling and does reduce to the Condon and Shortley phase for SU2 when consideration is restricted to bipartition irreps.

Because of the many interrelationships considered notation has been a vexing problem. In general Greek letters will indicate an irrep (λ) or a phase (θ,ϕ,Φ), while Latin letters will indicate a general basis member or multiplicity label. For the symmetric group we will use $\lambda \vdash L$ ($\vdash$ meaning a partition of), $\nu \vdash 3$, and $\sigma \vdash 2$. The product symbols $\times, \circ, \cdot, \otimes, \odot$ will be used to designate direct, inner, outer (often omitted), and symmetrized outer and inner products, respectively. Straight line brackets will indicate the positive integer value (i.e., dimension) of the quantity bracketed, e.g., $|G|$ is the group order or volume, $|\lambda|$ the irrep dimension, $|\lambda;\lambda_1\lambda_2|$ the outer product frequency, etc. However, in order to distinguish an irrep dimension in S_L from that in Un we will use $|\lambda|$ for the former and $\langle\lambda\rangle_n$ for the latter. Summation on repeated Latin indices will be assumed, while sums on Greek indices will be explicit. The placing of the multiplicity label within a triad has proved a bothersome point. Subgroup sequencing would suggest the placement $(\lambda:r\lambda_1\lambda_2)$, which placement make their role in the orthogonality of a DCME

$$\begin{bmatrix} \lambda & r & \lambda_1 & \lambda_2 \\ r' & & r_1 & r_2 \\ {}_1\lambda & {}_1r & {}_1\lambda_1 & {}_1\lambda_2 \\ {}_2\lambda & {}_2r & {}_2\lambda_1 & {}_2\lambda_2 \end{bmatrix}$$

(or a WDCME) more obvious, but such symbols appear cumbersome. So we have elected to use a posterior placement of the multiplicity w.r.t. its triad. Also a pre- or post index i or j will be assumed to range over two values unless specifically indicated that a range of three or more applies. Thus the above DCME is often rendered as

$$\begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix}^{r}_{r'} \quad {}_{,r} \quad {}_{r_j}$$

and where multiplicity labels are not crucial to the argument they will be dropped but their presence should always be implied.

Section II reviews the coupling algebra of Un in which we mainly follow Butler's[4] paper. Particular attention is given to questions of phase due to definitions, permutations, and complex conjugation. Double coset decomposition of the symmetric group is reviewed in Sec. III. Most of the considerations given here are applicable to any group. General group theoretical considerations place restrictions on the matrices representing involutive operations such as association, complex conjugation, and transpositions. These are given in Sec. IV. Section V reviews the identifications of duality and presents the symmetries which must hold if duality is to apply to the coupling algebra of SUn. Section VI argues that a resolution of multiplicities and a phase convention consistent with duality can be obtained. The properties a phase convention must have in order to be consistent with duality are specified. A phase convention for permutation and association is prescribed in Sec. VII and is shown to have the properties required in the previous section.

II. THE RACAH ALGEBRA OF Un

Following Butler[4] with some change in notation and convention we review the coupling algebra of Un. The formulation is sufficiently general to be applicable to any finite or compact group. The Wigner (Clebsch–Gordan or vector coupling) coefficients are elements of the unitary matrix that reduces the Kronecker product w.r.t. a group G being defined (up to a unitary transformation of the multiplicity index) by

$$\begin{bmatrix} \lambda_1 & g \\ M_1 & M_1' \end{bmatrix}\begin{bmatrix} \lambda_2 & g \\ M_2 & M_2' \end{bmatrix} = \sum_{\lambda_3} \langle \lambda_1 M_1 \cdot \lambda_2 M_2 | r\lambda_3 M_3 \rangle \times \begin{bmatrix} \lambda_3 & g \\ M_3 & M_3' \end{bmatrix} \langle r\lambda_3 M_3' | \lambda_1 M_1' \cdot \lambda_2 M_2' \rangle, \tag{2.1}$$

with

$$\langle r\lambda_3 M_3 | \lambda_1 M_1 \cdot \lambda_2 M_2 \rangle^* = \langle \lambda_1 M_1 \cdot \lambda_2 M_2 | r\lambda_3 M_3 \rangle.$$

The more symmetric $3jm$ symbol is similarly defined but using the complex conjugate irrep so that group orthogonality allows the equivalent definition

$$\int \begin{bmatrix} \lambda_1 & g \\ M_1 & M_1' \end{bmatrix}\begin{bmatrix} \lambda_2 & g \\ M_2 & M_2' \end{bmatrix}\begin{bmatrix} \lambda_3 & g \\ M_3 & M_3' \end{bmatrix} \frac{dg}{|G|} = \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ M_1 & M_2 & M_3 \end{pmatrix}^{r^*} \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ M_1' & M_2' & M_3' \end{pmatrix}^{r}. \tag{2.2}$$

While (2.1) is invariant under the transposition $1 \leftrightarrow 2$, the l.h.s. of (2.2) is invariant under S_3, the permutations in order of the three matrix elements. The r.h.s. has the form of a

scalar product over the multiplicity index. Therefore the multiplicity index may be adapted to correspond to the action of S_3 on the $3jm$ symbol. In all but one case the permutation matrix may be chosen diagonal in the multiplicity index as $\phi(\lambda_1\lambda_2\lambda_3 r)^p \delta rr'$, where $\phi(\lambda_1\lambda_2\lambda_3 r) = \pm 1$ and p is the parity of the permutation. In case two or more component irreps are equivalent Derome[5] has shown the representation theory of the symmetric group fixes ϕ as

$$\begin{aligned}&(a)\phi = +1 \text{ if } \lambda_1 \simeq \lambda_2 \not\simeq \lambda_3 \text{ and } 0 \in \lambda_3 \cdot (\lambda_1 \otimes [2])\\ &\quad \text{or } \lambda_1 \simeq \lambda_2 \simeq \lambda_3 \text{ and } 0 \in \lambda_1 \otimes [3],\\ &(b)\phi = -1 \text{ if } \lambda_1 \simeq \lambda_2 \not\simeq \lambda_3 \text{ and } 0 \in \lambda_3 \cdot (\lambda_1 \otimes [1^2])\\ &\quad \text{or } \lambda_1 \simeq \lambda_2 \simeq \lambda \simeq \lambda_3 \text{ and } 0 \in \lambda_1 \otimes [1^3],\end{aligned} \tag{2.3}$$

where 0 indicates the scalar irrep of the group. The only exception to diagonalization of the permutation phase matrix occurs for $\lambda_1 \simeq \lambda_2 \simeq \lambda_3$ and $0 \in \lambda_1 \otimes [2,1]$, in which case one must use for the permutational phase matrices the two by two matrices of the irrep $[2,1]$ of S_3 in any of their standard forms.

One of the component irreps in (2.2) being the scalar irrep defines the $1jm$ symbol relating the complex conjugate of an irrep to the complex conjugate irrep

$$\begin{bmatrix}\lambda^* & g\\ M^* & M'^*\end{bmatrix} = |\lambda| \begin{pmatrix}\lambda^* & 0 & \lambda\\ M^* & 0 & M\end{pmatrix}^* \begin{bmatrix}\lambda & g\\ M & M'\end{bmatrix}^* \begin{pmatrix}\lambda^* & 0 & \lambda\\ M'^* & 0 & M'\end{pmatrix}. \tag{2.4}$$

We may consider the $1jm$ symbol in real form and abbreviate to

$$|\lambda|^{1/2} \begin{pmatrix}\lambda^* & 0 & \lambda\\ M^* & 0 & M\end{pmatrix} \equiv (\lambda)_{MM^*} = \phi(\lambda)(\lambda^*)_{M^*M}. \tag{2.5}$$

This is just the Schur–Frobenius[6] result, where $\phi(\lambda)$ is the $1j$ phase satisfying $\phi(\lambda^*) = \phi(\lambda)^*$. The $3jm$ symbol with complex conjugate irreps and the complex conjugate $3jm$ symbol are related by

$$\begin{aligned}\begin{pmatrix}\lambda_1^* & \lambda_2^* & \lambda_3^*\\ M_1^* & M_2^* & M_3^*\end{pmatrix}^r = \begin{pmatrix}\lambda_1 & \lambda_2 & \lambda_3\\ M_1 & M_2 & M_3\end{pmatrix}^{r^*} (\lambda_1)_{M_1M_1^*}\\ \times (\lambda_2)_{M_2M_2^*} (\lambda_3)_{M_3M_3^*} A(\lambda_1\lambda_2\lambda_3)_{rr^*}.\end{aligned} \tag{2.6}$$

The unitary matrix A is the Derome–Sharp[7] matrix which because complex conjugaton is involutive has the symmetry

$$A(\lambda_1^*\lambda_2^*\lambda_3^*)_{r^*r} = \phi(\lambda_1)\phi(\lambda_2)\phi(\lambda_3) A(\lambda_1\lambda_2\lambda_3)_{rr^*}. \tag{2.7}$$

For most groups, including Un and SUn, Butler shows $\phi(\lambda_1)\phi(\lambda_2)\phi(\lambda_3) = +1$ for any triad. Because complex conjugation and permutation in order commute one can show A is block diagonal in the permutational symmetry types discussed above and independent of the order of the constituent irreps; i.e.,

$$A(\lambda_1\lambda_2\lambda_3)_{rr^*} = A(\lambda_a\lambda_b\lambda_c)_{rr^*}, \tag{2.8}$$

where r and r^* are of the same permutational type. Comparison of (2.1) to (2.2) using (2.4) shows the Wigner coefficient can be identified with the corresponding $3jm$ symbol by

$$\begin{aligned}&\begin{pmatrix}\lambda_1 & \lambda_2 & \lambda^*\\ M_1 & M_2 & M^*\end{pmatrix}^r\\ &\quad = \langle 0|\lambda^* M^* \cdot \lambda M\rangle \langle r\lambda M|\lambda_1 M_1 \cdot \lambda_2 M_2\rangle \theta(\lambda_1\lambda_2\lambda^* r)\\ &\quad = \frac{(\lambda^*)_{M^*M}}{|\lambda|^{1/2}} \langle r\lambda M|\lambda_1 M_1 \cdot \lambda_2 M_2\rangle \theta(\lambda_1\lambda_2\lambda^* r).\end{aligned} \tag{2.9}$$

The intermediate equality simply indicates the $3jm$ symbol can be considered to be (up to a phase) the vector coupling coefficient combining three component irreps in two successive pairwise steps to form the scalar irrep. Equation (2.9) fixes a standard coupling sequence of these pairwise steps for relating the $3jm$ symbol to the Wigner coefficient. The phase $\theta(\lambda_1\lambda_2\lambda^* r)$ is arbitrary and could be set equal to 1, which Butler adopts as the "sensible" choice. However, to correspond to choices in SU2 that have historical precedence[8] and to allow certain generalizations of these conventions it must be retained. For the groups we consider it is possible to uniquely establish a maximum coupling for which $\lambda = \lambda_1 + \lambda_2$. A generalization of the Condon and Shortley[9] phase which we adopt requires a positive phase for maximum coupling, in particular $\langle \lambda M|\lambda M \cdot 0\rangle = +1 = \langle \lambda M|0 \cdot \lambda M\rangle$. This convention *requires* an appropriate choice of the phase θ. The choice $\theta(\lambda_1\lambda_2\lambda_3 r) = \phi(\lambda_1)$ produces the simplest phase relations consistent with this convention and the phase rule we propose in Sec. VII, so we adopt that choice in all which follows. An additional phase $\theta(\lambda_1\lambda_2\lambda^*)$ has to be included in $\theta(\lambda_1\lambda_2\lambda^*)$ in order for our proposed phase convention to agree in SU2 with the Condon and Shortley phase for relating a $3jm$ symbol to the Wigner coefficient [viz. Ref. 8, p. 113 or Eq. (3.3), p. 39]. Such an additional phase depending on all three irreps of a triad has no effect on the relations given here and will effect only the overall sign ascribed to the $3jm$ symbol and is omitted in all which follows. If we designate the phase on transposition of the component irreps in the Wigner coefficient as $\Phi(r\lambda:\lambda_1\lambda_2)$, it is related to the $3j$ permutation phase by

$$\begin{aligned}\Phi(r\lambda:\lambda_1\lambda_2) &= \Phi(r\lambda:\lambda_2\lambda_1) = \phi(\lambda_1\lambda_2\lambda^* r)\phi(\lambda_1)\phi(\lambda_2)\\ &= \phi(\lambda_1\lambda_2\lambda^* r)\phi(\lambda).\end{aligned} \tag{2.10}$$

Since the $3j$ permutation phase is independent of order we must have

$$\begin{aligned}\Phi(r\lambda:\lambda_1\lambda_2) &= \Phi(r\lambda_1^*:\lambda^*\lambda_2)\phi(\lambda_2)\\ &= \Phi(r\lambda_2^*:\lambda_1\lambda^*)\phi(\lambda_1).\end{aligned} \tag{2.11}$$

For $\lambda_1 \simeq \lambda_2$ we have

$$\phi(\lambda_1 \otimes \sigma\lambda^* r) = \Phi(r\lambda:\lambda_1 \otimes \sigma) = \Phi(r\lambda_1^*:\lambda^*\lambda_1)\phi(\lambda_1) \tag{2.12}$$

and for $\lambda_1 \simeq \lambda_2 \simeq \lambda^*$ we have

$$\phi(\lambda_1 \otimes \nu r) = \Phi(r\lambda_1^*:\lambda_1 \otimes \sigma) \text{ with } \phi(\lambda_1) = +1. \tag{2.13}$$

The $1j$ phase is just the special case of the $3j$ phase with one of the component irreps the scalar irrep.

A Wigner coefficient with complex conjugate irreps is related to the complex conjugate Wigner coefficient using (2.6) and (2.9) to give

$$\langle r^*\lambda^*M\,|\,\lambda_1^*M_1^*\cdot\lambda_2^*M_2^*\rangle$$
$$=\langle r\lambda M\,|\,\lambda_1M_1\cdot\lambda_2M_2\rangle^*(\lambda_1)_{M,M_1^*}$$
$$\times(\lambda_2)_{M_2M_2^*}(\lambda)^*_{MM^*}A(\lambda:\lambda_1\lambda_2)_{rr^*}, \tag{2.14}$$

where

$$A(\lambda:\lambda_1\lambda_2)_{rr^*}\equiv A(\lambda_1\lambda_2\lambda^*)_{rr^*}=A(\lambda^*:\lambda_1^*\lambda_2^*)_{r^*r}.$$

Transposing and using (2.11) we conclude

$$\Phi(r\lambda:\lambda_1\lambda_2)=\Phi(r^*\lambda^*:\lambda_1^*\lambda_2^*). \tag{2.15}$$

For purposes of Sec. V we note the $9j$ recoupling coefficient (overlap integral between two different sequences for pairwise coupling of four component irreps) is defined by the symmetric sum on six Wigner coefficients as

$$\langle r'\lambda'M'|_i\lambda_iM\rangle\langle_ir_i\lambda_iM|_i\lambda_{jj}M_j\rangle\langle_i\lambda_{ji}M_j|r_j\lambda_jM_j\rangle\langle\lambda_jM_j|r\lambda M\rangle$$
$$=\delta_{\lambda\lambda'}\cdot\delta_{MM'}\begin{bmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\\ r' & r_j\end{bmatrix}^r_{\ {}_ir} \tag{2.16}$$

and is independent of the basis label M. A similar symmetrized sum on six $3j$ symbols (but including a sum on $M=M'$) defines the more symmetric $9j$ symbol related to the $9j$ recoupling coefficient by

$$\begin{bmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\\ r' & r_j\end{bmatrix}^r_{\ {}_ir}=[\,|{}_i\lambda\|\lambda_j|\,]^{1/2}\begin{bmatrix}\lambda & \lambda_j^*\\ {}_i\lambda^* & {}_i\lambda_j\\ r'^* & r_j\end{bmatrix}^{r^*}_{\ {}_ir}$$
$$\times A(\lambda^*:\lambda_j^*)^*_{r^*r}A(\lambda^*:{}_i\lambda^*)_{r^*r'}. \tag{2.17}$$

The $9j$ symbol has permutational symmetry on its rows and columns and reflection across its main diagonal corresponding to the group $S_2V(S_3)^2$ of order 72. For SU2 Regge showed the Wigner coefficient possessed surprising additional symmetries which now bear his name. We shall show in Sec. V that these derive from the higher symmetry of the $9j$ symbol. With one of the four component irreps being the scalar irrep the $9j$ recoupling coefficient contracts to the $6j$ recoupling coefficient. With two of the irreps being the scalar irrep the $9j$ recoupling coefficient contracts to a $3j$ permutation phase

$$\begin{bmatrix}\lambda & \lambda_1 & \lambda_2\\ \lambda_1 & \lambda_1 & 0\\ \lambda_2 & 0 & \lambda_2\\ r' & & \end{bmatrix}^r=\delta_{rr'}=\Phi(r\lambda:\lambda_1\lambda_2)\begin{bmatrix}\lambda & \lambda_1 & \lambda_2\\ \lambda_2 & 0 & \lambda_2\\ \lambda_1 & \lambda_1 & 0\\ & & r'\end{bmatrix}^r. \tag{2.18}$$

Irrep basis labelling is usually specified w.r.t. a chain of subgroups G/G_j. The Wigner coefficients can be expressed in terms of those of the subgroup times an isoscalar factor as

$$\langle r'\lambda\,(r\lambda_jM_j)|_i\lambda\,(_ir_i\lambda_{ji}M_j)\rangle_G$$
$$=\begin{pmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\\ r' & r_j\end{pmatrix}^r_{\ {}_ir}\langle r_j\lambda_jM_j|_i\lambda_{ji}M_j\rangle_{G_j}. \tag{2.19}$$

Permutation phases and $1jm$ symbols similarly factor. The transposition phase factors as

$$\Phi(r\lambda:_i\lambda)_G=\Phi(\lambda,r_i'\lambda;\lambda_j,r_{ji}\lambda_j)\Phi(r_j\lambda_j:_i\lambda_j)_{G_j}, \tag{2.20}$$

where $\Phi(r'_i\lambda;\lambda_j,r_{ji}\lambda_j)$ is the transposition phase of the isoscalar factor. The $1jm$ symbol factors as

$$(\lambda)_{r\lambda_jM_j,r^*\lambda_j^*M_j^*}$$
$$=\left(\frac{|\lambda|_G}{|\lambda_j|_{G_j}}\right)^{1/2}\begin{pmatrix}0 & 0_j\\ \lambda & \lambda_j\\ \lambda^* & \lambda_j^*\end{pmatrix}^r_{\ r^*}(\lambda_j)_{M_jM_j^*}, \tag{2.21}$$

where the dimensions, the scalar irreps, and compex conjugation are to be considered w.r.t. the corresponding group. For the groups we shall consider the $1jm$ symbol can be made a pure phase so the dimensional and magnitude factors of (2.21) cancel and we define the $1jm$ phase factor as

$$(G\lambda:G_j\lambda_j)\equiv\left(\frac{|\lambda|_G}{|\lambda_j|_{G_j}}\right)^{1/2}\begin{pmatrix}0 & 0_j\\ \lambda & \lambda_j\\ \lambda^* & \lambda_j\end{pmatrix}^r_{\ r^*}. \tag{2.22}$$

Establishing a scheme for evaluating the isoscalar factor and its permutation properties establishes by iteration the evaluation of all coupling constants and their permutation phase convention.

For the unitary and unitary unimodular groups factoring the maximum abelian subgroup $Un/U1\simeq SUn/Cn$, where Cn is the cyclic group on n objects, corresponds to mapping the pseudoscalar irreps $[l^n]$ of Un onto the scalar irrep 0 of SUn. For $u\in Un$ the pseudoscalar irreps transform according to the determinant $u|[l^n]|>=|[l^n]>(\det u)^l$ and determinantal powers couple trivially,

$$u|[l_1^n]\rangle|[l_2^n]\rangle$$
$$=|[l_1^n]\rangle|[l_2^n]\rangle\langle\det u)^{l_1+l_2}=u|[(l_1+l_2)^n]\rangle. \tag{2.23}$$

So one may identify the coupling algebra of Un with that of SUn. The question remains: Is this trivial identification still possible if one wishes to demand additional symmetries of the coupling coefficients which follow from duality? We answer this question in the affirmative in Sec. VI.

III. DOUBLE COSET DECOMPOSITION OF S_L

Let S_{L_j} symbolize the direct product group $\times S_{L_j}$ such that $\Sigma L_j=L$, then double cosets (DC) $S_{iL}\backslash S_L/S_{L_j}$ are in one to one correspondence with DC symbols

$$\begin{bmatrix}L & L_j\\ {}_iL & {}_iL_j\end{bmatrix},$$

the ${}_iL_j$ taking on all non-negative integral values such that $\Sigma\,{}_iL_j=L_j$ and $\Sigma_j\,{}_iL_j={}_iL$. The intertwining subgroups $S_L\cap qS_{L_j}q^{-1}\simeq q^{-1}S_{iL}q\cap S_{L_j}$ is isomorphic to the direct product subgroup $S_{{}_iL_j}$. The DC representative q can be chosen as that permutation which rearranges the ordered integers 1 to L arranged in subsets $({}_iL_j)$ ordered on the first index i and then on the second index j to an ordering on the second index j and then on the first index i, the order within each subset being maintained; i.e.,

$$q=\begin{pmatrix}{}_1L_1\ {}_2L_1\ {}_3L_1\cdots\\ {}_1L_1\ {}_1L_2\ {}_1L_3\cdots\end{pmatrix}.$$

For our purposes we will restrict the range of i and j to two (or later when indicated to three) since all other DC decompositions can be built in steps from decompositions into two parts. These DC we will call fundamental and q is the cyclic permutation $({}_2L_1+{}_1L_2)^{{}_1L_2}$ with parity $(-1)^{{}_1L_2\,{}_2L_1}$. The dou-

ble coset matrix elements (DCME) representing q in a (generally mixed) basis symmetry adapted to the two subgroup sequences on the left and on the right possess special block diagonal form because of Schur's lemma

$$\begin{bmatrix} \lambda & q \\ r'_i\lambda_i r_i\lambda'_{j\,i}m'_j & r\lambda_j r_{j\,i}\lambda_{j\,i}m_j \end{bmatrix} = \delta_{\lambda_{ji}\lambda'_j}\delta_{{}_im_{ji}m'_j}\begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \\ r' & r_j \end{bmatrix}_{ir} . \tag{3.1}$$

The irreps are labels from the respective groups and the rows (columns) of the DCME must couple to the left (up) in accordance with the Littlewood–Richardson rules for outer products. The indexes r label any multiplicity ($|\lambda:\lambda_1\lambda_2| > 1$) connected with the various outer products. For ease in printing multiplicities may be omitted when their role is not crucial to the discussion. Since all subgroups we consider are ambivalent and have only irreps of the first Schur type (orthogonal irreps), the DCME may be chosen as elements of an orthogonal matrix; i.e., they are real with the inverse element given by the transpose (interchange i and j indexes) such that

$$\sum_{\lambda_j}\begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \\ r' & r_j \end{bmatrix}_{ir'}\begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda' & {}_i\lambda_j \\ r'' & r_j \end{bmatrix}_{ir'} = \delta_{\lambda\lambda'}\delta_{r'r''}\delta_{i'_ji''_j} . \tag{3.2}$$

The DCME is orthogonal on the label sets $(r'_i\lambda_i r|r\lambda_j r_j)$, where λ, ${}_i\lambda_j$, and q act as fixed parameters. The DCME may be regarded as the orthogonal (overlap) transformation between equivalent bases symmetry adapted to the different subgroup chains.

The group orthogonality and completeness conditions for finite or compact groups causes the weighted double coset matrix element (WDCME) defined as

$$\begin{pmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \\ r' & r_j \end{pmatrix}_{ir} \equiv \left(\frac{|\lambda\,||{}_i\lambda_j|{}_iL\,!L_j!}{|{}_i\lambda\,||\lambda_j|L\,!{}_iL_j!}\right)^{1/2}\begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \\ r' & r_j \end{bmatrix}_{ir} \tag{3.3}$$

(note a + 1 phase is chosen in this definition) to be orthogonal on the label sets $(r\lambda r'|{}_ir_i\lambda_j r_j q)$, where ${}_i\lambda$ and λ_j act as fixed parameters. The WDCME may be regarded as the orthogonal transformation between two induced bases known to be equivalent by Mackey's[10] subgroup theorem.

In the case ${}_iL = L_i$ the DCME is expressed in a non-mixed bases, the DC representative belongs to the self-inverse class $2^{{}_1L_2}1^{({}_1L_1+{}_2L_2)}$, and the character is given by the trace

$$[\lambda\,]_{2^{1L_2}} = \sum_{\lambda_j\lambda_j}|{}_1\lambda_1||{}_1\lambda_2||{}_2\lambda_2| \times\begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ \lambda_1 & {}_1\lambda_1 & {}_1\lambda_2 \\ \lambda_2 & {}_1\lambda_2 & {}_2\lambda_2 \\ r & r_1 & r_2 \end{bmatrix}_{r\,r_1\,r_2} . \tag{3.4}$$

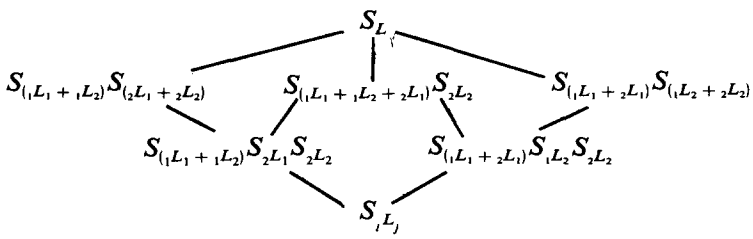
The trace is independent of the particular basis so we may choose $L_2 = {}_1L_2$ with the more compact result

$$[\lambda\,]_{2^{L_2}} = \sum_{\lambda_{j_1}\lambda_1}|{}_1\lambda_1||\lambda_2| \times\begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ \lambda_1 & {}_1\lambda_1 & \lambda_2 \\ \lambda_2 & \lambda_2 & 0 \\ r & r_1 & \end{bmatrix}_{r\,r_1} . \tag{3.5}$$

Since the trace may be evaluated by Murnaghan's[11] formula, (3.5) provides a useful constraint on the number of ± 1 roots for the DCME as being $(|\lambda\,| \pm [\lambda\,]_{2^{L_2}})/2$.

Whenever the order of two subsets become equal as for $L_1 = L_2$ one may insert the semidirect product $S_2 \wr S_{L_1}{}^2$ into the subgroup chain and utilize its irreps to label the DCME. If $\lambda_1 \simeq \lambda_2$ the symmetrized outer products $\lambda_1 \otimes \sigma$ ($\sigma \vdash 2$) may partially resolve otherwise arbitrary muliplicities. It is important to note that the orthogonality properties of the DCME and the WDCME are invariant under unitary transformations on the multiplicities.

One may use the transitivity of equivalence transformations to progress through a sequence of subgroup chains. Thus via the subgroup nesting



one may express any fundamental DCME in terms of DCME with one vanishing component

$$\begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \\ r' & r_j \end{bmatrix}_{i\,r} = \sum_{\epsilon}\begin{bmatrix} \lambda & \epsilon & {}_2\lambda_2 \\ {}_1\lambda & {}_1\lambda & 0 \\ {}_2\lambda & {}_2\lambda_1 & {}_2\lambda_2 \\ r^1 & s_1 & \end{bmatrix}_{2r}{}^{s}\begin{bmatrix} \epsilon & \lambda_1 & {}_1\lambda_2 \\ {}_1\lambda & {}_1\lambda_1 & {}_1\lambda_2 \\ {}_2\lambda_1 & {}_2\lambda_1 & 0 \\ s_1 & r_1 & \end{bmatrix}_{1}{}^{s}{}_{1r}\begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ \epsilon & \lambda_1 & {}_1\lambda_2 \\ {}_2\lambda_2 & 0 & {}_2\lambda_2 \\ s & & r_2 \end{bmatrix}_{1}{}^{r}{}_{1s} \tag{3.6}$$

In such expansions each subgroup link corresponds to a DCME factor. Since the first and third DCME on the rhs of (3.6) are related by the transposition of a row or column to the form of the middle DCME, the basic task is to evaluate DCME of the latter form.

On the basis of sequential equivalence transformations two back-coupling relations have been developed which allow any

DCME to be evaluated from known DCME of lower rank. These relations are compatible with the symmetries of duality discussed in Sec. V and the phase convention proposed in Sec. VII so DCME evaluated with their use will also be compatible with these symmetries and phases. A particular application of the first back-coupling relation shows upon transposition of a column (similarly for a row) one has

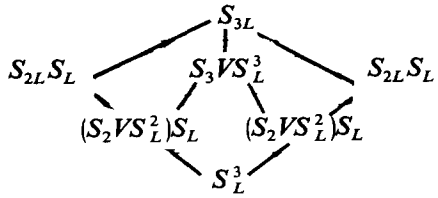
$$\begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ {}_1\lambda & {}_1\lambda_1 & {}_1\lambda_2 \\ {}_2\lambda & {}_2\lambda_1 & {}_2\lambda_2 \\ r' & r_1 & r_2 \end{bmatrix} \begin{matrix} r \\ {}_1r \\ {}_2r \end{matrix} = \begin{bmatrix} \lambda & \lambda_2 & \lambda_1 \\ {}_1\lambda & {}_1\lambda_2 & {}_1\lambda_1 \\ {}_2\lambda & {}_2\lambda_2 & {}_2\lambda_1 \\ r' & r_2 & r_1 \end{bmatrix} \begin{matrix} s \\ {}_1s \\ {}_2s \end{matrix} \Phi(\lambda:\lambda_1\lambda_2)_{rs}\ \Phi({}_1\lambda:{}_1\lambda_1\ {}_1\lambda_2)_{{}_1r{}_1s}\Phi({}_2\lambda:{}_2\lambda_1{}_2\lambda_2)_{{}_2r{}_2s}, \tag{3.7}$$

where

$$\Phi(\lambda:\lambda_1\lambda_2)_{rs} \equiv \begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ \lambda_2 & 0 & \lambda_2 \\ \lambda_1 & \lambda_1 & 0 \end{bmatrix}^{r}_{s}$$

is the fundamental transposition matrix to be fixed by a phase convention.

When several subsets are of equal size the semidirect product groups offer more convenient subgroup nesting for expanding a fundamental DCME. The subgroup nesting



gives the expansion for equivalent component irreps λ,

$$\begin{bmatrix} \Gamma & \Lambda' & \bar{\lambda} \\ \Lambda & \lambda & \lambda \\ \lambda & \lambda & 0 \end{bmatrix}^{\sigma} = \sum_{\nu} \begin{bmatrix} \Gamma & \lambda\otimes\nu \\ \Lambda & \lambda\otimes\sigma \\ \lambda & \lambda \end{bmatrix}^{s} \times \begin{bmatrix} \nu & \sigma' & [1] \\ \sigma & [1] & [1] \\ [1] & [1] & 0 \end{bmatrix} \begin{bmatrix} \Gamma & \Lambda' & \lambda \\ \lambda\otimes\nu & \lambda\otimes\sigma' & \lambda \end{bmatrix}_{s}. \tag{3.8}$$

The use of semidirect product subgroups and symmetrized products introduces some ambiguity into the designation of the DC representative since under $S_2VS_L^2 \backslash S_{2L}/S_2VS_L^2$ the formerly distinct DC's with symbols

$$\begin{bmatrix} 2L & L & L \\ L & L-K & K \\ L & K & L-K \end{bmatrix}$$

and

$$\begin{bmatrix} 2L & L & L \\ L & K & L-K \\ L & L-K & K \end{bmatrix}$$

are coalesced into one DC.

If the parity of K and $L-K$ are the same there is no ambiguity in the phase convention for association. If they have opposite parity, we choose the DC representative to have the same parity as L. When using semidirect product subgroups care must be exercised in identifying equivalent irreps of the isomorphic intertwining subgroups since the action of q may cause the interchange of distinct but similar classes. This nuance does not affect the phase arguments presented here.

IV. ASSOCIATION AND 1/m FACTORS

An involution is an automorphism on the group which when carried out twice results in the identity map. Under the involution, which we shall initially represent by a $\overline{\text{bar}}$, a matrix irrep may become equivalent to another (barred) irrep by

$$\begin{bmatrix} \lambda g \\ mn \end{bmatrix} = (\lambda)_{m\bar{m}} \begin{bmatrix} \bar{\lambda} g \\ \bar{m}\bar{n} \end{bmatrix}^{-} (\lambda)^{*}_{n\bar{n}}. \tag{4.1}$$

Carried out twice Schur's lemma requires

$$(\lambda)_{m\bar{n}}\,(\bar{\lambda})_{\bar{m}m'})^{-} = \delta_{mm'}\theta(\lambda), \tag{4.2}$$

where $\theta(\lambda) = \theta(\bar{\lambda})^{-}$ is a basis independent phase. In case $\lambda \simeq \bar{\lambda}$ Eq. (4.2) places a restriction on the phase transformation required by the involution.

Complex conjugation is an antiunitary involution applicable to all groups. The Schur invariant

$$\theta(\lambda) = \sum_{g} \frac{\text{trace}\,[\lambda g^2]}{|G|}$$

classifies irreps into three types:

(i) $\lambda^* \simeq \lambda$ and the scalar irrep 0 is contained in the symmetrized square product $|0{:}\lambda \otimes [2]| = 1$. In this case $\theta(\lambda) = +1$ and the matrix representation can be made real (orthogonal) by a symmetric unitary transformation.

(ii) $\lambda^* \simeq \lambda$ and the scalar irrep is contained in the antisymmetrized square product $|0{:}\lambda \otimes [1^2]| = 1$. In this case $\theta(\lambda) = -1$ and the matrix representation cannot be brought to real form (it is called symplectic) and (4.2) restricts the phase transformation to skew symmetric unitary matrices.

(iii) $\lambda^* \not\simeq \lambda$ and there is no restriction on the choice of $\theta(\lambda)$ and the representation is complex.

For complex conjugation the transformation is the 1/m matrix and $\theta(\lambda)$ is the 1/j phase which for Un we label $\phi(n\lambda)$.

The alternating group A_L is a normal subgroup of index two in S_L which give rise to the involutive operation of association w.r.t. A_L. Designating association by a tilde we have for $\pi \epsilon S_L$

$$\begin{bmatrix} \lambda\ \pi \\ m\ n \end{bmatrix} = (-1)^P (\lambda)_{m\tilde{m}} \begin{bmatrix} \tilde{\lambda}\ \pi \\ \tilde{m}\ \tilde{n} \end{bmatrix} (\lambda)^*_{n\tilde{n}}, \tag{4.3}$$

where P is the parity of the permutation π and

$$(\lambda)_{m\tilde{m}} (\tilde{\lambda})_{\tilde{m}m'} = \theta(\lambda)\delta mm'. \tag{4.4}$$

Note (4.3) is valid even when the matrix representation is in a mixed basis. As is the case for the Schur invariant association distinguishes three types of irreps in S_L:

(i) $\lambda \simeq \tilde{\lambda}$ and the antisymmetric irrep $[1^L]$ is contained in the symmetrized square inner product $|[1^L]:\lambda \odot [2]| = 1$, in which case $\theta(\lambda) = +1$,

(ii) $\lambda \simeq \tilde{\lambda}$ and $|[1^L]:\lambda \odot [1^2]| = 1$, in which case $\theta(\lambda) = -1$, and

(iii) $\lambda \not\simeq \tilde{\lambda}$, in which case there is no restriction on the choice $\theta(\lambda)$.

The association transformation is essentially a $1jm$ factor for S_L defined w.r.t. the one-dimensional antisymmetric irrep $[1^L]$. From $|[L]:\lambda \odot [2]| = 1$ and $|[L]:\lambda \odot [1^2]| = 0$ and the above inner product frequencies one can show the three association types correspond to the following Schur (complex conjugation) types in A_L:

(i) $\lambda\downarrow = \langle\lambda\rangle_+ + \langle\lambda\rangle_-$ and both are orthogonal irreps,

(ii) $\lambda\downarrow = \langle\lambda\rangle_+ + \langle\lambda\rangle_-$ with $\langle\lambda\rangle^*_+ \simeq \langle\lambda\rangle_-$, and

(iii) $\lambda\downarrow = \langle\lambda\rangle$ is orthogonal.

The association phase may be obtained by taking the Young tableau in standard order (the integers 1 to L entered in order from left to right proceeding in order from the top to bottom row) and interchanging rows and columns by transposing the frame about its main diagonal to obtain a Young tableau for the associate irrep. The association phase is $\theta(\lambda) = (-1)^P$, where P is the parity of the permutation required to restore standard order for the associated irrep tableau.[12] For self-associate irreps $\lambda \simeq \tilde{\lambda}$ this become $\theta(\lambda) = (-1)^{(L-r)/2}$ where r is the number of nodes along the main diagonal of the Young frame (sometimes called the rank). This agrees with a result previously obtained by Butler and King.[13]

On symmetry adapting a basis to a subgroup chain both the $1jm$ matrix and the association matrix factor, thus defining the corresponding $1jm$ and association factors. Before writing down their defining equations we need to consider a notational difficulty related to complex conjugation. In discussing the Derome–Sharp factor we have used the star (*) symbol in a triad to denote complex conjugation w.r.t. the same group Un. For the $1jm$ factor we need to denote the complex conjugate irreps of a triad w.r.t. the subgroup chain $U(n_1 + n_2)/Un_1 \times Un_2$. To distinguish this meaning we will use a bar ($\bar{}$) symbol. In a formal manner we can extend our labelling by using partitions with negative components to distinguish contravariant tensors from covariant tensors in Un. Because all irreps labelled by n rowed rectangular Young frames are equivalent to the scalar irrep in SUn (i.e., $[l^n] \simeq [0]$), one may always label the complex conjugate of an irrep in SUn by a partition with positive integral components as

$$\lambda^* \equiv [l^n/\lambda] \equiv [l_i^*] = [l - l_{n-i+1}]. \tag{4.5}$$

Since outer products forming rectangular irreps are multiplicity free $|[l^n]:\lambda\lambda^*| = 1$, this provides a unique way to label the irrep complex conjugate to λ in SUn. Note equivalence in SUn also allows one to reduce all irrep labels in this group to partitions with less than n parts ($l_n = 0$). Using this notation the * and $\bar{}$ symbols distinguish the triads $([(l_1 + l_2)^n/\lambda]:[l_j^n/\lambda_j]r^*)$ and $([l^{(n_1+n_2)}/\lambda]:[l^{n_j}/\lambda_j]\bar{r})$. Especially note the multiplicity labels, although related, must be distinguished in the two triads. Because irreps labelled by rectangular patterns have a central role in what follows, we note the outer product of two rectangular pattern irreps may be decomposed as

$$[l^n]\cdot[l'^{n'}] = \sum_\lambda \begin{bmatrix} l^n \bar{\lambda} \\ \lambda \end{bmatrix}. \tag{4.6}$$

The irrep symbol used in the sum is introduced to denote the pattern with row components $\mathscr{L}_i = l + l' - l_{n'-i+1}$, $\mathscr{L}_{n+i} = l_i$ for $1 \leqslant i \leqslant n'$, and $\mathscr{L}_i = l$ for $n' \leqslant i \leqslant n$, where $\lambda = [l_i]$. For $l' = l$ and $n' = n$ one separates the symmetrized product by

$$[l^n] \otimes \sigma = \sum_\lambda \begin{bmatrix} l^n \bar{\lambda} \\ \lambda \end{bmatrix},$$

where the rank L ($\lambda \vdash L$) is even or odd as σ is $[2]$ or $[1^2]$. One should distinguish the symbol

$$\begin{bmatrix} l^n \\ \lambda \end{bmatrix}$$

introduced in (4.6) from $[l^n/\lambda]$ introduced earlier.

Using the above notation we define the $1jm$ and association factors by

$$(\lambda)_{r\lambda_j m, \bar{r}\bar{\lambda}\bar{m}} = (n\lambda : n_j\lambda_j)r\bar{r}(\lambda_j)_{m\bar{m}} \tag{4.7}$$

and

$$(\lambda)_{r\lambda_j m, \tilde{r}\tilde{\lambda}_j\tilde{m}} = (\lambda : \lambda_j)r\tilde{r}(\lambda_j)_{m\tilde{m}} \tag{4.8}$$

with orthogonality

$$(n\lambda : n_j\lambda_j)_{r\bar{r}}(n\bar{\lambda} : n_j\bar{\lambda}_j)^*_{\bar{r}r'} = \delta rr'\phi(n\lambda)\phi(n_j\lambda_j) \tag{4.9}$$

and

$$(\lambda : \lambda_j)_{r\tilde{r}}(\lambda : \lambda_j)_{\tilde{r}r'} = \delta rr'\theta(\lambda)\theta(\lambda_j). \tag{4.10}$$

V. DUALITY OF Un and S_L

The Racah algebra of the general linear group Gln and its unitary Un and unitary unimodular SUn subgroups possesses structure over and above that discussed in Sec. II. The Schur–Weyl result that symmetrized homogeneous L th rank tensors of a basic n-dimensional defining vector space span the integer irreps of Gln and the irreps of S_L, allows a tensor basis to be completely labelled by a common irrep λ and basis indexes m and M specifying their transformation properties under action of S_L and Gln, respectively. We use the ket symbol

$$\left| \begin{matrix} m \\ \lambda \\ M \end{matrix} \right\rangle$$

for a tensor basis member in which the basis indexes m and M will be enlarged to indicate symmetry adaption to a subgroup as, for example, $m \to r_i \lambda_i m$ for S_L/S_{iL} or $M \to r\lambda_j M_j$ for $\mathrm{Gl}(n_1+n_2)/\mathrm{Gl}n_j$. Duality exists between outer products in S_L/S_{L_j} and inner products in Gln in that the former may be carried out by the orthogonal matrix basis projectors in S_L,

$$P^{\lambda}{}_{m,r\lambda_j m_j} \left| \begin{matrix} m_j \\ \lambda_j \\ M_j \end{matrix} \right\rangle = \left| \begin{matrix} m \\ \lambda \\ M \end{matrix} \right\rangle \left\langle \begin{matrix} r\lambda \\ M \end{matrix} \middle| \begin{matrix} \lambda_j \\ M_j \end{matrix} \right\rangle, \tag{5.1}$$

and the latter may be carried out by Wigner coefficients of Un (identical to those for the integer irreps of Gln),

$$\left| \begin{matrix} {}_i m \\ r\,{}_i\lambda \\ \lambda \\ M \end{matrix} \right\rangle = \left| \begin{matrix} {}_i m \\ {}_i\lambda \\ {}_i M \end{matrix} \right\rangle \left\langle \begin{matrix} {}_i\lambda \\ {}_i M \end{matrix} \middle| \begin{matrix} r\lambda \\ M \end{matrix} \right\rangle. \tag{5.2}$$

A similar duality[14] exists between outer products in $\mathrm{Gl}n_1 n_2/\mathrm{Gl}n_j$ and inner product coupling in S_L but we do not develop this aspect here. It is important to note that the tensors in (5.2) are by supposition basis members in Gln, while the dimensions of the defining spaces underlying the component tensors in the product on the l.h.s. of (5.1) need only be such that their union spans the defining vector space of the tensor on the right. A particular linear combination coefficient can be extracted from the sum on the r.h.s. of (5.1) by taking the scalar product with a particular tensor, giving

$$\left\langle \begin{matrix} r\lambda_j m_j \\ \lambda \\ M \end{matrix} \middle| \begin{matrix} m_j \\ \lambda_j \\ M_j \end{matrix} \right\rangle = \left\langle \begin{matrix} r\lambda \\ M \end{matrix} \middle| \begin{matrix} \lambda_j \\ M_j \end{matrix} \right\rangle. \tag{5.3}$$

The independence of the overlap on the basis index m_j follows from the independence of the projection (5.1) on the particular selection of basis member m_j in the subgroup. If all the tensors in (5.1) have the same defining vector space $n_j = n$, the linear combination coefficients on the r.h.s. of (5.1) are the Wigner coefficient of Un since (5.3) corresponds to their definition as an overlap integral in Un. If the defining vector space on the r.h.s. of (5.1) is simply the sum of the defining vector spaces for the tensor product of the left $n = n_1 + n_2$, we may symmetry adapt to the subgroup sequence $\mathrm{Gl}n/\mathrm{Gl}n_j$ such that $m \to r'\lambda' {}_j M'_j$ and use (5.2) in (5.3) to obtain

$$\delta_{rr'}\delta_{\lambda_j\lambda'_j}\delta_{M_j M'_j} \left\langle \begin{matrix} r\lambda \\ r'\begin{pmatrix} \lambda_1 & \lambda_2 \\ M_1 & M_2 \end{pmatrix} \end{matrix} \middle| \begin{matrix} \lambda_1 \\ \begin{pmatrix} \lambda_1 & 0 \\ M_1 & 0 \end{pmatrix} \end{matrix} \middle| \begin{matrix} \lambda_2 \\ \begin{pmatrix} 0 & \lambda_2 \\ 0 & M_2 \end{pmatrix} \end{matrix} \right\rangle$$
$$= \left\langle \begin{matrix} r\lambda \\ r'\lambda'_j \\ M'_j \end{matrix} \middle| \begin{matrix} \lambda_j \\ M_j \end{matrix} \right\rangle; \tag{5.4}$$

i.e., the linear combination coefficient on the r.h.s. of (5.1) is again a Wigner coefficient but in a basis symmetry adapted to the specified subgroup chain. The delta factor on the multiplicity means we have chosen to resolve any multiplicity arising in the lower index M in exactly the same way it is resolved in the upper index label m, i.e., the manner of resolving the multiplicity within a triad will be the same and does not depend on whether we are considering Gln or S_L. Also any given tensor which is a basis member for some Gln is also a basis member for any Gln' with $n' > n$ and we choose it to be so with the same sign. This is the same as choosing all equivalent irreps to have identical matrix representation in S_L which, of course, is the usual choice.

For clarity we explicitly indicate the dimensions of the defining vector spaces and consider the scalar product

$$\left\langle \begin{matrix} {}_i m \\ {}_i\lambda\ {}_i n \\ {}_i M \end{matrix} \middle| P^{\lambda}{}_{r',\lambda_i m, r\lambda_j m_j} \middle| \begin{matrix} m_j \\ \lambda_j n_j \\ M_j \end{matrix} \right\rangle = \left\langle \begin{matrix} {}_i\lambda \\ {}_i M \end{matrix} \middle| \begin{matrix} r'\lambda \\ M \end{matrix} \right\rangle \left\langle \begin{matrix} r\lambda \\ M \end{matrix} \middle| \begin{matrix} \lambda_j \\ M_j \end{matrix} \right\rangle. \tag{5.5}$$

Using DC decomposition the l.h.s. can be reduced to the DCME times appropriate Wigner coefficients and weighting factors. We consider three cases for the defining vectors spaces entering (5.5):

(a) ${}_i n = n = n_j$. All couplings correspond to inner products in Un and the coupling coefficients are Wigner coefficients of that group. One may use orthogonality of the WDCME and orthogonality on the Wigner coefficients to rearrange (5.5) to

$$\delta_{\lambda\lambda'}\cdot\delta_{MM'}\begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \\ r' & r_j \end{bmatrix}_{r}{}_{i}r$$
$$= \left\langle \begin{matrix} r'\lambda' \\ M' \end{matrix} \middle| \begin{matrix} {}_i\lambda \\ {}_i M \end{matrix} \right\rangle_n \left\langle \begin{matrix} {}_i r_i \lambda \\ {}_i M \end{matrix} \middle| \begin{matrix} {}_i\lambda_j \\ {}_i M_j \end{matrix} \right\rangle_n \left\langle \begin{matrix} {}_i\lambda_j \\ {}_i M_j \end{matrix} \middle| \begin{matrix} r_j\lambda_j \\ M_j \end{matrix} \right\rangle_n \left\langle \begin{matrix} \lambda_j \\ M_j \end{matrix} \middle| \begin{matrix} r\lambda \\ M \end{matrix} \right\rangle_n$$
$$= \delta_{\lambda\lambda'}\cdot\delta_{MM'}\begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \\ r' & r_j \end{bmatrix}_{r}{}_{i}r \,. \tag{5.6}$$

The apparent repetition of symbols on the right and left sides is to emphasize that the DCME on the l.h.s., an entity of S_L, is being identified with the $9j$ recoupling coefficient of Un defined by the expression in the middle. All labels, multiplicities, orthogonality relations, and values of these entities are identical. Kramer[15] first demonstrated such an identification for U2 and bipartition irreps in S_L and noted it must be true even for $n > 2$. We later formulated the identification in its present form.[3a] On the basis of this identification it is easy to recognize (3.6) as the expansion of a $9j$ recoupling coefficient in terms of $6j$ recoupling coefficients and the first of our back-coupling relations as the usual back-coupling relation for such coefficients. But many relations such as (3.5) or the second back-coupling relation have no previously developed partner in the usual formulation of a coupling algebra.

(b) ${}_i n = n = n_1 + n_2$. Half the couplings on the l.h.s. of (5.5) correspond to inner products in the respective groups Un and Un_j and the other half correspond to generalized branching for $\mathrm{U}n/\mathrm{U}n_j$. We expand all basis labels M to form $r\lambda_j M_j$, etc., appropriate to symmetry adaption in $\mathrm{U}n/\mathrm{U}n_j$ to obtain

$$\delta_{\lambda_j\lambda_j'}\delta_{M_jM_j'}\begin{pmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\\ r' & r_j\end{pmatrix}^{r}_{\ {}_ir}$$
$$=\left\langle\begin{matrix}r' & \lambda\\ r & (\lambda_jM_j)\end{matrix}\middle|\begin{matrix}{}_i\lambda\\ {}_ir({}_i\lambda_j{}_iM_j)\end{matrix}\right\rangle_{n_1+n_2}\left\langle\begin{matrix}{}_i\lambda_j\\ {}_iM_j\end{matrix}\middle|\begin{matrix}r_j\lambda_j'\\ M_j'\end{matrix}\right\rangle_{n_j}$$
$$=\delta_{\lambda_j\lambda_j'}\delta_{M_jM_j'}\begin{pmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\\ r' & r_j\end{pmatrix}^{r}_{\ {}_ir}. \tag{5.7}$$

In deriving this result we always choose as in (5.4) to resolve the multiplicity in a given triad $(\lambda:\lambda_j)$ in the same way no matter what the group context, i.e., S_L/S_{L_j}, $\mathrm{U}n/\mathrm{U}n_j$, or inner product in U$n$, so that

$$\left\langle\begin{matrix}\lambda_1 & \lambda_2\\ (\lambda_1M_1,0) & (0,\lambda_2M_2)\end{matrix}\middle|\begin{matrix}r & \lambda\\ r' & (\lambda_jM_j)\end{matrix}\right\rangle_{n_1+n_2}=\delta_{rr'}\left(\frac{|\lambda|L_j!}{L!|\lambda_j|}\right)^{1/2}, \tag{5.8}$$

which is consistent with (5.7) and a positive phase for maximum coupling. Again the apparent repetition of symbols on the right and left sides of (5.7) is to emphasize that the WDCME on the l.h.s., an entity of S_L, is being identified with the isoscalar factor of Un defined by the expression in the middle. All labels, multiplicities, orthogonality relations, and values of these entities are identical. This identification was first made in Ref. 3e, although for bipartition irreps it underlies Kramer and Seligman's[16] derivation of the Regge symmetries of SU2. The identification (3.3) between the DCME and WDCME thus relates the $9j$ recoupling coefficient with the isoscalar factor with the high symmetry of the former giving the Regge symmetries of the latter. (For SU2 the isoscalar factor is the Wigner coefficient.) Restricting some of the irreps to the totally symmetric irrep $\lambda=[L]$, Jucys[17] has obtained similar results.

(c) ${}_1n+{}_2n=n=n_1+n_2$ with ${}_in\cap n_j\equiv{}_in_j$. All couplings correspond to generalized branching, so expanding ${}_iM$ and M_j to correspond to the appropriate subgroup chain and using (5.8) we obtain

$$\begin{bmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\\ r' & r_j\end{bmatrix}^{r}_{\ {}_ir}=\left\langle\begin{matrix}m & m\\ \lambda & \lambda\\ r'({}_i\lambda) & r(\lambda_j)\\ {}_ir({}_i\lambda_j{}_iM_j) & r_j({}_i\lambda_j{}_iM_j)\end{matrix}\right\rangle. \tag{5.9}$$

The r.h.s. is a transformation matrix between equivalent bases in Un and thus is a DCME in Un corresponding to the DC decomposition $\mathrm{U}_in\setminus\mathrm{U}n/\mathrm{U}n_j$. Relation (5.9) could equally have been derived by considering DC decomposition of a scalar product involving a matrix basis projector in Un. The DC representative of Un actually can be chosen to belong to the symmetric subgroup $S_n\subset\mathrm{U}n$ in that it interchanges the distinct dimension ${}_1n_2\leftrightarrow{}_2n_1$ and could be designated by the DC symbol

$$\begin{bmatrix}n & n_j\\ {}_in & {}_in_j\end{bmatrix}.$$

This identification of DCME in the two groups was first made in Ref. 3b.

These identifications between S_L and Un must hold in

TABLE I. Symmetries the DCME and WDCME must satisfy due to equivalence and complex conjugation in SUn and the identifications (a) (5.8), (b) (5.12), and (c) (5.13). The factors $a(n\lambda{:}\lambda_jr)$ and $(n\lambda{:}n_j\ \lambda_jr)$ are the real phase forms (±1) of the Derome–Sharp and $1/m$ factors. r^* and $\bar r$ distinguish complex conjugation of the triad $(\lambda{:}\lambda_j)$ in SUn and SU$n/$SUn_j, respectively.

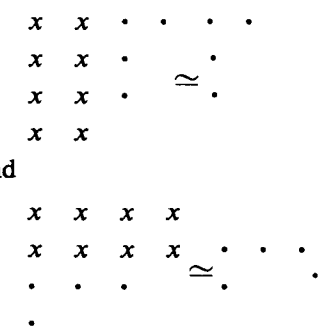
	Equivalence	Complex conjugation
(a)	$\begin{bmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\\ r' & r_j\end{bmatrix}^{r}_{\ {}_ir}=\begin{bmatrix}[l^n\lambda] & [l_j^n\lambda_j]\\ [{}_il^n{}_i\lambda] & [{}_il_j^n{}_i\lambda_j]\\ r' & r_j\end{bmatrix}^{r}_{\ {}_ir}$	$\begin{bmatrix}[l^n/\lambda] & [l_j^n/\lambda_j]\\ [{}_il^n/{}_i\lambda] & [{}_il_j^n/{}_i\lambda_j]\\ r'^* & r_j^*\end{bmatrix}^{r^*}_{\ {}_ir^*}\ a(n\lambda{:}\lambda_jr)a(n_i\lambda{:}{}_i\lambda_{j\,i}r)a(n\lambda{:}{}_i\lambda r')a(n\lambda_j{:}{}_i\lambda_jr_j)$
(b)	$\begin{pmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\\ r' & r_j\end{pmatrix}^{r}_{\ {}_ir}=\begin{pmatrix}[l^n\lambda] & [l^{n_j}\lambda_j]\\ [{}_il^n{}_i\lambda] & [{}_il^{n_j}{}_i\lambda_j]\\ r' & r_j\end{pmatrix}^{r}_{\ {}_ir}$	$\begin{pmatrix}[l^n/\lambda] & [l^{n_j}/\lambda_j]\\ [{}_il^n/{}_i\lambda] & [{}_il^{n_j}/{}_i\lambda_j]\\ r'^* & r_j^*\end{pmatrix}^{\bar r}_{\ {}_i\bar r}\ (n\lambda{:}n_j\lambda_jr)(n_i\lambda{:}n_{j\,i}\lambda_{j\,i}r)a(n\lambda{:}{}_i\lambda r')a(n_j\lambda_j{:}{}_i\lambda_jr_j)$
(c)	$\begin{bmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\\ r' & r_j\end{bmatrix}^{r}_{\ {}_ir}=(-1)^{l_i(n_2:n_1)}\begin{bmatrix}[l^n\lambda] & [l^{n_j}\lambda_j]\\ [{}_il^n{}_i\lambda] & [{}_il^{n_j}{}_i\lambda_j]\\ r' & r_j\end{bmatrix}^{r}_{\ {}_ir}=(-1)^{l_i(n_2:n_1)}\begin{bmatrix}[l^n/\lambda] & [l^{n_j}/\lambda_j]\\ [{}_il^n/{}_i\lambda] & [{}_il^{n_j}/{}_i\lambda_j]\\ \bar r' & \bar r_j\end{bmatrix}^{\bar r}_{\ {}_i\bar r}\ (n\lambda{:}n_j\lambda_jr)({}_in_i\lambda{:}{}_in_j\ {}_i\lambda_j\ {}_ir)(n_j\lambda_j{:}{}_in_j\ {}_i\lambda_jr_j)(n\lambda{:}{}_in_i\lambda r')$

magnitude and phase. Since restriction from Un to SUn involves at most only a phase and a possible transformation on the multiplicity indexes, these relations must hold at least in magnitude for the coupling algebra of SUn. In the next section we argue multiplicities may be resolved so the relations hold in magnitude and phase even on restriction to SUn. We therefore set down in Table I the symmetries the DCME and WDCME must possess due to equivalence and complex conjugation in SUn. Under complex conjugation we use the real phase form (a single ± 1 element in any row or column) of the Derome–Sharp and 1jm factors discussed in the next section. We use the notation $\bar{r}$ to distinguish the multiplicity related to r by taking the complex conjugate of the triad $(\lambda: \lambda_j)$ in SUn/SUn_j. This is in general distinct from r^* related to r by taking the complex conjugate of the triad $(\lambda: \lambda_j)$ in SUn. The explicit sign factor $(-1)^{l_1 n_2 2 n_1}$ enters part (c) of Table I because the DCME is a matrix element of an operator in $S_n \subset$ Un (and not in SUn) with determinant $(-1)^{n_2 2 n_1}$ in the defining irrep.

The DCME and WDCME by being entities in S_L also possess the symmetry of association w.r.t. A_L. Under association $[l^n\lambda]$ goes to

$$\begin{bmatrix} n^l \\ \tilde{\lambda} \end{bmatrix}$$

and $[l^n/\lambda]$ goes to $[n^l/\tilde{\lambda}]$. Equivalence can be considered a process of adding (or stripping) vertical rectangular segments to a Young frame. Association extends this to adding (or stripping) horizontal rectangular segments to a Young frame as suggested by the following sketch:



and

The adding (or stripping) of the rectangular segments must always be accomplished so that the resulting DCME satisfies the outer product rules in its rows and columns. Phases introduced by adding (or stripping) vertical rectangular segments are specifically indicated under the equivalence column of Table I. Adding (or stripping) horizontal rectangular segments requires association phases as discussed in the next section. Under association and with $l \leftrightarrow n$ and $\bar{r} \leftrightarrow r^*$ the symmetries under (a) in Table I become those of (c) and vice versa while those under (b) are self-associate.

VI. THE PHASE MATRICES

The foregoing is to a large extent a review in preparation for the multiplicity and phase arguments advanced in this and the next sections. Utilizing arguments based only on the theory of S_L we first show it is always possible to choose a real phase form for the matrix representing the transposition of two component irreps in a triad. The phase moreover is independent of the multiplicity resolution and so by a simple reordering of the multiplicity the matrix could be brought to the form of $\pm E$ within a given triad. Again utilizing arguments based only on the theory of S_L we next show there exists a large set of equivalence relations essentially identifying the multiplicity resolution in different triads. That is, if we consider the multiplicity as being labelled by the nodal placements in carrying out the Littlewood–Richardson rules for outer product decomposition in S_L, many triads give the same combinatorial possibilities and we choose to resolve the multiplicities in these triads in an identical manner. We say these belong to the same triad class. A graphical presentation of the triad condition helps to illustrate these equivalence relations and immediately demonstrates the real phase form for the transposition matrix and the association factor. The equivalence relations so demonstrated subsume all those required by duality. It is not clear from this graphical presentation, however, that the Derome–Sharp matrix and the 1jm factor are also in real phase form. Utilizing arguments based on duality we show these two phase matrices are related by the process of association. Thus if either of these phase matrices is put in real phase form then the other also is since all association factors are in real phase form. For distinct triad classes we argue the Derome–Sharp factor can always be put in real phase form without affecting our previous choices. This leaves us to examine the particular cases in which the triad classes become identical. For these cases we utilize the phase conventions proposed in the next section to show consistency is maintained when real phase forms are assumed.

For any binary coupling $(\lambda: \lambda_j)$ of S_L/S_{L_j} where λ has a Young's frame with l (n) nodes in the first row (column) (respectively, l_j and n_j for λ_j) we instead form the triad $([l^n]: \lambda_1\lambda_2\lambda_3)$, where $\lambda_3 \equiv [l^n/\lambda]$. We relate transposition phase matrices by the expansion (3.6) giving

$$\begin{bmatrix} [l^n] & \left[\frac{l^n}{\lambda_1}\right] & \lambda_1 \\ \left[\frac{l^n}{\lambda_2}\right] & \lambda_3 & \lambda_1 \\ \lambda_2 & \lambda_2 & 0 \end{bmatrix}_{r_1} r_2$$

$$= \begin{bmatrix} [l^n] & \lambda_3 & \left[\frac{l^n}{\lambda_3}\right] \\ \left[\frac{l^n}{\lambda_2}\right] & \lambda_3 & \lambda_1 \\ \lambda_2 & 0 & \lambda_2 \end{bmatrix}_{s'} r_2 \begin{bmatrix} \left[\frac{l^n}{\lambda_3}\right] & \lambda_2 & \lambda_1 \\ \lambda_1 & 0 & \lambda_1 \\ \lambda_2 & \lambda_2 & 0 \end{bmatrix}_{s'} s$$

$$\times \begin{bmatrix} [l^n] & \left[\frac{l^n}{\lambda_1}\right] & \lambda_1 \\ \lambda_3 & \lambda_3 & 0 \\ \left[\frac{l^n}{\lambda_3}\right] & \lambda_2 & \lambda_1 \end{bmatrix}_{r_1} s. \tag{6.1}$$

Each DCME appearing in (6.1) is orthogonal on the multiplicity indexes only, the other labels being fixed. For triads of

type I, $\lambda_1 \not\simeq \lambda_2 \not\simeq \lambda_3 \not\simeq \lambda_1$, one may use orthogonal transformations on the multiplicity indexes in the distinct triad spaces to bring each DCME factor to the form of a real phase times the identity matrix. Note the phase is independent of the multiplicity index and we have

$$\begin{bmatrix} [l^n] & \left[\frac{l^n}{\lambda_1}\right] & \lambda_1 \\ \left[\frac{l^n}{\lambda_2}\right] & \lambda_3 & \lambda_1 \\ \lambda_2 & \lambda_2 & 0 \end{bmatrix}_{r_1 r_2} = \delta_{r_1 r_2} \Phi\left(\left[\frac{l^n}{\lambda_3}\right]: \lambda_1 \lambda_2\right)$$

$$= \begin{bmatrix} \left[\frac{l^n}{\lambda_3}\right] & \lambda_2 & \lambda_1 \\ \lambda_1 & 0 & \lambda_1 \\ \lambda_2 & \lambda_2 & 0 \end{bmatrix}_{r_2 r_1}$$

$$= \delta_{r_1 r_2} \Phi\left(\left[\frac{l^n}{\lambda_3}\right]: \lambda_2 \lambda_1\right). \tag{6.2}$$

This is consistent with the identification of the DCME with a recoupling coefficient in SUn and the phase relations (2.10). For type II triads $\lambda_1 \simeq \lambda_2 \not\simeq \lambda_3$ the DCME is self-transpose and can be brought to diagonal form by an orthogonal similarity transformation. By using the symmetrized products of the semidirect product subgroup $S_2 \backslash S_{L_1}{}^2$ one requires the diagonal form to be a real phase determined by the symmetrized product times the identity matrix so that

$$\begin{bmatrix} [l^n] & \left[\frac{l^n}{\lambda_1}\right] & \lambda_1 \\ \left[\frac{l^n}{\lambda_1}\right] & \lambda_3 & \lambda_1 \\ \lambda_1 & \lambda_1 & 0 \end{bmatrix}_{\sigma r \, \sigma' r'} = \begin{bmatrix} \left[\frac{l^n}{\lambda_3}\right] & \lambda_1 \otimes \sigma' \\ \lambda_1 \otimes \sigma & \lambda_1 \otimes \sigma \end{bmatrix}_{r r'}$$

$$= \delta_{rr'} \delta_{\sigma\sigma'} \phi_\sigma, \tag{6.3}$$

where $\phi_\sigma = \pm 1$ as $\sigma = [2]$ or $[1^2]$. The character theory of S_L requires the forms $\delta r r'$ with phase independent of r since

$$\operatorname{trace}[l^n]_{2^{L_1} 1^{L_1}} = \sum_{\substack{\lambda_1 \lambda_3 \\ \sigma}} |\lambda_1|\,|\lambda_3| \begin{bmatrix} [l^n] & \left[\frac{l^n}{\lambda_1}\right] & \lambda_1 \\ \left[\frac{l^n}{\lambda_1}\right] & \lambda_3 & \lambda_1 \\ \lambda_1 & \lambda_1 & 0 \end{bmatrix}_{\sigma r}$$

$$= \sum_{\substack{\lambda_1 \lambda_3 \\ \sigma}} \phi_\sigma |\lambda_1|\,|\lambda_3| \left|\left[\frac{l^n}{\lambda_3}\right]: \lambda_1 \otimes \sigma\right|, \tag{6.4}$$

which may be checked using Murnaghan's formula for the characters of S_L. Type III triads have $\lambda_1 \simeq \lambda_2 \simeq \lambda_3$ so we may use the symmetrized products of the semidirect product subgroup $S_3 \backslash S_L{}^3$ to symmetry adapt the multiplicity index via the expansion (3.8),

$$\begin{bmatrix} [l^n] & \left[\frac{l^n}{\lambda}\right] & \lambda \\ \left[\frac{l^n}{\lambda}\right] & \lambda & \lambda \\ \lambda & \lambda & 0 \end{bmatrix}_{\sigma r \, \sigma' r'}$$

$$= \sum_\nu \begin{bmatrix} [l^n] & \lambda \otimes \nu \\ \left[\frac{l^n}{\lambda}\right] & \lambda \otimes \sigma \\ \lambda & \lambda \end{bmatrix}_{s r} \begin{bmatrix} \nu & \sigma' & [1] \\ \sigma & [1] & [1] \\ [1] & [1] & 0 \end{bmatrix}$$

$$\times \begin{bmatrix} [l^n] & \left[\frac{l^n}{\lambda}\right] & \lambda \\ \lambda \otimes \nu & \lambda \otimes \sigma' & \lambda \\ s & r' & \end{bmatrix}. \tag{6.5}$$

The DCME

$$\begin{bmatrix} [l^n] & \lambda \otimes \nu \\ \left[\frac{l^n}{\lambda}\right] & \lambda \otimes \sigma \\ \lambda & \lambda \end{bmatrix}_{s r}$$

is orthogonal on the indexes $(r|s\nu)$, the other labels being fixed. By orthogonal transformations of the multiplicity indexes we diagonalize this DCME to obtain

$$\begin{bmatrix} [l^n] & \left[\frac{l^n}{\lambda}\right] & \lambda \\ \left[\frac{l^n}{\lambda}\right] & \lambda & \lambda \\ \lambda & \lambda & 0 \end{bmatrix}_{\nu\sigma r \, \nu'\sigma' r'} = \delta_{\nu\nu'} \delta_{rr'} \begin{bmatrix} \nu & \sigma' & [1] \\ \sigma & [1] & [1] \\ [1] & [1] & 0 \end{bmatrix}. \tag{6.6}$$

Thus for $\nu = [3]$ or $[1^3]$ one has the same result as for type II triads (i.e., $\delta\sigma\sigma'\phi_\sigma$) but for $\nu = [2,1]$ one has the phase matrix

$$\begin{bmatrix} [2,1] & \sigma' & [1] \\ \sigma & [1] & [1] \\ [1] & [1] & 0 \end{bmatrix} = \begin{array}{c|cc} \sigma \backslash \sigma' & [2] & [1^2] \\ \hline [2] & -\frac{1}{2} & +\sqrt{3}/2 \\ [1^2] & +\sqrt{3}/2 & +\frac{1}{2} \end{array}. \tag{6.7}$$

Note a pure phase form for the transpose DC representative $S_2 \backslash S_3 / S_2$ could be chosen as

$$[[2,1](23)] = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

but then other matrices in this basis involve complex phases as

$$[[2,1](123)] = \begin{pmatrix} \omega & 0 \\ 0 & \omega^* \end{pmatrix}$$

with $\omega = \exp i\frac{2}{3}\pi$. Moreover, the form (6.7) is in accord with the WDCME equivalence

$$\begin{pmatrix} [2,1] & [2] & [1] \\ [1^2] & [1] & [1] \\ [1] & [1] & 0 \end{pmatrix} = \begin{pmatrix} [1] & [1] & 0 \\ 0 & 0 & 0 \\ [1] & [1] & 0 \end{pmatrix} = +1$$

and so is in accord with the full symmetries of duality and real coupling coefficients.

The above treatment is just a derivation via DC decomposition in S_L of the results obtained by Derome for the permutation matrices of the coupling algebra of any group. Note the form of a real phase times the identity matrix remains invariant under an orthogonal similarity transformation so some freedom in manipulating the multiplicity index still remains.

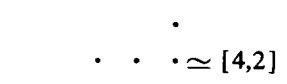
The equivalence relations in SUn require us to resolve multiplicities in triads related by these equivalences in identical ways and to have a transposition phase convention satisfying

$$\Phi(\lambda:\lambda_1\lambda_2) = \Phi([(l_1+l_2)^n\lambda\,]:[l_1^n\lambda_1][l_2^n\lambda_2]) = (-1)^{ln_1n_2}\Phi([l^{(n_1+n_2)}\lambda\,]:[l^{n_1}\lambda_1][l^{n_2}\lambda_2]). \tag{6.8}$$

In fact the set of equivalence relations under which multiplicities can be resolved in identical ways can be extended well beyond those required by duality. The application of the equivalence relations required by duality to the expansion (6.1), for example, shows that such must be the case. Triads of different rank which have identical multiplicity labelling form a class which we say has identical combinatorial structure. To have identical combinatorial structure the nodes added to or stripped from a triad via an equivalence relation enter the outer product in a fixed way and thus do not effect the range of nodal placements which we can and do use to label the multiplicities via the Littlewood–Richardson rules (i.e., the intermediate composite irrep patterns at the various stages of carrying out the outer product label the multiplicities). The triad condition (triangular condition of SU2) and the general equivalence relations can be given a useful graphical presentation. Given a triad with $|[l^n]:\lambda_K\lambda_L\lambda_M| \geqslant 1$, where $\lambda_K = [l_i(K)]$, etc., we consider the inverse outer product $[l^n/\lambda_L\,]/\lambda_K$. Following Robinson[18] this may be expanded as a determinantal outer product of one rowed irreps,

$$[l^n/\lambda_L\,]/\lambda_K \simeq |[l-l_{n-i+1}(L)-l_j(K)-i+j]|, \tag{6.9}$$

which graphically corresponds to a skew diagram also discussed by Robinson. The skew diagram is formed by deleting in a l by n rectangular array the nodes corresponding to the irrep λ_K from the upper left hand corner and the nodes corresponding to the irrep λ_L (inverted) from the lower right hand corner. The skew diagram labels a representation whose reduction is obtained by separating it into disjoint parts, each of which is a regular right diagram (perhaps inverted) and forming the outer product of the parts subject to the restriction: No two nodes appearing in the same row (column) of the skew diagram may appear in the same column (row) of a resolvent diagram. This is equivalent to the determinantal expansion (6.9), but the graphical presentation is more transparent for handling the symmetries which hold between rows and columns due to association. Note the separation of the skew diagram into disjoint right parts is not unique but the resolution is. As a simple example we resolve



$$\simeq [4,2] + [4,1^2] + [3^2] + [3,2,1]. \tag{6.10}$$

Our starting triad then corresponds to a skew diagram with boundaries determined by λ_K and λ_L *and* the specification of a resolvant diagram λ_M. Obviously the resolvent diagram must contain any of the disjoint parts into which the skew diagram may be divided. This is the general triad condition. The general equivalence relations follow by noting that anytime a row or column of the resolvent diagram just matches that of the skew diagram when λ_M is superposed at any of the exterior corners of the skew diagram (λ_M is superposed in an erect position for upper left exterior corners and in an inverted position for lower right exterior corners), then that row or column may be deleted with a corresponding reduction in the component irreps and l or n. The resulting lower rank triad has identical multiplicity resolution as in the initial triad so they belong to the same triad class. The reader should note this includes all the equivalence relations required by duality but is more general than these. In particular this identification of multiplicities allows the equalities (6.8) to be satisfied. As an example the triad with $|[6^6]:[5,4,2^2,1][4,3,1^2][5,3^2,1^2]| = 7$ has skew diagram as shown in the center frame of Fig. 1. Superposing the resolvent irrep $[5,3^2,1^2]$ as indicated shows the fourth row of the diagram may be stripped yielding the triad $|[6^5]:[5,4,2,1][4,3,1][5,3,1^2]| = 7$ of the same class. Clearly the multiplicity resolution for triads obtained by rotating all diagrams by 180° is in one to one correspondence with the original justifying the pure phase form (6.2) adopted for the transposition matrix. Reflecting all diagrams across the main diagonal characterize the triads formed by the associate irreps. The multiplicity resolution is again in one to one correspondence with the original, justifying a pure phase form for the association factor.

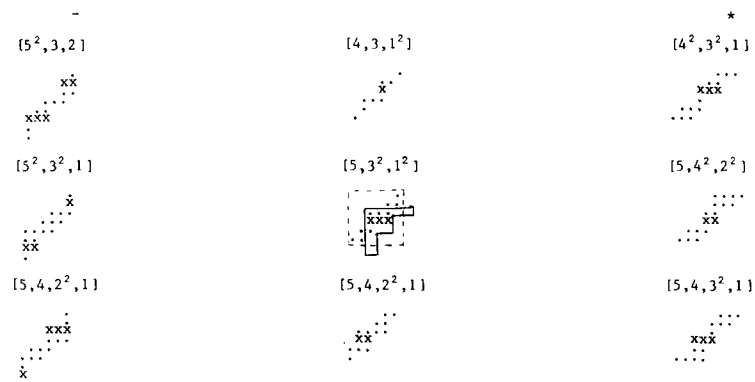
It is not obvious from this graphical presentation, however, that the multiplicity resolution for triads obtained by cyclic permutation of all three component irreps or by complex conjugation are related (not identical) to the original. That such is the case follows as we have already discussed from the theory of S_L and Un. Nor are the equivalence relations which may allow the stripping of rows or columns obvious in these triads obtained by cyclic permutations or complex conjugation. In Fig. 1 we indicate the skew diagrams with resolvent irreps for all triads related to $|[6^6]:[5,4,2^2,1][4,3,1^2][5,3^2,1^2]|$ by cyclic permutations or complex conjugation. We indicate by an x those rows which must strip in order to be consistent with the equivalence reduction already discussed for the central diagram of the figure. Graphically the two processes of complex conjugation denoted by * and — correspond to interchanging the left and right (or upper and lower) boundaries of the skew diagram by a horizontal (or vertical) shift equal to the number of columns (or rows) in the resolvent irrep, respectively. This also transposes the indices K and L in the conjugate triads.

It is easy to verify that the operations of equivalence commute among themselves and with the operations of permutation or complex conjugation. Moreover, permutation commutes with complex conjugation. The two processes of complex conjugation, however, do not commute and used alternately can generate a sequence of triad classes with equi-

FIG. 1. Skew diagrams and resolvent irreps for triads equivalent by cyclic permutation or complex conjugation. × denotes rows which can be stripped as indicated in the center diagram with the resolvent irrep shown overlaid on the skew diagram.

$|[6^6]:[5,4^2,2,1][5^2,3,2][5^2,3^2,1]| \simeq |[6^6]:[5,4,2^2,1][4,3,1^2][5,3^2,1^2]| \simeq |[8^6]:[5,4,3^2,1][4^2,3^2,1][5,4^2,2^2]|$

$|[6^6]:[5,4,2,1][5,3,2][5^2,3,1]| \simeq |[6^5]:[5,4,2,1][4,3,1][5,3,1^2]| \simeq |[8^5]:[5,4,3,1][4^2,3,1][5,4^2,2]|$



valent multiplicity resolution. There must be one triad class of which $l_1(K)+l_1(L)+l_1(M)-l \geqslant l$ and $n_1(K)+n_1(L)+n_1(M)-n \geqslant n$ with all equivalent triad classes being of equal or higher rank. It is convenient to consider the multiplicity resolution as fixed first in the triad of lowest rank and then fixed in equivalent triads of higher rank by requiring real phase form of the transformations relating the irreps.

The Derome–Sharp matrix may be expressed as a recoupling coefficient in SUn and therefore as a DCME of the form

$$A\left(\left[\frac{l^n}{\lambda_3}\right]:\lambda_1\lambda_2\right)_{rr^*}$$

$$\equiv\begin{bmatrix} [(l_1+l_2)^n] & [l_j{}^n] \\ [l^n/\lambda_3] & \lambda_j \\ [(l_1+l_2-l)^n\lambda_3] & [l_j{}^n/\lambda_j] \end{bmatrix}_{r^*}^{r}\left(\frac{\langle\lambda_j\rangle_n}{\langle\lambda_3\rangle_n}\right)^{1/2}. \quad (6.11)$$

This is actually the phase matrix for complex conjugation of a vector coupling coefficient, but since a positive sign can be shown by induction to hold for

$$A\left([l^n]:\lambda\left[\frac{l^n}{\lambda}\right]\right)$$

$$=\begin{bmatrix} [(2l)^n] & [l^n]\otimes[2] & \\ [l^n]\otimes[2] & \lambda & \left[\frac{l^n}{\lambda}\right] \\ & \left[\frac{l^n}{\lambda}\right] & \lambda \end{bmatrix}\langle\lambda\rangle_n = +1, \quad (6.12)$$

one has $A([l^n]:\lambda_1\lambda_2\lambda_3)_{rr^*} = A([l^n/\lambda_3]:\lambda_1\lambda_2)_{rr^*}$ in agreement with the identity immediately following (2.14). Similarly the 1jm factor of SUn is proportional to an isoscalar factor and therefore to a WDCME of the form

$((n_1+n_2)\lambda:n_j\lambda_j)_{r\bar{r}}$

$$\equiv\begin{pmatrix} [l^{(n_1+n_2)}] & [l^{n_j}] \\ \lambda & \lambda_j \\ \left[\frac{l^{(n_1+n_2)}}{\lambda}\right] & \left[\frac{l^{n_j}}{\lambda_j}\right] \end{pmatrix}_{\bar{r}}^{r}\left(\frac{\langle\lambda_j\rangle_{n_j}}{\langle\lambda\rangle_{(n_1+n_2)}}\right)^{1/2}. \quad (6.13)$$

By transposition of the component rows or columns in (6.11) and using $\phi(n\lambda_1)\phi(n\lambda_2)\phi(n\lambda_3) = +1$ for any triad in SUn one has

$$A([l^n]:\lambda_j)_{rr^*} = A([l^n]:\lambda_{\pi j})_{rr^*}$$

$$= A([(l_1+l_2+l_3-l)^n]:[l_j{}^n/\lambda_j])_{r^*r}, \quad (6.14)$$

in agreement with (2.7) and (2.8), where π is any permutation in component order. Similarly transposition of the rows and columns of (6.13) using $\phi(n\lambda) = (-1)^{L(n-1)}$ of the next section gives

$((n_1+n_2)\lambda:n_j\lambda_j)_{r\bar{r}}$

$$= (-1)^{L_1n_2+L_2n_1}\left((n_1+n_2)\left[\frac{l^{(n_1+n_2)}}{\lambda}\right]:n_j\left[\frac{l^{n_j}}{\lambda_j}\right]\right)_{\bar{r}r}$$

$$= (-1)^{ln_1n_2}\Phi(\lambda:\lambda_j)\Phi\left(\left[\frac{l^{(n_1+n_2)}}{\lambda}\right]:\left[\frac{l^{n_j}}{\lambda_j}\right]\right)$$

$$\times((n_1+n_2)\lambda:n_2\lambda_2n_1\lambda_1)_{\bar{r}\bar{r}}. \quad (6.15)$$

The 1jm factor has a form associate to the DCME giving the Derome–Sharp matrix. Since the association factor $(\lambda:\lambda_j)_{r\bar{r}}$ is in real phase form when the transposition phase matrix has the real phase form $\pm E$, then the 1jm factor will be in real phase form iff the Derome–Sharp matrix is also in real phase form. If a triad class and its * complex conjugate class are distinct, separate orthogonal transformations may be used on the multiplicity resolution in the two classes to bring the Derome–Sharp matrix to almost any form, in particular the identity matrix. If a triad class and its * complex conjugate class are identical, the Derome–Sharp matrix is symmetric and orthogonal so it can be brought to diagonal form with ± 1 eigenvalues via an orthogonal similarity transformation on the multiplicity resolution. By restricting ourselves to orthogonal transformations to retain real values we depart from Butler, who uses congruence transformations to always choose the Derome–Sharp factor as the identity. Just as symmetrized products distinguish separate permutation phase species they also separate them in the complex conjugate triads and in the associate triad. In the latter instance one must use the association $\lambda\otimes\sigma = \tilde{\lambda}\otimes\sigma$ or $\tilde{\sigma}$ as L is even or odd.

Assuming real phase form and taking the associate of (6.11) gives via (6.13) an identity which can be separated into a dimensional relation and a phase relation. The dimensional relation is

$$\frac{|[l^n]|(ln-L)!L!}{(ln)!|[l^n/\lambda]|\,|\lambda|}\langle\lambda\rangle_n\langle\tilde{\lambda}\rangle_l = 1, \quad (6.16)$$

which identity is verified by substitution of known dimensional formulas for the irreps. As alluded to above the phase identity relates the Derome–Sharp matrix to the $1jm$ factor via the association factors such that

$$A(n\lambda:\lambda_j)_{rr^*} = (-1)^{L_2(l_1 n - L_1)}([l^n]:[l_j^n])\left([l^n]:\lambda\left[\frac{l^n}{\lambda}\right]\right)$$
$$\times\left([l_j{}^n]:\lambda_j\left[\frac{l^{n_j}}{\lambda_j}\right]\right)(\lambda:\lambda_j)_{r\tilde{r}}$$
$$\times\{[(l_1+l_2)^n/\lambda\,]:[l_j{}^n/\lambda_j\,]\}_{r^*\tilde{r}^*}(l\tilde{\lambda}:l_j\tilde{\lambda}_j)_{\tilde{r}\tilde{r}}, \quad (6.17)$$

where $\tilde{r}^* = \bar{\tilde{r}}$ and $\bar{\tilde{r}} = \tilde{r}^*$.

A question of consistency arises only in case a triad, its complex conjugate, and its associate all become equivalent so the triad classes are identical and there is only one multiplicity space in which orthogonal transformations may be carried out. In this case the transposition phase matrix, the Derome–Sharp matrix, and the $1jm$ factor are all symmetric orthogonal matrices, while the association factor is symmetric or skew symmetric as

$$(-1)^{\binom{l}{2}}\theta(\lambda_1)\theta(\lambda_2)\theta(\lambda_3)$$

is even or odd [$\binom{l}{2}$ is the binomial coefficient]. Putting the transposition phase matrix in the form $\pm E$ and the Derome–Sharp matrix in diagonal form with eigenvalues ± 1 essentially limits any further manipulation of the multiplicity space to orthogonal similarity transformations on multiplicities with the same eigenvalue in the Derome–Sharp matrix. For this case $n = l$ and $\lambda_j \simeq \tilde{\lambda}_j \simeq [l_j{}^l/\lambda_j] \simeq [l^{l_j}/\lambda_j]$, where the last equivalence is not independent but follows from the first three. These equivalences require the Young frames labelling each irrep have regular step boundaries, i.e., the irrep frame itself can be considered to be composed of square subarrays $c_j b$ nodes on a side to give the partition label $\lambda_j = [(kc_j b)^{c_j b}]$ with k taking on all integral values below some upper limit $p_j \geqslant k \geqslant 1$. Each component irrep λ_j has integer parameters p_j and c_j with integer b being a scaling factor common to the triad so that

$$L_j = \binom{p_j+1}{2} c_j{}^2 b^2.$$

To form a triad one must have

$$l = (p_j+1)c_j b = (p_1 c_1 + p_2 c_2 + p_3 c_3)b/2, \quad 1 \leqslant j \leqslant 3. \quad (6.18)$$

There are only three solutions to (6.18), each corresponding to a triad of type I, II, or III. These solutions and their characteristic triads ($b = 1$) are listed in Table II. For type I and II the association factor is symmetric or skew symmetric as the scaling factor b is even or odd. For type III triads it is always symmetric. Using the phase convention proposed in the next section one can show the transposition phase is self-associate for types I and II and is consistent with the value ϕ_σ for type III. Relation (6.17) for this case takes the special form

$$\begin{bmatrix} [(l_1+l_2)^l] & [l_j{}^l] \\ \lambda\otimes\sigma & \lambda_j\otimes\sigma_j \end{bmatrix}_{r r^*}$$
$$= (\lambda:\lambda_j)_{r\tilde{r}}(\lambda:\lambda_j)_{r^*\tilde{r}^*}\begin{bmatrix} [l^{(l_1+l_2)}] & [l^{l_j}] \\ \lambda\otimes\sigma' & \lambda_j\otimes\sigma_j{}' \end{bmatrix}_{\tilde{r}\,\tilde{r}^*}$$
$$= \begin{bmatrix} [l^{(l_1+l_2)}] & [l^{l_j}] \\ \lambda\otimes\sigma' & \lambda_j\otimes\sigma_j{}' \end{bmatrix}_{\tilde{r}} \delta_{\tilde{r}\tilde{r}^*}\delta_{rr^*}, \quad (6.19)$$

where the last equality holds when the Derome–Sharp matrix is in diagonal form. Here $\sigma'(\sigma_j{}')$ is the same as $\sigma(\sigma_j)$ if $L\,(L_j)$ is even, or is the associate of $\sigma(\sigma_j)$ if $L\,(L_j)$ is odd. The σ's are related by the inner product so that $|\sigma:\sigma_1 \circ \sigma_2| > 0$. Equation (6.19) identifies the Derome–Sharp phase with the $1jm$ phase for type I and II triads of this case. For type III triads an additional phase factor $(-1)^b$ must be included on the r.h.s. of (6.19). So even in the most restrictive case a real phase form convention consistent with duality is possible.

In summary we conclude it is possible to have a phase convention with the transposition matrix, the Derome–Sharp, the $1jm$ factor, and the association factor all in real phase form and consistent with duality. The latter requires a transposition phase independent of multiplicity satisfying equivalence relations

$$\Phi(\lambda:\lambda_1\lambda_2) = \Phi(\lambda:\lambda_2\lambda_1) = \Phi([(l_1+l_2)^n\lambda\,]:[l_j{}^n\lambda_j\,])$$
$$= (-1)^{ln_1n_2}\Phi([l^{(n_1+n_2)}\lambda\,]:[l^{n_j}\lambda_j\,])$$
$$= \Phi\left(\begin{bmatrix}(l_1+l_2)^n \\ \lambda\end{bmatrix}:\begin{bmatrix}l_j{}^n \\ \lambda_j\end{bmatrix}\right)$$
$$= (-1)^{ln_1n_2+n_1L_2+n_2L_1}\Phi\left(\begin{bmatrix}l^{(n_1+n_2)} \\ \lambda\end{bmatrix}:\begin{bmatrix}l^{n_j} \\ \lambda_j\end{bmatrix}\right). \quad (6.20)$$

The $1j$ phase is given by $\phi(n\lambda) \equiv \Phi([l^n]:\lambda\,[l^n/\lambda\,])$ such that $\phi(n\lambda_1)\phi(n\lambda_2)\phi(n\lambda_3) = +1$ for any triad. Under complex conjugation of the irreps the transposition phase satisfies

TABLE II. Triads $([l^l]:\lambda_j)$ with equivalence under association and complex conjugation $\lambda_j \simeq \tilde{\lambda}_j \simeq [l_j{}^l/\lambda_j]$ parametrized by $\lambda = [(kcb)^{cb}]$, $p \geqslant k \geqslant 1$, b an arbitrary scale factor. Characteristic triads ($b = 1$) are given for each case with $3j$ permutation phase ϕ, association phase θ, and the labels m_{ij} which distinguish the multiplicity for the characteristic triad.

Type	(p_j) (c_j)	$([l^l]:\lambda_j)$	ϕ	θ	Multiplicity labels m_{ij} r	s^*
I	(1,2,5) (3,2,1)	$([6^6]:[3^3][4^2,2^2][5,4,3,2,1])$	$(-1)^b$	$(-1)^b$	$m_{23}=1=m_{34};m_{24}=1=m_{33}$	$m_{45}=1=m_{34};m_{35}=1=m_{44}$
II	(1,3,3) (2,1,1)	$([4^4]:[2^2][3,2,1]\otimes[2])$	$+1$	$(-1)^b$	$m_{23}=1=m_{34};m_{24}=1=m_{33}$	$m_{23}=1=m_{12};m_{13}=1=m_{22}$
III	(2,2,2)	$([3^3]:[2,1]\otimes\nu)$	ϕ_ν	$+1$	$m_{12}=1=m_{23};m_{13}=1=m_{22}$	$m_{23}=1=m_{12};m_{13}=1=m_{22}$

$$\Phi(\lambda:\lambda_j) = \Phi([(l_1 + l_2)^n/\lambda\,]:[\,l_j{}^n/\lambda_j\,]) \tag{6.21}$$

and under association one has

$$\Phi(\lambda:\lambda_j) = (-1)^{L_1 L_2}(\lambda:\lambda_1\lambda_2)(\lambda:\lambda_2\lambda_1)\Phi(\tilde{\lambda}:\tilde{\lambda}_j). \tag{6.22}$$

Moreover, the ljm factor is related to the Derome–Sharp phase by the association relation (6.17).

VII. THE PHASE CONVENTION

In this section we propose a convention for the association phase factor and for the transposition phase which satisfies all the requirements established in the previous section. To do this we extend our notation to give a more detailed description of carrying out an outer product according to the Littlewood–Richardson rules. Let $m_{ij}(p)$, $i \leqslant j$, be the number of nodes transferred from the ith row of the component irrep λ_p to row j of the composite irrep at the ith stage of carrying out the Littlewood–Richardson rules. In this notation specification of the set $\{m_{ij}(p)\}$ implicitly distinguishes all multiplicities by specifying the composite irrep at each stage of the outer product so the index r may be dropped in writing a triad as

$$(\lambda:\lambda_1\lambda_2 r) = \left(\left[\sum_i m_{ij}(1) + \sum_k m_{jk}(2)\right]:\left[\sum_k m_{jk}(1)\right]\left[\sum_k m_{jk}(2)\right]\right), \tag{7.1}$$

where $\lambda_p = [l_j(p)]$ with $l_j(p) = \Sigma_k m_{jk}(p)$. To satisfy the Littlewood–Richardson rules we assume the component irrep tableaus are in standard order and each row application is carried out in a standard sequence. The standard sequence applies the nodes m_{ij} in order starting with lowest i, proceeding from highest to lowest j, and then advancing in numerical order on i. The product tableau will be lexical at all stages. The remaining outer product rules are satisfied iff

$$\sum_k m_{jk}(p') + \sum_i^l m_{ij}(p) \leqslant \sum m_{(j-1)k}(p') + \sum_i^{l-1} m_{i(j-1)}(p) \tag{7.2}$$

and

$$\sum_j^l m_{ij}(p) \geqslant \sum_j^{l-i} m_{(i+1)j}(p) \tag{7.3}$$

for any value l.

The convention for the association phase factor is a direct extension of that used by Hamermesh (viz. p. 266) which assigns ± 1 based on the parity of the permutation required to rearrange a Young tableau to standard order. This procedure correctly gives the required sign $\theta(\lambda)$ for self-associate irreps. The permutation to standard order may be factored into permutations which bring the composite irrep to standard order after each stage of one row applications in the process of forming the outer product. So we may proceed by iterating the result for a one row application which is accomplished by successive cyclic permutations $(\Sigma_{k>j} l_k - \Sigma_{i<j} m_i)^{m_j}$ to give the association phase

$$\left([l_j + m_j]:[l_j]\left[\sum m_i\right]\right) = (-)\left[\sum_i m_i \sum_{j>i}(l_j + m_j)\right]. \tag{7.4}$$

In this secion we use a notation $(-)(\text{argument}) = \pm 1$ as the argument is even or odd. This iterates in the general case to

$$\left(\left[\sum_k m_{jk}(p') + \sum_i m_{ij}(p)\right]:\left[\sum_k m_{jk}(p')\right]\left[\sum_k m_{jk}(p)\right]\right) = (-)\left(\sum_{i<j<k} m_{ij}(p)\left(\sum_l m_{kl}(p') + \sum_{h<i} m_{hk}(p)\right)\right), \tag{7.5}$$

which is the convention for determining any association phase. Particularly important association phase factors involve coupling to irreps with rectangular Young frames for which

$$m_{ij} = m_{1(j-i+1)} = l_{j-i} - l_{j-i+1} \quad \text{with } l_0 \equiv l_1$$

so that the association phase factor reduces to

$$([l^n]:[l_j]\,[l - l_{n-j+1}\,]) = (-)\left((l-1)\sum(n-i)l_i\right). \tag{7.6}$$

One also has the useful relation under equivalence

$$\left(\begin{bmatrix} l'^{n'} \\ \lambda \end{bmatrix}:\begin{bmatrix} l'^{n'} \\ \lambda_1 \end{bmatrix}\lambda_2\right) = (\lambda:\lambda_1\lambda_2). \tag{7.7}$$

The convention we propose for the transposition phase is

$$\Phi(\lambda:\lambda_j) = (-)\left(\sum_{i,j}(j-i)m_{ij}\right), \tag{7.8}$$

which reduces to the Condon–Shortley convention for SU2 in the case of two rowed irreps. Because m_{ii} does not contribute to the phase, the standard convention of positive phase for maximum coupling ($m_{ij} = \delta_{ij} l_j$) is obtained and the invariances

$$\Phi(\lambda:\lambda_j) = \Phi([(l_1 + l_2)^n\lambda\,]:[\,l_j^n\lambda_j\,]) = \Phi\left(\begin{bmatrix} (l_1 + l_2)^n \\ \lambda \end{bmatrix}:\begin{bmatrix} l_j^n \\ \lambda_j \end{bmatrix}\right) \tag{7.9}$$

are trivially satisfied. A short calculation also shows

$$\Phi(\lambda:\lambda_j) = (-1)^{ln_1n_2}\Phi([\,l^{(n_1+n_2)}\lambda\,]:[\,l^{n_j}\lambda_j\,]) = (-1)^{ln_1n_2 + n_1L_2 + n_2L_1}\Phi\left(\begin{bmatrix} l^{(n_1+n_2)} \\ \lambda \end{bmatrix}:\begin{bmatrix} l^{n_j} \\ \lambda_j \end{bmatrix}\right) \tag{7.10}$$

and

$$\Phi([l^n]:\lambda\,[l^n/\lambda\,]) = (-)(L(n-1)) \equiv \phi(n\lambda) = \phi(n\lambda^*) \tag{7.11}$$

so $\phi(n\lambda_1)\phi(n\lambda_2)\phi(n\lambda_3) = +1$ for any triad in SUn as required. That the transposition phase is independent of the multiplicity or the component order follows from expressing the product irrep λ in either of two ways,

$$\lambda = \left[\sum_k m_{jk}(p') + \sum_i m_{ij}(p)\right] = \left[\sum_k m_{jk}(p) + \sum_i m_{ij}(p')\right]. \tag{7.12}$$

Multiplying by j and then summing on this index, transposing two of the sums, and redefining some summation indexes shows $\Sigma(j-i)m_{ij}(p)$ is independent of p and hence is independent of both the multiplicity and the component order. Examination of the coupling process for $([(l_1 + l_2)^n/\lambda\,]:[\,l_2^n/\lambda_2][\,l_1^n/\lambda_1])$ in terms of the coupling for $(\lambda:\lambda_1\lambda_2)$ leads to the

identification

$$m_{ij}^{*}(p') = m_{(n-j+1)(n-i+1)}(p) + \delta_{ij}(l_1 + l_2 - l_j(p')), \tag{7.13}$$

which essentially gives the multiplicity mapping under complex conjugation and transposition. Using (7.13) in (7.8) shows

$$\Phi(\lambda:\lambda_j) = \Phi([(l_1 + l_2)^n/\lambda\,]:[l_j^n/\lambda_j\,]). \tag{7.14}$$

Finally, we must show the phase convention (7.8) is compatible with association. We proceed by iterating the result for a one row or column application

$$\Phi(\lambda:\lambda_1[l\,]) = (-1)^{lL_1}(\lambda:\lambda_1[l\,])(\lambda:[l\,]\lambda_1)\Phi(\tilde{\lambda}:\tilde{\lambda}_1[1^l\,]). \tag{7.15}$$

To prove this result we note $(\lambda:[l\,]\lambda_1)$ has $m'_{ij} + m_{j(j+1)} = l_j$ and $m'_{j(j+1)} = \Sigma_{k>j} m_k$ so that

$$\left([l_j + m_j\,]:[l_j\,]\left[\sum m_i\right]\right)\left([l_j + m_j\,]:\left[\sum m_i\right][l_j\,]\right) = (-)\sum_j\left(l_j m_j + \sum_{i<j} m_i m_j + l l_j + (j-1)m_j\right). \tag{7.16}$$

Similarly

$$\Phi(\tilde{\lambda}:\tilde{\lambda}_1[1^l\,]) = (-)\left(\sum l_j m_j - \sum_{i<j} m_i m_j\right). \tag{7.17}$$

Substitution of (7.8), (7.16), and (7.17) verifies (7.15). For iterating to the general case, we factor each association phase to $(\lambda:\lambda_1\lambda_2 r) = \Pi_i(\lambda_1{}^i:\lambda_1{}^{(i-1)}[l_i(2)\,])$ and each transposition phase to

$$\Phi(\lambda:\lambda_j) = \Pi_i(-)^{(i-1)l_i(2)}\Phi(\lambda_1{}^i:\lambda_1{}^{(i-1)}[l_i(2)\,]),$$

where $\lambda_1{}^i$ is the intermediate composite irrep obtained after the ith application in carrying out the outer product and the first factor in the transposition phase adjusts the row count to start at i for the ith application. Analogous consideration of $(\lambda_1{}^i:[l_i(2)]\lambda_1{}^{(i-1)})$ along with the identities

$$m^i_{ij} + m^i_{j(j+1)} = l_j{}^{i-1} = l_j(1) + \sum_{h<i} m_{hj}$$

and

$$m^i_{j(j+1)} = l_i(2) - \sum_{h<j} m_{ih} = \sum_{k>j} m_i$$

gives

$$(\lambda_1{}^i:\lambda_1{}^{i-1}[l_i(2)\,])(\lambda_1{}^i:[l_i(2)]\lambda_1{}^{i-1}) = (-)\left(L_1 l_i(2) - \sum_j l_j(1)m_{ij} - \sum_j m_{ij}(j-1) + \sum_{h<i<j<k} m_{ij}m_{hk} + \sum_{\substack{h<i<j \\ k<j}} m_{ij}m_{hk}\right). \tag{7.18}$$

Similarly

$$\Phi(\tilde{\lambda}_1{}^i:\tilde{\lambda}_1{}^{(i-1)}[1^{l_i(2)}]) = (-)\sum_k\left(\left(l_k{}^{(i-1)} - \sum_{j<k} m_{ij}\right)m_{ik}\right) = (-)\sum_k\left(\left(l_k(1) + \sum_{j<i} m_{jk} - \sum_{j<k} m_{ij}\right)m_{ik}\right). \tag{7.19}$$

Using $\Phi(\tilde{\lambda}:\tilde{\lambda}_j) = \Pi_i\Phi(\tilde{\lambda}_1{}^i:\tilde{\lambda}_1{}^{(i-1)}[1^{l_i(2)}])$, $(\lambda:\lambda_2\lambda_1 s) = \Pi_i(\lambda_1{}^i:[l_i(2)]\lambda_1{}^{(i-1)})$ and the above identities one verifies

$$\Phi(\lambda:\lambda_j) = (-1)^{L_1L_2}(\lambda:\lambda_1\lambda_2 r)(\lambda:\lambda_2\lambda_1 s(r))\Phi(\tilde{\lambda}:\tilde{\lambda}_j). \tag{7.20}$$

Thus the phase conventions (7.5) and (7.8) satisfy all the requirements established in the previous section for a real phase convention compatible with all the symmetries of duality. For $\lambda_1 = \lambda_2$ the transposition phase (7.8) has been found to correctly distinguish the symmetrized product frequencies $|\lambda:\lambda_1 \otimes [2]|$ and $|\lambda:\lambda_1 \otimes [1^2]|$ in all cases examined, but we can offer no independent proof of this assertion. In the case of unsymmetrized products $|\lambda:\lambda_1\lambda_1| > 1$ the convention (7.8) will give the same phase for all multiplicities but it is only by a transformation to the symmetrized powers that the transposition matrix is brought to the diagonal form $\Phi(\lambda:\lambda_1 \otimes \sigma) = \phi_\sigma$. We illustrate this point by using the characteristic triad of type III from Table II, ([3,2,1]:[2,1][2,1]), with multiplicity two. All phase matrices are symmetric with

$$\Phi = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = A$$

and the association matrix

$$([3,2,1]:[2,1]^2) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

such that the $1jm$ factor is also off-diagonal given by

$$(-1)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Transformation to symmetrized products ([3,2,1]:[2,1] ⊗ [2]) and ([3,2,1]:[2,1] ⊗ [1²]) simultaneously diagonalizes Φ, A, and the $1jm$ factor to the form

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

while the association factor becomes nondiagonal of the form

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

such that

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

as required by (6.19). It is not possible to simultaneously diagonalize all four phase matrices. Incidentally $A([3,2,1]:[2,1] \otimes [1^2]) = -1$ provides us with the first nontrivial case in which the convention we propose requires a negative value for the Derome–Sharp factor. The unsymmetrized form of the phase matrices for the characteristic triad of type II of Table II, ([4²,2²]:[3,2,1][3,2,1]), has

$$A(4[4^2,2^2]:[3,2,1][3,2,1]) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \Phi([4^2,2^2]:[3,2,1][3,2,1])$$

$$= (6[4^2,2^2]:3[3,2,1]3[3,2,1])$$ and association factor

$$([4^2,2^2]:[3,2,1][3,2,1]) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

These latter forms remain invariant under transformation to

symmetrized products because the multiplicity is within the same symmetrized species, i.e., $|[4^2,2^2]:[3,2,1] \otimes [2]| = 2$ and $|[4^2,2^2]:[3,2,1] \otimes [1^2]| = 0$.

Since the composition of this paper Bickerstaff *et al.*[19] have published tables of $3j$ phases and $6j$ coefficients for SU3 and SU6, and $3jm$ factors for SU3/U1$\times$SU2 and SU6/SU3$\times$SU2. These tables exhibit symmetries of duality, the last factor involving duality with inner products of S_L. The phases, however, have not been chosen to be consistent with the conventions proposed here. For SU3 there is no distinction between Butler's "sensible" choice of phase in (2.10) and our choice since the $1j$ phase is always positive. Our convention (7.8) agrees with their $3j$ Table 3 with due regard for symmetrized powers. The one exception to this agreement is in the resolution of the multiplicity for the triad $|[3,2][3,1][2,1]:[4^3]|$ for which (7.8) gives a positive phase for both multiplicities while they list opposite phases. The equivalence considerations of Sec. VI show the triad multiplicity is in the same triad class with $|[2,1]^3:[3^3]|$ and thus can be separated as with the symmetrized products for that case, but there is no symmetrization that is required in this instance. As noted above we require $A([2,1] \otimes [1^3]:[3^3]) = -1$ whereas these authors always choose $A = +1$. If our conventions were used [i.e., a negative A for their triads $([2,1]^3)$ and $([3,2][3,1][2,1]1)$] all SU3 $6j$ coefficients listed in their Table 4 become real. This same choice of A will make all $3jm$ factors for SU3/U1$\times$SU2 listed in their Table 10 real if, in addition, one chooses the $1jm$ factor

$$\begin{pmatrix} [2^3] & [2] & [2^2] \\ [2,1] & [1] & [2] \\ [2,1] & [1] & [2] \end{pmatrix}$$

$\left[\binom{[2,1]}{[2]}\text{in their notation}\right]$ as negative. This latter choice is required of us by duality as shown by considering the associate form. Their $3j$ phases for SU6 differ in several instances from our convention (7.8). Although we have made no detailed analysis of the situation, it seems plausible from the occurrence of only purely real or purely imaginary coefficents in their tables that all values could be brought to real form by a suitable choice of phase.

VIII. CONCLUSION

Excluding the nonsimple phase case $0 \in \lambda \otimes [2,1]$ we argue that it is possible to arrange a separation of multiplicities and a phase convention so that the symmetries derived from duality with S_L may be required of the coupling algebra of SUn. All coupling coefficients can be made real and all phase transformation matrices can be brought to real phase form. We have listed the properties the phases related to transposition, association, the Derome–Sharp matrix, and the $1jm$ factor must have to be consistent with duality. We have proposed a transposition phase convention (7.8) along with the direct extension of the usual association phase convention (7.5) which possess these properties. The $1jm$ factor phase convention has been related to the Derome–Sharp matrix phase convention via association (6.17). We have not developed a convention similar to (7.5) or (7.8) for assigning either the $1jm$ factor or the Derome–Sharp matrix directly. To fix a complete phase convention it would seem necessary to have a prescription for one or the other factor. We have determined that the simple choice $A = +E$ is not consistent with duality. We consider the convention fixing A or the $1jm$ factor to be an unresolved question. Finally, because of the noncommutitivity of the two operations of complex conjugation, we can offer no direct proof that no closed transformation loop exists which is inconsistent with the proposed phase convention. The requirements enumerated at the close of Sec. VI at least are necessary conditions for a phase convention consistent with duality between S_L and SUn.

ACKNOWLEDGMENT

The author gratefully acknowledges discussion with P. H. Butler on the question of phase consistency on restricting from Un to SUn.

[1] I. Schur, Sitzungsber. K. Preuss. Akad. Wiss. 58 (1927); 100 (1928); H. Weyl, *The Classical Groups* (Princeton U. P., Princeton, 1946).
[2] A. P. Stone, Proc. Cambridge Philos. Soc. **57**, 3, 46, 469 (1961).
[3] J. J. Sullivan, J. Math. Phys. (a) **14**, 387 (1973); (b) **16**, 756 (1975); (c) **16**, 1707 (1975); (d) **19**, 1674 (1978); and (e) *Proceedings of the International Symposium on Mathematical Physics*, Mexico City, Jan. 5–8, 1976, Vol. I, p. 253.
[4] P. H. Butler, Philos. Trans. R. Soc. London **277**, 545 (1975).
[5] J. R. Derome, J. Math. Phys. **7**, 612 (1966); **8**, 714 (1967).
[6] C. Frobenius and I. Schur, Berl. Ber. 186 (1906).
[7] J. R. Derome and W. T. Sharp, J. Math. Phys. **6**, 1584 (1965).
[8] D. M. Brink and G. R. Satchler, *Angular Momentum* (Clarendon, Oxford, 1962).
[9] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge U.P., New York, 1935).
[10] G. W. Mackey, *The Theory of Group Representations*, University of Chicago Lecture Notes (University of Chicago Press, Chicago, 1955).
[11] F. D. Murnaghan, *The Theory of Group Representations* (John Hopkins, Baltimore, MD, 1938).
[12] M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1962).
[13] P. H. Butler and R. C. King, J. Math. Phys. **14**, 1176 (1973).
[14] J. Quan Chen, J. Math. Phys. **22**, 1 (1981).
[15] P. Kramer, Z. Phys. **205**, 181 (1967); **216**, 68 (1968).
[16] P. Kramer and T. H. Seligman, Z. Phys. **219**, 105 (1969).
[17] A. A. Jucys, Liet. Fiz. Rinkinys (Lith. Phys. Papers) **9**, 629 (1969); **10**, 5 (1970).
[18] G. de B. Robinson, *Representation Theory of the Symmetric Group* (U. of Toronto, Toronto, 1961).
[19] R. P. Bickerstaff, P. H. Butler, M. B. Butts, R. W. Haase, and M. F. Reid, J. Phys. A **15**, 1087 (1982).

The Racah algebra for groups with time reversal symmetry. III

J. D. Newmarch[a)]
Department of Physics, Universiti Pertanian Malaysia, Serdang, Selangor, Malaysia

R. M. Golding
Department of Physical Chemistry, University of New South Wales, PO Box 1, Kensington, New South Wales 2033, Australia

(Received 21 July 1981; accepted for publication 18 September 1981)

The coupling of basis vectors and the recoupling of irreducible corepresentations is considered. The difference between the coupling coefficient and $3jm$ tensors leads to a variety of $6j$ and recoupling tensors.

PACS numbers: 02.20.Hj, 03.65.Fd

1. INTRODUCTION

In two previous papers,[1,2] henceforth denoted as I and II respectively, the elements of the Racah algebra for compact groups with time reversal symmetry were developed. These groups contain antilinear operators and it was found convenient to use Wigner's theory of corepresentations[3] to achieve a proper group-theoretical treatment. In many respects this algebra parallels that for linear groups in that symmetrized $3jm$ symbols can be found and both the Wigner–Eckart theorem and Racah's lemma hold. In this paper the theory is further developed by considering the coupling of basis vectors and various possibilities for the n-j symbols. Throughout the notations of I and II are used.

Despite the many similarities that exist between representation and corepresentation theory, an important distinction arises at an early stage, in Schur's Lemmas: whereas a matrix commuting with an irreducible representation (IR) is constant, one commuting with an irreducible corepresentation (ICR) need not be. The set of all such commuting matrices forms an algebra isomorphic to $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Q}$.[4] This has many consequences, among which is the fact that the multiplicity of j_3 in $j_1 \otimes j_2$ may be less than the multiplicity of the identity ICR $\mathbf{0}$ in $j_1 \otimes j_2 \otimes j_3$ (these groups are ambivalent). Thus the coupling coefficient tensor which reduces $j_1 \otimes j_2$ to j_3 may have a smaller multiplicity range to the $3jm$ tensor which reduces $j_1 \otimes j_2 \otimes j_3$ to $\mathbf{0}$. There is, then, an essential difference between the two tensors.

In any applications of Racah algebra methods which do not trivially involve the definitions of these tensors, it must first be established which tensor is used. Thus in I $3jm$ tensors were used in the Wigner–Eckart theorem and for Racah's lemma (although considerably less tractable forms exist using the coupling coefficient.[5–7]) After a brief review in Sec. 2 of the tensor notation and a minor correction to the Wigner–Eckart theorem in Sec. 3, Sec. 4 considers the coupling of basis vectors. Some results peculiar to corepresentation theory occur here. This is also true in Sec. 5, where the transformation properties of the two tensors under basis transformations are given. These two sections further illustrate the differences between the coupling coefficient and $3jm$ tensors and in particular the care that must be exercised in using the coupling coefficient to avoid undesirable results.

Following on from the $3jm$ tensor, the next major symbol is the $6j$. Due to its importance in representation theory, a variety of different approaches to it have been devised: as the recoupling coefficient of three IRs to a fourth, as the recoupling coefficient of four IRs to the identity $\mathbf{0}$, as the first nontrivial invariant under basis transformations, and as the coefficient occurring in certain coupled tensor operator schemes. In Secs. 6–8, respectively, the first three of these approaches are considered for the gray groups and three distinct symbols are found, each made up of a different number of coupling coefficient and $3jm$ tensors. Each of these symbols is shown to possess certain properties but none of them exhibit the full range of properties of the essentially unique symbol in representation theory.[8] Indeed, the identities which use both permutational symmetry and orthogonality of coupling coefficient tensors (such as the Biedenharn identity and the Racah backcoupling rule) are in doubt for all of these symbols. Schur's lemma has some far-reaching effects!

In Sec. 9 a coupled tensor operator scheme is discussed which yields a type of $9j$ symbol. Special cases which in representation theory yield a $6j$ symbol are also dealt with.

To illustrate the theory given here and in I and II a particular gray group is dealt with in detail in Sec. 10. This gray group possesses a $3jm$ tensor with $\{21\}$ symmetry and although the group is not physically important it provides the first such numerical example to be given (at the time of writing).

2. THE TENSOR NOTATION

We assume a complex vector space S with basis $\{e_{(m)}: m = 1,2,\ldots\}$. The operation of complex conjugation may be applied to these basis vectors to give $\{e_{(\dot{m})}: \dot{m} = 1,2,\ldots\}$ as a basis for S^*. The dot above the suffix denotes complex conjugation. The following summation convention is used: a summation is implied over two identical indices if they are (a) one inner and one outer, (b) one upper and one lower, and (c) both dotted or both undotted. The matrix of a linear operator is given by

$$ue_{(n)} = e_{(m)}\, j(u)^m{}_n \tag{2.1}$$

and of an antilinear operator by

[a)] Present address: School of Electrical Engineering and Computer Science, University of New South Wales, P.O. Box 1, Kensington, New South Wales, Australia.

 0022-2488/83/030441-10$02.50

$$ae_{(n)} = e_{(m)} j(a)^m{}_{\dot n}. \tag{2.2}$$

Using the dot notation for complex conjugation,

$$j(u)^m{}_n{}^* = j(u)^{\dot m}{}_{\dot n} \quad \text{and} \quad j(a)^m{}_{\dot n}{}^* = j(a)^{\dot m}{}_n. \tag{2.3}$$

The transpose is obtained by changing inner indices to outer indices and vice versa. This may be accomplished by use of the tensors $\delta^{m_1}_{m_2}$ and $\delta^{\ m_1}_{m_2}$:

$$e_{(m_1)} = e^T{}_{(m_1)} = e_{(m_2)}\delta^{m_2}_{m_1}, \tag{2.4}$$

$$j(u)_{n_1}{}^{m_1} = j(u)^{m_1}{}_{n_1}{}^T = j(u)^{m_2}{}_{n_2}\,\delta^{\ m_1}_{m_2}\delta^{n_2}_{n_1}, \tag{2.5}$$

and

$$j(a)_{\dot n_1}{}^{m_1} = j(a)^{m_2}{}_{\dot n_2}\delta^{\ m_1}_{m_2}\delta^{\dot n_2}_{\dot n_1}. \tag{2.6}$$

Using an orthonormal basis, the metric for S is

$$e^T{}_{(\dot m)}e_{(n)} = e_{(\dot m)}e_{\ (n)} = \delta_{\dot m\ n} = \begin{pmatrix} 1 & & 0 \\ & 1 & \\ & & 1 \\ 0 & & \ddots \end{pmatrix}. \tag{2.7}$$

This may be combined with the transposition tensor to produce other δ tensors:

$$\delta^{\dot m_1 m_2} = \delta^{m_3\ \dot m_1}\delta^{\ m_2}_{m_3}, \tag{2.8}$$

and so on. This allows the unitary properties of the linear operators to be written in a variety of forms:

$$j(u)_{\dot n_1}{}^{\dot m_1}\delta_{\dot m_1\ m_2}\,j(u)^{m_2}{}_{n_2} = \delta_{\dot n_1\ n_2}, \tag{2.9}$$

$$j(u)^{\dot m_1}{}_{\dot n_1}\delta^{\dot n_1 n_2}j(u)^{m_2}{}_{n_2} = \delta^{\dot m_1 m_2}, \tag{2.10}$$

and

$$(j(u)^{m_1}{}_{n_1})^{-1} = j(u)^{\dot m_2}{}_{\dot n_2}\delta^{n_1\dot n_2}\delta_{m_1\dot m_2}, \tag{2.11}$$

and the antiunitary properties of the antilinear operators as

$$j(a)_{n_1}{}^{\dot m_1}\delta_{\dot m_1\ m_2}\,j(a)^{m_2}{}_{\dot n_2} = \delta_{n_1\ \dot n_2} \tag{2.12}$$

$$j(a)^{\dot m_1}{}_{n_1}\delta^{n_1\dot n_2}j(a)^{m_2}{}_{\dot n_2} = \delta^{\dot m_1 m_2} \tag{2.13}$$

and

$$(j(a)^{m_1}{}_{\dot n_1})^{-1} = j(a)^{\dot m_2}{}_{n_2}\delta^{\dot n_1 n_2}\delta_{\ m_1\dot m_2}. \tag{2.14}$$

Two further δ tensors were given in I, namely $\delta^{r_1}{}_{r_2}$ and $\delta^{r_1}{}_{\dot r_2}$. These are distinct from those already given in that they are really shorthand for the direct sum of copies of the identity ICR **0** which always has value 1:

$$\delta^{r_1}{}_{r_2} = \begin{pmatrix} \mathbf{0}(u) & & \\ & \mathbf{0}(u) & \\ & & \ddots \end{pmatrix}^{r_1}_{r_2} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & \ddots \end{pmatrix}^{r_1}_{r_2} \tag{2.15}$$

and

$$\delta^{r_1}{}_{\dot r_2} = \begin{pmatrix} \mathbf{0}(a) & & \\ & \mathbf{0}(a) & \\ & & \ddots \end{pmatrix}^{r_1}_{\dot r_2} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & \ddots \end{pmatrix}^{r_1}_{\dot r_2}. \tag{2.16}$$

It is understood that no change of basis is to be applied to these tensors as the antilinearity of a would change $\delta^{r_1}{}_{\dot r_2}$ [I, Eq. (2.7)].

For definitions and properties of the symbols particular to Racah algebra, the reader is referred to I. We note only that in the definitions (5.1), (5.2), (7.1), and (7.2) of I an abuse of notation was committed by denoting the inverse of $(j_1j_2j_3)^r{}_{m_1m_2m_3}$ by $(j_1j_2j_3)^{m_1m_2m_3}{}_r$ and similarly for the coupling coefficient, but that this is consistent with the tensor notation and affords considerable simplification in many expressions.

3. THE WIGNER–ECKART THEOREM

A transposition tensor should have been included in Eqs. (9.1) and (9.2) of I. If the identifications

$$e_{\ (m)} = |jm\rangle$$

and

$$e_{(\dot m)} = \langle jm|$$

are made, then

$$|u(jm)\rangle = j(u)^n{}_m\,|jn\rangle \tag{3.1}$$

and

$$\langle u(jm)| = j(u)_{\dot m}{}^{\dot n}\langle jn| = \delta^{\dot m_1}_{\dot m}\delta^{\ \dot n}_{\dot n_1}\,j(u)^{\dot n_1}{}_{\dot m_1}\langle jn|. \tag{3.2}$$

Thus Eq. (9.1) of I should read

$$\langle j_1m_1|T(kq_1)|j_2m_2\rangle = \frac{2}{|G|}\int_H \delta^{\ \dot n_1}_{\dot n_3}\,j_1(u)^{\dot n_3}{}_{\dot m_3}\,k(u)^{q_2}{}_{q_2}\,j_2(u)^{n_2}{}_{m_2}\delta^{\dot m_3}_{\dot m_1} \times\langle j_1n_1|T(kq_2)|j_2n_2\rangle\,du, \tag{3.3}$$

with a similar change in Eq. (9.2). The complex conjugate of Eq. (6.6) of I was then used to transform j_1^* to j_1. It turns out that this introduces an annoying 1-j phase factor into coupled tensor schemes (cf. Sec. 9). It is better to use the *inverse* of this equation

$$j(u)^{\dot n_1}{}_{\dot n_2} = \begin{bmatrix} \dot n_1 & \\ & m_1 \end{bmatrix} j(u)^{m_1}{}_{m_2}\begin{pmatrix} m_2 & \\ & \dot n_2 \end{pmatrix}. \tag{3.4}$$

These modify Eqs. (9.3) and (9.4) of I to

$$\langle j_1\|T(k)\|j_2\rangle_r = \delta^{\ \dot n_1}_{\dot n_3}\begin{bmatrix} \dot n_3 & \\ & n_4 \end{bmatrix}(j_1kj_2)^{n_4q_2n_2}{}_r \times\langle j_1n_1|T(kq_2)|j_2n_2\rangle \tag{3.5}$$

and

$$\langle j_1m_1|T(kq_1)|j_2m_2\rangle = \delta^{\dot m_3}_{\dot m_1}\begin{pmatrix} m_4 & \\ & \dot m_3 \end{pmatrix}(j_1kj_2)^r{}_{m_1q_1m_2} \times\langle j_1\|T(k)\|j_2\rangle_r. \tag{3.6}$$

This change has the effect of transferring a phase factor into the reduced matrix element which is, nevertheless, still real.

4. THE COUPLING OF BASIS VECTORS

In I and II attention was mainly concentrated on the ICR matrices, as this afforded the simplest means of access to the Racah algebra. However, in many cases it is not so much these matrices that are of interest but the basis vectors carrying the ICRs, and to these we now turn. It was shown in I and II that the reduction of the direct product of three ICRs to **0** led to the $3jm$ symbol with better properties than the coupling coefficient of the reduction of two ICRs to a third. In particular, the $3jm$ tensor satisfies a completeness property not shared by the coupling coefficient. It may be expected that this similarly holds for the coupling of basis vectors and our first step is to verify this completeness for the triple product.

It is trivial to verify that if

$$\{e^1{}_{(m)} : m = 1, \ldots [j_1]\}, \quad \{e^2{}_{(m)} : m = 1, \ldots, [j_2]\},$$

and

$$\{e^3{}_{(m)} : m = 1, \ldots [j_3]\}$$

form bases for j_1, j_2, and j_3, respectively, then

$$\{e^1{}_{(m_1)} e^2{}_{(m_2)} e^3{}_{(m_3)} : m_1, m_2, m_3 = 1,2, \cdots\}$$

forms a basis for $j_1 \otimes j_2 \otimes j_3$. Suppose there exists a linear combination of these which forms a basis for **0** so that

$$u(a^{n_1 n_2 n_3}{}_r e^1{}_{(n_1)} e^2{}_{(n_2)} e^3{}_{(n_3)}) = a^{m_1 m_2 m_3}{}_r e^1{}_{(m_1)} e^2{}_{(m_2)} e^3{}_{(m_3)} \tag{4.1}$$

and

$$a(a^{n_1 n_2 n_3}{}_r e^1{}_{(n_1)} e^2{}_{(n_2)} e^3{}_{(n_3)}) = \delta^r{}_{r'} a^{m_1 m_2 m_3}{}_{r'} e^1{}_{(m_1)} e^2{}_{(m_2)} e^3{}_{(m_3)}. \tag{4.2}$$

A multiplicity label r is included in this because there may be more than one such linear combination, and the tensor $\delta^r{}_{r'}$ is the tensor of Eq. (2.16).

By the linearity of u and Eq. (2.1), Eq. (4.1) may be rewritten

$$a^{m_1 m_2 m_3}{}_r e^1{}_{(m_1)} e^2{}_{(m_2)} e^3{}_{(m_3)} = a^{n_1 n_2 n_3}{}_r j_1(u)^{m_4}{}_{n_1} j_2(u)^{m_5}{}_{n_2} j_3(u)^{m_6}{}_{n_3} e^1{}_{(m_4)} e^2{}_{(m_5)} e^2{}_{(m_6)}. \tag{4.3}$$

As this is true for all u, it may be integrated over the linear subgroup H and divided by the volume of H to give, via Eq. (5.3) of I,

$$a^{m_1 m_2 m_3}{}_r e^1{}_{(m_1)} e^2{}_{(m_2)} e^3{}_{(m_3)} = a^{n_1 n_2 n_3}{}_r (j_1 j_2 j_3)^{m_4 m_5 m_6}{}_{r_1} (j_1 j_2 j_3)^{r_1}{}_{n_1 n_2 n_3} e^1{}_{(m_4)} e^2{}_{(m_5)} e^3{}_{(m_6)}. \tag{4.4}$$

From the orthogonality of the basis vectors it follows that

$$a^{m_1 m_2 m_3}{}_r = B^{r_1}{}_r (j_1 j_2 j_3)^{m_1 m_2 m_3}{}_{r_1}, \tag{4.5}$$

with

$$B^{r_1}{}_r = a^{n_1 n_2 n_3}{}_r (j_1 j_2 j_3)^{r_1}{}_{n_1 n_2 n_3}. \tag{4.6}$$

B is real from the antilinear equation. This is similar to the completeness condition for the 3 jm tensor as it follows that a complete set of orthonormal bases for **0** may be found by setting the tensor a equal to each column of the 3 jm tensor in turn.

Turning to the coupling of two basis vectors to a third, it is straightforward to show that with bases for j_1 and j_2 as above and

$$e^3{}_{r(m_3)} = \langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r m_3} e^1{}_{(m_1)} e^2{}_{(m_2)}, \tag{4.7}$$

the set $\{e^3{}_{r(m_3)} : m_3 = 1,2, \ldots\}$ forms a basis for j_3. Different bases exist for different values of r, and the vectors are all orthogonal

$$e^3{}_{r_1(m_3)} e^3_{r_2(m_4)} = \delta_{r_2 r_1} \delta_{m_4 m_3} \tag{4.8}$$

by unitarity of the coupling coefficient tensor. If a single coupling coefficient tensor is used, everything is fine. However, if j is of type (b) or (c) then there exist other coupling coefficient tensors that are *not* related by a transformation in the multiplicity label. These may be used to produce further sets of bases and by example we now show that these are not orthogonal to the first set.

Consider gray C_3^* as in Sec. 7 of I, with direct product $E' \otimes A_1 = E'$. Although the multiplicity of the coupling coefficient is only 1, there are four such tensors as E' is of type (b):

$$\langle E' A_1 | E' \rangle_{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \langle E' A_1 | E' \rangle_{(2)} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

$$\langle E' A_1 | E' \rangle_{(3)} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad \langle E' A_1 | E' \rangle_{(4)} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{4.9}$$

where a dummy label in brackets is used to distinguish the tensors. Using these tensors in turn in Eq. (4.7) gives the bases

$$\{e^1{}_{(1)} e^2{}_{(1)}, e^1{}_{(2)} e^2{}_{(1)}\}, \quad \{i e^1{}_{(1)} e^2{}_{(1)}, -i e^1{}_{(2)} e^2{}_{(1)}\},$$
$$\{i e^1{}_{(2)} e^2{}_{(1)}, i e^1{}_{(1)} e^2{}_{(1)}\}, \quad \text{and} \quad \{-e^1{}_{(2)} e^2{}_{(1)}, e^1{}_{(1)} e^2{}_{(1)}\}, \tag{4.10}$$

which are certainly not orthogonal. This is no cause for alarm, but only for care. It means that where a coupling coefficient tensor is required only one should be used and not some mixture of two or more.

5. TRANSFORMATION PROPERTIES

Transformation properties under a change of basis are always of importance, especially when symmetry adaptation to various subgroup chains is used. For the 3 jm tensor they are very easily derived. Let P_1, P_2, and P_3 be three unitary change of basis matrices for j_1, j_2, and j_3, respectively, which act according to Eqs. (2.5) and (2.7) of I. Substitution into Eq. (5.3) of I gives immediately

$$(j_1 j_2 j_3)^{m_1' m_2' m_3'}{}_{r'} (j_1 j_2 j_3)^{r'}{}_{n_1' n_2' n_3'} = P^{m_1'}{}_{1\ m_1} P^{m_2'}{}_{2\ m_2} P^{m_3'}{}_{3\ m_3} (j_1 j_2 j_3)^{m_1 m_2 m_3}{}_r \times (j_1 j_2 j_3)^r{}_{n_1 n_2 n_3} P^{n_1}{}_{1\ n_1'} P^{n_2}{}_{2\ n_2'} P^{n_3}{}_{3\ n_3'},$$

or

$$(j_1 j_2 j_3)^{m_1' m_2' m_3'}{}_{r'} = E^r{}_{r'} P^{m_1'}{}_{m_1} P^{m_2'}{}_{m_2} P^{m_3'}{}_{m_3} (j_1 j_2 j_3)^{m_1 m_2 m_3}{}_r, \tag{5.1}$$

with

$$E^r{}_{r'} = (j_1 j_2 j_3)^r{}_{n_1 n_2 n_3} P^{n_1}{}_{1\ n_1'} P^{n_2}{}_{2\ n_2'} P^{n_3}{}_{3\ n_3'} (j_1 j_2 j_3)^{n_1' n_2' n_3'}{}_{r'}. \tag{5.2}$$

The antilinear equation (5.4) of I shows E is real and hence orthogonal.

An interesting special case of Eq. (5.1) occurs when all P_i belong to the set of matrices commuting with j_i, for then

$$(j_1 j_2 j_3)^{m_1' m_2' m_3'}{}_{r'} = E^r{}_{r'} \langle j_1 0 | j_1 \rangle^{m_1' 0}{}_{m_1} \langle j_2 0 | j_2 \rangle^{m_2' 0}{}_{m_2} \times \langle j_3 0 | j_3 \rangle^{m_3' 0}{}_{m_3} (j_1 j_2 j_3)^{m_1 m_2 m_3}{}_r. \tag{5.3}$$

However, as far as the ICR matrices are concerned these commuting transformations have no effect, i.e., they do not alter the matrices at all. Thus the 3 jm tensor on the left reduces the same triple product as the 3 jm tensor on the right and hence by completeness the two are related by an orthogonal transformation in the multiplicity label:

$$(j_1 j_2 j_3)^{m_4 m_5 m_6}{}_{r_2} = F^{r_1}{}_{r_2} \langle j_1 0 | j_1 \rangle^{m_4 0}{}_{m_1} \langle j_2 0 | j_2 \rangle^{m_5 0}{}_{m_2} \times \langle j_3 0 | j_3 \rangle^{m_6 0}{}_{m_3} (j_1 j_2 j_3)^{m_1 m_2 m_3}{}_{r_1}. \tag{5.4}$$

In representation theory the equation corresponding to this is trivial, but here it is not because for ICR's of types (b) and (c) the coupling coefficient tensors are not necessarily constant. This allows the production of other columns of a $3jm$ tensor from a given one and thus acts as a general $3jm$ version of the paragraph following Eq. (7.10) of I. It also shows that the tensor F cannot always be taken as diagonally constant.

Turning now to the coupling coefficient a different approach needs to be adopted. This is characteristic of the methods applicable to gray groups, that in dealing with the $3jm$ tensors Eqs. (5.3) and (5.4) of I are used whereas for the coupling coefficient an intertwining or commutation relation with an ICR is required. In this case we consider the transformed form, and the transformations leading to it, of Eq. (7.1) of I:

$$
\begin{aligned}
&j_1(u)^{m_1'}{}_{n_1'}\, j_2(u)^{m_2'}{}_{n_2'} \\
&\quad = \sum_{j_3} \langle j_1 j_2 | j_3 \rangle^{m_1' m_2'}{}_{r' m_3'}\, j_3(u)^{m_3'}{}_{n_3'} \langle j_1 j_2 | j_3 \rangle^{r' n_3'}{}_{n_1' n_2'} \\
&\quad = \sum_{j_4} P_1^{m_1'}{}_{m_1} P_2^{m_2'}{}_{m_2} \langle j_1 j_2 | j_4 \rangle^{m_1 m_2}{}_{r m_4} P_4^{m_4}{}_{m_4'}\, j_4(u)^{m_4'}{}_{n_4'} \\
&\qquad \times P_4^{n_4'}{}_{n_4} \langle j_1 j_2 | j_4 \rangle^{r m_4}{}_{n_1 n_2} P_1^{n_1}{}_{n_1'} P_2^{n_2}{}_{n_2'}. \qquad (5.5)
\end{aligned}
$$

Simple manipulations using the orthogonality of the P_i and the coupling coefficients turn this into an intertwining relation of the form

$$G j_3(u)' = j_4(u)' G.$$

Doing the same with the antilinear equations,

$$G j_3(a)' = j_4(a)' G^*,$$

from which G is zero unless $j_3' = j_4'$, in which case it is one of the matrices commuting with j_3'. More explicitly,

$$
\begin{aligned}
&\langle j_1 j_2 | j_3 \rangle^{m_1' m_2'}{}_{r' m_3'} \\
&\quad = G^{r m_4'}{}_{r' m_3'}\, P_1^{m_1'}{}_{m_1} P_2^{m_2'}{}_{m_2} P_3^{m_3}{}_{m_4'} \langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r m_3}, \qquad (5.6)
\end{aligned}
$$

with

$$
\begin{aligned}
&G^{r m_4'}{}_{r' m_4'} \\
&\quad = \langle j_1 j_2 | j_3 \rangle^{r m_3}{}_{m_1 m_2} P_1^{m_1}{}_{m_1'} P_2^{m_2}{}_{m_2'} P_3^{m_4'}{}_{m_3} \langle j_1 j_2 | j_3 \rangle^{m_1' m_2'}{}_{r' m_3}, \qquad (5.7)
\end{aligned}
$$

where G is a tensor of the form $\langle j_3 0 | j_3 \rangle$ for each pair of multiplicity labels. It need not be diagonal in the m-values nor even be the same commuting matrix for different pairs of multiplicity labels. This introduces a high degree of freedom into the transformed coupling coefficient. Out of the possibilities two recommend themselves for further examination. The first is to multiply the transformed tensor by the inverse of $\langle j_3 0 | j_3 \rangle$. This diagonalizes G in the m-values. The second case is when all P_i are the identity transformation. This gives

$$\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r_2 m_4} = G^{r_1 m_3}{}_{r_2 m_4} \langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r_1 m_3}. \qquad (5.8)$$

It is not hard to show that this is the completeness property for coupling coefficient tensors although of a less manageable form than that for $3jm$ tensors. It should be used with caution because when G is nondiagonally constant in the m-values it forms a relation between different and nonorthogonal coupling coefficient tensors.

6. THE RECOUPLING COEFFICIENT

The first of the n-j or n-j related symbols we shall consider is the recoupling coefficient. This relates two different coupling schemes in reducing $j_1 \otimes j_2 \otimes j_3$ to j_4. Performing this through the intermediate schemes of $j_1 \otimes j_2$ to j_{12} and of $j_2 \otimes j_3$ to j_{23} by Eq. (7.1) of I gives

$$
\begin{aligned}
&j_1(u)^{m_1}{}_{n_1}\, j_2(u)^{m_2}{}_{n_2}\, j_3(u)^{m_3}{}_{n_3} \\
&\quad = \sum_{j_{12}, j_4} \langle j_1 j_2 | j_{12} \rangle^{m_1 m_2}{}_{r_1 m_{12}} \langle j_{12} j_3 | j_4 \rangle^{m_{12} m_3}{}_{r_2 m_4} \\
&\qquad \times j_4(u)^{m_4}{}_{n_4} \langle j_1 j_2 | j_{12} \rangle^{r_1 n_{12}}{}_{n_1 n_2} \langle j_{12} j_3 | j_4 \rangle^{r_2 n_4}{}_{n_{12} n_3} \\
&\quad = \sum_{j_{23}, j_5} \langle j_2 j_3 | j_{23} \rangle^{m_2 m_3}{}_{r_3 m_{23}} \langle j_1 j_{23} | j_5 \rangle^{r_4 m_1 m_{23}}{}_{r_4 m_5} \\
&\qquad \times j_5(u)^{m_5}{}_{n_5} \langle j_2 j_3 | j_{23} \rangle^{r_3 n_{23}}{}_{n_2 n_3} \langle j_1 j_{23} | j_5 \rangle^{r_4 n_5}{}_{n_1 n_{23}}. \qquad (6.1)
\end{aligned}
$$

By orthogonality of the coupling coefficient this may be rearranged to

$$
\begin{aligned}
&j_4(u)^{m_4}{}_{n_4} \langle j_1 j_2 | j_{12} \rangle^{r_1 n_{12}}{}_{n_1 n_2} \langle j_{12} j_3 | j_4 \rangle^{r_2 n_4}{}_{n_{12} n_3} \\
&\qquad \times \langle j_2 j_3 | j_{23} \rangle^{n_2 n_3}{}_{r_3 n_{23}} \langle j_1 j_{23} | j_5 \rangle^{n_1 n_{23}}{}_{r_4 n_5} \\
&\quad = \langle j_1 j_2 | j_{12} \rangle^{r_1 m_{12}}{}_{m_1 m_2} \langle j_{12} j_3 | j_4 \rangle^{r_2 m_4}{}_{m_{12} m_3} \\
&\qquad \times \langle j_2 j_3 | j_{23} \rangle^{m_2 m_3}{}_{r_3 m_{23}} \langle j_1 j_{23} | j_5 \rangle^{m_1 m_{23}}{}_{r_4 m_5}\, j_5(u)^{m_5}{}_{n_5}, \qquad (6.2)
\end{aligned}
$$

with a similar equation for the antilinear operators.

For each set of values of the multiplicity labels the two equations form an intertwining relation

$$j_4(u) R = R j_5(u) \quad \text{and} \quad j_4(u) R^* = R j_5(u), \qquad (6.3)$$

and hence R is zero unless $j_4 = j_5$, in which case it commutes with j_4. We therefore define the recoupling coefficient as

$$
\begin{aligned}
&\langle (j_1 j_2) j_{12} j_3, j_4 | j_1 (j_2 j_3) j_{23}, j_4 \rangle^{r_1 r_2 m_4}{}_{r_3 r_4 m_5} \\
&\quad = \langle j_1 j_2 | j_{12} \rangle^{r_1 m_{12}}{}_{m_1 m_2} \langle j_{12} j_3 | j_4 \rangle^{r_2 m_4}{}_{m_{12} m_3} \\
&\qquad \times \langle j_2 j_3 | j_{23} \rangle^{m_2 m_3}{}_{r_3 m_{23}} \langle j_1 j_{23} | j_4 \rangle^{m_1 m_{23}}{}_{r_4 m_5}. \qquad (6.4)
\end{aligned}
$$

An alternative form may be found by setting u equal to the identity of G in Eq. (6.1) and transferring just two coupling coefficients

$$
\begin{aligned}
&\langle j_1 j_2 | j_{12} \rangle^{r_1 n_{12}}{}_{n_1 n_2} \langle j_{12} j_3 | j_4 \rangle^{r_2 n_4}{}_{n_{12} n_3} \\
&\quad = \sum_{j_{23}} \langle (j_1 j_2) j_{12} j_3, j_4 | j_1 (j_2 j_3) j_{23}, j_4 \rangle^{r_1 r_2 n_4}{}_{r_3 r_4 n_5} \\
&\qquad \times \langle j_2 j_3 | j_{23} \rangle^{r_3 n_{23}}{}_{n_2 n_3} \langle j_1 j_{23} | j_4 \rangle^{r_4 n_5}{}_{n_1 n_{23}}. \qquad (6.5)
\end{aligned}
$$

From this immediately follows the orthogonality property

$$
\begin{aligned}
&\sum_{j_{23}} \langle (j_1 j_2) j_{12} j_3, j_4 | j_1 (j_2 j_3) j_{23}, j_4 \rangle^{r_1 r_2 n_4}{}_{r_3 r_4 n_5} \\
&\qquad \times \delta^{r_3 r_3'} \delta^{r_4 r_4'} \delta^{n_5 n_5'} \langle (j_1 j_2) j_{12}' j_3, j_4 | j_1 (j_2 j_3) j_{23} j_4 \rangle_{r_3' r_4' n_5'}{}^{r_1' r_2' n_4'} \\
&\quad = \delta^{r_1 r_1'} \delta^{r_2 r_2'} \delta^{n_4 n_4'} \delta(j_{12}, j_{12}'). \qquad (6.6)
\end{aligned}
$$

If we were using representation theory, Schur's lemma on Eq. (6.3) would give directly that the recoupling coefficient is diagonally constant in the m-values. However if j_4 is on ICR of type (b) or (c) it need not be, and it is only necessary to give one suitable example to show that in general it is not constant. Such an example is readily found in gray C_4 with $j_1 = A_1, j_2 = j_{12} = j_{23} = j_4 = E$, and $j_3 = A_2$. The multiplicity

of each coupling coefficient tensor is only 1, whereas the multiplicity of A_1 in each triple product is 2. Thus for each triple there are two coupling coefficient tensors related by a commuting matrix. Distinguishing between these by a *dummy* multiplicity label in parentheses (r) we have

$$\langle A_1 E|E\rangle^{a_1 x}{}_{(1)x} = \langle AE|E\rangle^{a_1 y}{}_{(1)y} = 1,$$
$$\langle A_1 E|E\rangle^{a_1 x}{}_{(2)x} = -\langle A_1 E|E\rangle^{a_1 y}{}_{(2)y} = i,$$

and

$$\langle EA_2|E\rangle^{x a_2}{}_{(1)y} = \langle EA_2|E\rangle^{y a_2}{}_{(1)x} = 1,$$
$$\langle EA_2|E\rangle^{x a_2}{}_{(2)y} = -\langle EA_2|E\rangle^{y a_2}{}_{(2)y} = i,$$

with all other elements zero. Then

$$\langle (A_1 E)EA_2, E|A_1(EA_2)E, E\rangle^{(1)(1)}{}_{(1)(2)} = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix},$$

as required.

It should be clear that in this example we have done a "silly" thing. The more sensible choice would have been to use the coupling coefficient with dummy label (1) instead of (2). This would have given a diagonal recoupling coefficient. In any particular recoupling coefficient such an adjustment can always be made *post hoc* to one or more of the coupling coefficients, but it is not clear whether a set of coupling coefficient tensors, one for each triad j_1, j_2, j_3 of G, can be found which give all recoupling coefficient tensors of G diagonal. The importance of being able to find such a set lies in the breakdown of such results as this: if the recoupling coefficient here is not diagonal in the m-values, then the recoupling coefficient of four ICRs to a fifth cannot be expressed as a product of three of these recoupling coefficients.

7. REDUCTION OF THE QUARTIC PRODUCT

In the next approach to the recoupling problem, we consider the reduction of the quartic product $j_1 \otimes j_2 \otimes j_3 \otimes j_4$ to the identity **0**. A variety of intermediate coupling schemes exist of which we choose $j_1 \otimes j_2$ to j_{12}, $j_{12} \otimes j_3 \otimes j_4$ to **0** and $j_2 \otimes j_3$ to j_{23}, $j_1 \otimes j_{23} \otimes j_4$ to **0**. Although it may be hoped that this leads to a satisfactory 6*j* with properties completely analogous to the 6*j* symbol in representation theory, this turns out not to be the case. For example, if other intermediate schemes are used, the various symbols found are not essentially equivalent to each other.

To proceed, the coupling of two to a third may as before be effected by Eq. (7.1) of I, and of three to **0** by Eq. (5.1) of I. This gives

$$\begin{aligned} &j_1(u)^{m_1}{}_{n_1} j_2(u)^{m_2}{}_{n_2} j_3(u)^{m_3}{}_{n_3} j_4(u)^{m_4}{}_{n_4} \\ &= \sum_{j_{12}} \langle j_1 j_2|j_{12}\rangle^{m_1 m_2}{}_{r_1 m_{12}} (j_{12} j_3 j_4)^{m_{12} m_3 m_4}{}_{r_2} \\ &\quad \times (j_{12} j_3 j_4)^{r_2}{}_{n_{12} n_3 n_4} \langle j_1 j_2|j_{12}\rangle^{r_1 n_{12}}{}_{n_1 n_2} \oplus \text{other ICRs} \\ &= \sum_{j_{23}} \langle j_2 j_3|j_{23}\rangle^{m_2 m_3}{}_{r_3 m_{23}} (j_1 j_{23} j_4)^{m_1 m_{23} m_4}{}_{r_4} \\ &\quad \times (j_1 j_{23} j_4)^{r_4}{}_{n_1 n_{23} n_4} \langle j_2 j_3|j_{23}\rangle^{r_3 n_{23}}{}_{n_2 n_3} \oplus \text{other ICRs}. \end{aligned} \tag{7.1}$$

Upon integration over H all these other ICRs not equal to **0** drop out [I, Eqs. (3.1c) and (3.6)] and the two summations given may be equated. Orthogonality of the coupling coefficient tensors gives

$$\begin{aligned} &\langle j_1 j_2|j_{12}\rangle^{r_1 n_{12}}{}_{n_1 n_2} (j_{12} j_3 j_4)^{r_2}{}_{n_{12} n_3 n_4} \\ &= \sum_{j_{23}} \phi_{j_3} [j_{12}]^{1/2} [j_{23}]^{1/2} \begin{Bmatrix} j_1 & j_2 & \cdot & j_{12} \\ & & \cdot & \\ j_3 & j_4 & \cdot & j_{23} \end{Bmatrix}^{r_1 r_2}{}_{r_3 r_4} \\ &\quad \times \langle j_2 j_3|j_{23}\rangle^{r_3 n_{23}}{}_{n_2 n_3} (j_1 j_{23} j_4)^{r_4}{}_{n_1 n_{23} n_4}, \end{aligned} \tag{7.2}$$

with

$$\begin{aligned} &\begin{Bmatrix} j_1 & j_2 & \cdot & j_{12} \\ & & \cdot & \\ j_3 & j_4 & \cdot & j_{23} \end{Bmatrix}^{r_1 r_2}{}_{r_3 r_4} \\ &= \phi_{j_3} [j_{12}]^{-1/2} [j_{23}]^{-1/2} \langle j_1 j_2|j_{12}\rangle^{r_1 m_{12}}{}_{m_1 m_2} (j_{12} j_3 j_4)^{r_2}{}_{m_{12} m_3 m_4} \\ &\quad \times \langle j_2 j_3|j_{23}\rangle^{m_2 m_3}{}_{r_3 m_{23}} (j_1 j_{23} j_4)^{m_1 m_{23} m_4}{}_{r_4}. \end{aligned} \tag{7.3}$$

The antilinear equations give this "quasi-6*j*" as a real tensor. The 1-*j* phase $\phi_{j_3} = j_3(\theta^2)$ is inserted for later convenience. Two immediate points distinguish the quasi-6*j* from the recoupling coefficient: first, it is always independent of the m-values and, second, the range of the multiplicity labels r_2 and r_4 is larger if j_4 is of type (b) or (c). This tensor satisfies an orthogonality relation that is also different in the multiplicity range

$$\begin{aligned} &\sum_{j_{23}} [j_{12}]\,[j_{23}] \begin{Bmatrix} j_1 & j_2 & \cdot & j_{12} \\ & & \cdot & \\ j_3 & j_4 & \cdot & j_{23} \end{Bmatrix}^{r_1 r_2}{}_{r_3 r_4} \\ &\quad \times \delta^{r_3 \dot{r}_3} \delta^{r_4 \dot{r}_4} \begin{Bmatrix} j_1 & j_2 & \cdot & j'_{12} \\ & & \cdot & \\ j_3 & j_4 & \cdot & j_{23} \end{Bmatrix}^{\dot{r}_1 \dot{r}_2}{}_{\dot{r}_3 \dot{r}_4} \\ &= \delta^{r_1 \dot{r}_1} \delta^{r_2 \dot{r}_2} \delta(j_{12}, j'_{12}). \end{aligned}$$

Here is an appropriate point to begin discussion of two criteria for the 3*jm* tensor which have not yet been satisfactorily resolved. The Racah backcoupling rule *almost* holds for the quasi-6*j* symbol. If the definition were changed by transposing j_2 and j_3 in the coupling coefficient $\langle j_2 j_3|j_{23}\rangle$, then j_2 would occupy a sufficiently symmetric position that no permutations would be needed to give the rule. However, the standard order has become so firmly embedded in the literature that we have to consider an alternative question: can the order be permuted in a coupling coefficient? Permutations involve the 3*jm* tensor and so this question resolves into one concerning the relation between the coupling coefficient and the 3*jm* tensor. To be "sensible" (that is, to avoid unnecessary work), the relation between the coupling coefficient tensor should be of the form of Eq. (7.9) of I. Due to the large variability of the 3*jm* tensor under multiplicity label transforms this requirement must act as a restrictive condition, as we can see from the tighter orthogonality condition for the coupling coefficient. On the other hand, the permutation properties of the 3*jm* symbol have their own simple form which in turn acts as a restrictive condition. In every numerical case so far considered, 3*jm* tensors can be found that satisfy both of these simplicity criteria but neither a demonstration that it can *always* be done nor a counterexample have yet been found. Caution urges that we refrain from permuting the coupling coefficient until the situation becomes clearer.

8. THE INVARIANT 6j TENSOR

We define the $6j$ tensor by

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}_{r_1 r_2 r_3 r_4}$$
$$= \delta_{m_1\dot{m}_1}\delta_{m_2\dot{m}_2}\delta_{m_3\dot{m}_3}\delta_{m_4\dot{m}_4}\delta_{m_5\dot{m}_5}\delta_{m_6\dot{m}_6}$$
$$\times \begin{bmatrix} \dot{m}_1 & \\ & m_7 \end{bmatrix}\begin{bmatrix} \dot{m}_2 & \\ & m_8 \end{bmatrix}\begin{bmatrix} \dot{m}_3 & \\ & m_9 \end{bmatrix}\begin{bmatrix} \dot{m}_4 & \\ & m_{10} \end{bmatrix}$$
$$\times \begin{bmatrix} \dot{m}_5 & \\ & m_{11} \end{bmatrix}\begin{bmatrix} \dot{m}_6 & \\ & m_{12} \end{bmatrix}$$
$$\times (j_1 j_2 j_3)^{m_7 m_8 m_9}{}_{r_1}(j_3 j_4 j_5)^{m_9 m_{10} m_{11}}{}_{r_2}$$
$$\times (j_2 j_4 j_6)^{m_2 m_4 m_{12}}{}_{r_3}(j_1 j_6 j_5)^{m_1 m_6 m_5}{}_{r_4}. \tag{8.1}$$

This differs in the ordering of j values to the definition given by Butler[9] for linear groups and was chosen so as to be simply related to the quasi-$6j$. Under a change of basis it transforms according to

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}_{r_1' r_2' r_3' r_4'}$$
$$= E_{123}{}^{r_1}{}_{r_1'} E_{345}{}^{r_2}{}_{r_2'} E_{246}{}^{r_3}{}_{r_3'} E_{165}{}^{r_4}{}_{r_4'} \begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}_{r_1 r_2 r_3 r_4}, \tag{8.2}$$

where the suffixes label the j-values, and hence

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}^{r_1 r_2 r_3 r_4} \begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}_{r_1 r_2 r_3 r_4} \tag{8.3}$$

is invariant under such transformations. We note that an alternative definition similar to Eq. (7.2) does not exist due to orthogonality breakdown in a sum over two m-values.

The two other major properties of this tensor are its reality and its behavior under permutations. To show it is real we need some preliminary results. First, from Eq. (2.14),

$$\begin{bmatrix} \dot{m}_1 & \\ & m_2 \end{bmatrix}\delta_{\dot{m}_1 m_1} = \begin{pmatrix} \dot{m}_2 & \\ & m_1 \end{pmatrix}\delta_{m_2\dot{m}_2}. \tag{8.4}$$

Second, a very strong relation exists between the fully conjugated and unconjugated $3jm$ tensors. From Eq. (5.2) of I,

$$(j_1 j_2 j_3)^{m_1 m_2 m_3}{}_{r_1}\delta^{r_1}{}_{\dot{r}_2} = j_1(a)^{m_1}{}_{\dot{n}_1} j_2(a)^{m_2}{}_{\dot{n}_2} j_3(a)^{m}{}_{\dot{n}_3}$$
$$\times (j_1 j_2 j_3)^{\dot{n}_1\dot{n}_2\dot{n}_3}{}_{\dot{r}_2}. \tag{8.5}$$

Specializing this by setting a equal to θ,

$$(j_1 j_2 j_3)^{m_1 m_2 m_3}{}_{r_1} = \delta^{r_2}{}_{r_1}\begin{pmatrix} m_1 & \\ & \dot{n}_1 \end{pmatrix}\begin{pmatrix} m_2 & \\ & \dot{n}_2 \end{pmatrix}$$
$$\times \begin{pmatrix} m_3 & \\ & \dot{n}_3 \end{pmatrix}(j_1 j_2 j_3)^{\dot{n}_1\dot{n}_2\dot{n}_3}{}_{\dot{r}_2}. \tag{8.6}$$

This is an improvement on Eq. (8.3) of I. Substitution of this into Eq. (8.1) and some manipulations give the $6j$ real.

The symmetry properties of the $6j$ tensor under permutations follow in the same manner as for linear groups[9] with the simplication that the gray groups are ambivalent. The results are collected in Table I where the M_i are the permutation matrices of the $3jm$ tensors, the π_i are the cycle structures of the permutations of S_3 and ϕ is a product of 1-j phases. Thus for example the first entry in this table gives

$$\begin{Bmatrix} j_1 & j_3 & j_2 \\ j_4 & j_6 & j_5 \end{Bmatrix}_{r_1' r_3' r_2' r_4'}$$
$$= \phi_{j_2}\phi_{j_3}\phi_{j_4} M_1(23)^{r_1}{}_{r_1'} M_2(1)^{r_2}{}_{r_2'} M_3(1)^{r_3}{}_{r_3'} M_4(23)^{r_4}{}_{r_4'}$$
$$\times \begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}_{r_1 r_2 r_3 r_4}. \tag{8.7}$$

TABLE I. Permutations of the $6j$ symbols.[a]

$j_1'\ j_2'\ j_3'\ j_4'\ j_5'\ j_6'$	$a\ b\ c\ d$	π_1	π_2	π_3	π_4	ϕ
$j_1\ j_3\ j_2\ j_4\ j_6\ j_5$	1 3 2 4	(23)	(1)	(1)	(23)	$\phi_2\phi_3\phi_4$
$j_1\ j_5\ j_6\ j_4\ j_2\ j_3$	4 3 2 1	(23)	(13)	(13)	(23)	$\phi_1\phi_2\phi_4\phi_5$
$j_3\ j_1\ j_2\ j_6\ j_4\ j_5$	1 3 4 2	(132)	(23)	(23)	(I)	$\phi_2\phi_3\phi_4\phi_5$
$j_2\ j_1\ j_3\ j_5\ j_4\ j_6$	1 2 4 3	(12)	(23)	(23)	(23)	ϕ_6
$j_4\ j_3\ j_5\ j_1\ j_6\ j_2$	2 4 1 3	(132)	(12)	(12)	(132)	$\phi_1\phi_5\phi_6$
$j_4\ j_5\ j_3\ j_1\ j_2\ j_6$	2 1 4 3	(132)	(123)	(123)	(132)	$\phi_3\phi_6$
$j_2\ j_3\ j_1\ j_5\ j_6\ j_4$	1 4 2 3	(123)	(23)	(I)	(23)	$\phi_1\phi_3\phi_5\phi_6$
$j_5\ j_6\ j_1\ j_2\ j_3\ j_4$	4 1 3 2	(I)	(13)	(132)	(13)	$\phi_3\phi_4\phi_5\phi_6$
$j_4\ j_6\ j_2\ j_1\ j_3\ j_5$	3 1 4 2	(12)	(123)	(123)	(12)	$\phi_3\phi_4\phi_5$
$j_6\ j_4\ j_2\ j_3\ j_1\ j_5$	3 1 2 4	(123)	(12)	(13)	(123)	$\phi_3\phi_4$
$j_6\ j_2\ j_4\ j_3\ j_5\ j_1$	3 2 1 4	(123)	(12)	(132)	(12)	ϕ_2
$j_3\ j_5\ j_4\ j_6\ j_2\ j_1$	2 3 4 1	(132)	(23)	(123)	(13)	$\phi_1\phi_2\phi_4$
$j_1\ j_6\ j_5\ j_4\ j_3\ j_2$	4 2 3 1	(I)	(13)	(13)	(I)	$\phi_1\phi_2\phi_6$
$j_6\ j_1\ j_5\ j_3\ j_4\ j_2$	4 2 1 3	(23)	(132)	(132)	(12)	$\phi_1\phi_6$
$j_5\ j_1\ j_6\ j_2\ j_4\ j_3$	4 3 1 2	(I)	(132)	(132)	(132)	$\phi_1\phi_2\phi_3\phi_4\phi_5$
$j_3\ j_2\ j_1\ j_6\ j_5\ j_4$	1 4 3 2	(13)	(I)	(23)	(I)	$\phi_1\phi_3\phi_4\phi_5\phi_6$
$j_6\ j_5\ j_1\ j_3\ j_2\ j_4$	4 1 2 3	(23)	(132)	(13)	(123)	$\phi_3\phi_5\phi_6$
$j_4\ j_2\ j_6\ j_1\ j_5\ j_3$	3 4 1 2	(12)	(12)	(12)	(12)	$\phi_1\phi_2\phi_3\phi_4\phi_5\phi_6$
$j_3\ j_4\ j_5\ j_6\ j_1\ j_2$	2 4 3 1	(13)	(I)	(123)	(13)	$\phi_1\phi_2\phi_5\phi_6$
$j_2\ j_4\ j_6\ j_5\ j_1\ j_3$	3 4 2 1	(123)	(123)	(I)	(123)	$\phi_1\phi_2\phi_4\phi_5\phi_6$
$j_5\ j_4\ j_3\ j_2\ j_1\ j_6$	2 1 3 4	(13)	(13)	(12)	(13)	ϕ_3
$j_2\ j_6\ j_4\ j_5\ j_3\ j_1$	3 2 4 1	(12)	(123)	(23)	(123)	$\phi_1\phi_2$
$j_5\ j_3\ j_4\ j_2\ j_6\ j_1$	2 3 1 4	(13)	(132)	(12)	(132)	$\phi_2\phi_4$

[a]The values of $j_1', j_2', j_3', j_4', j_5', j_6'$; a, b, c, d; $\pi_1, \pi_2, \pi_3, \pi_4$ and ϕ which is a product of $1-j$ phases are given, where

$$\begin{Bmatrix} j_1' & j_2' & j_3' \\ j_4' & j_5' & j_6' \end{Bmatrix}_{r_1' r_2' r_3' r_4'} = M_1(\pi_1)^{r_1}{}_{r_1'} M_2(\pi_2)^{r_2}{}_{r_2'} M_3((\pi_3)^{r_3}{}_{r_3'} M_4(\pi_4)^{r_4}{}_{r_4'}\ \phi \begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}_{r_1 r_2 r_3 r_4}.$$

In Sec. 10, where a group is considered with a 3jm tensor with $\{21\}$ symmetry, it is found that although these symmetries are complete, they are not always convenient. Due to the number of probable cases and the special nature of the gray groups we shall not consider this further here.

Despite these permutational properties, the lack of an orthogonality property in a sum over only two m-values blocks the various 6j identities.

In Secs. 6–8 we have presented the most pessimistic aspects of the various symbols. However, as we have not yet found examples necessarily possessing such bad behavior, we consider in brief the "ideal" situation. A number of criteria need to be satisfied. First, the symmetrized 3jm tensor must fall into blocks, with each block simply giving a coupling coefficient tensor. Further, under permutations this block structure must be preserved to the extent that a block in one order yields one, two or four blocks in another or vice versa. To be more explicit, in one order we may have ICRs of type (a),(b),(c). As the third ICR is of type (c) the 3jm tensor should split into two coupling coefficient tensor blocks. In the order (a),(c),(b) it should split into four blocks. What we require is that when one of the blocks in the first order is permuted it should break into exactly two of the blocks of the second order without any "mixing up" of the blocks in the second order. Secondly, with the coupling coefficient tensors effectively symmetrized by this, that one may be taken as a "basic" coupling coefficient, with all others derivable from it by commuting tensors with trace zero. By computation these two conditions may be satisfied by the ICRs of the gray double point groups. The third condition is that the recoupling coefficient derived from these basic coupling coefficient tensors be diagonally constant in the m-values.

If we are granted all these conditions, then as far as recoupling problems are concerned we need only consider the recoupling coefficient derived from the basic coupling coefficients. For the quasi-6j and 6j symbols many of the extra values caused by increasing the ranges of the multiplicity variables may be found by multiplying a pair of commuting matrices and taking their trace. Many will be zero. The permutations of the coupling coefficient will allow the 6j identities to hold.

Finally in this section, indices may be raised on the 6j tensor by

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}^{r_5}_{r_2 r_3 r_4} = \delta^{r_5}{}_{r_1} \delta^{r_1 \bar{r}_1} \begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}_{r_1 r_2 r_3 r_4}, \quad (8.8)$$

and so on. If the special choice of 3jm symbols of Eq. (7.9) in I is made (with the positive sign chosen for convenience) some messy but straightforward manipulations give the quasi-6j as part of a 6j with two indices raised. This provides the relation required between the two tensors if we have the "ideal" situation above.

9. COUPLED TENSOR OPERATORS

We now turn to the formulas of Racah[9] involving products of tensor operators acting on coupled schemes. In what follows we shall assume that a final coupled state is a "good" state, and consider the case of the matrix elements of

$$\langle (j_3 j_4) j_1 m_1 || (TU) X(kq) || (j_5 j_6) j_2 m_2 \rangle,$$

where T acts only between j_3 and j_5, and U between j_4 and j_6. In accord with Eq. (4.7) multiplicity labels must be inserted into this matrix element to label the couplings and we have

$$\begin{aligned} &\langle (j_3 j_4) r_1 j_1 m_1 | (TU) r_3 X(kq) | | (j_5 j_6) r_2 j_2 m_2 \rangle \\ &\quad = \langle j_3 j_4 | j_1 \rangle_{\bar{r}_1 \bar{m}_1}{}^{\bar{m}_3 \bar{m}_4} \langle k_1 k_2 | k \rangle^{q_1 q_2}{}_{r_3 q} \langle j_5 j_6 | j_2 \rangle^{m_5 m_6}{}_{r_2 m_2} \\ &\qquad \times \langle j_3 m_3 | T(k_1 q_1) | j_5 m_5 \rangle \langle j_4 m_4 | U(k_2 q_2) | j_6 m_6 \rangle. \quad (9.1) \end{aligned}$$

Both sides of this expression may be expanded by the Wigner–Eckart theorem of Sec. 3. The resultant expression is rather messy but manipulations of the type already dealt with give a quasi-9j relation between the reduced matrix elements

$$\begin{aligned} &\langle (j_3 j_4) r_1 j_1 || (TU) r_3 X(k) || (j_5 j_6) r_2 j_2 \rangle_{r_4} \\ &\quad = [j_1]^{1/2} [j_2]^{1/2} [k]^{1/2} \begin{pmatrix} j_3 & j_4 & \cdot & j_1 \\ k_1 & k_2 & \cdot & k \\ j_5 & j_6 & \cdot & j_2 \end{pmatrix}^{r_5 r_6}_{r_8 r_2 r_3 r_4} \\ &\qquad \times \delta^{r_7}_{r_1} \delta^{r_8}{}_{\bar{r}_7} \langle j_3 || T(k_1) || j_5 \rangle_{r_5} \langle j_4 || U(k) || j_6 \rangle_{r_6}, \quad (9.2) \end{aligned}$$

with

$$\begin{aligned} &\begin{Bmatrix} j_3 & j_4 & \cdot & j_1 \\ k_1 & k_2 & \cdot & k \\ j_5 & j_6 & \cdot & j_2 \end{Bmatrix}^{r_5 r_6}_{r_8 r_2 r_3 r_4} \\ &\quad = [j_1]^{-1/2} [j_2]^{-1/2} [k]^{-1/2} \langle j_3 j_4 | j_1 \rangle^{m_3 m_4}{}_{r_8 m_1} \\ &\qquad \times \langle k_1 k_2 | k \rangle^{q_1 q_2}{}_{r_3 q} \langle j_5 j_6 | j_2 \rangle^{m_5 m_6}{}_{r_2 m_2} \\ &\qquad \times (j_3 k_1 j_5)^{r_5}{}_{m_3 q_1 m_5} (j_4 k_2 j_6)^{r_6}{}_{m_4 q_2 m_6} (j_1 k j_2)^{m_1 q m_2}{}_{r_4}. \quad (9.3) \end{aligned}$$

Two special cases exist for this coupling scheme. First, let U be the identity operator. Although the tensor $(j_4 k_2 j_6)$ reduces to $(j_4 0 j_4)$ it may have a multiplicity greater than 1, and as the summation in Eq. (9.2) is over this multiplicity range the quasi-9j does not reduce to a quasi-6j. Second, if X is the identity operator we have the same situation in the sum over r_4 in equation (3.6) and the quasi-9j again will not reduce to a quasi-6j.

10. AN EXAMPLE

As an illustration to the theory developed here and in I and II, we consider the case where G is the direct product $H \times C_2$ and the linear subgroup H is the metacyclic group of order 21 with presentation

$$H = \langle x, y | x^7 = y^3 = e, xy = yx^2 \rangle.$$

Specializing from the general case for metacyclic groups given by Basmaji,[10,11] we find that H has five IRs—$\mathbf{0}, k_1, k_2, k_3, k_4$, with $k_1^* = k_2$ and $k_3^* = k_4$. The character table is given in Table II and the generators for k_3 are

$$k_3(x) = \begin{pmatrix} \zeta & 0 & 0 \\ 0 & \zeta^2 & 0 \\ 0 & 0 & \zeta^4 \end{pmatrix} \quad \text{and} \quad k_3(y) = \begin{pmatrix} 0 & 0 & I \\ I & 0 & 0 \\ 0 & I & 0 \end{pmatrix},$$

where ζ is a primitive seventh root of 1. As θ^2 here equals the identity there are three ICRs—$0, j_1$ and j_2, with

$$\mathbf{0}(u) = \mathbf{0}(a) = 1,$$

$$j_1(u) = \begin{pmatrix} k_1(u) & 0 \\ 0 & k_1(u)^* \end{pmatrix}, \quad j_1(\theta) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

and

TABLE II. The character table for $H = \langle x,y|x^7 = y^3 = e,\ xy = yx^3\rangle$.[a]

Class / IR	1 $\{e\}$	3 $\{x\}$	3 $\{x^3\}$	7 $\{y\}$	7 $\{y^2\}$
0	1	1	1	1	1
k_1	1	1	1	ω	ω^2
k_2	1	1	1	ω^2	ω
k_3	3	$\xi+\xi^2+\xi^4$	$\xi^3+\xi^5+\xi^6$	0	0
k_4	3	$\xi^3+\xi^5+\xi^6$	$\xi+\xi^2+\xi^4$	0	0

[a] The classes are $\{e\}$, $\{x, x^2x^4\}$, $\{x^6,x^5,x^3\}$, $\{yx^n : n = 0,...,6\}$ and $\{y^2x^n : n = 0,...,6\}$. ξ is a primitive seventh root of one, and ω a primitive cube root of one.

TABLE III. The magnetic character Table for $G = \langle x,y|x^7 = y^3 = e,\ xy = yx^3\rangle$.[a]

C-class / ICR	1 $\{e\}$	6 $\{x\}$	14 $\{y\}$	Intertwining number
0	1	1	1	1
j_1	2	2	-1	2
j_2	6	-1	0	2

[a] The C-classes are $\{e\}$, $\{x^n : n = 1,...,6\}$, $\{y^mx^n : m = 1,2;\ n = 0,...,6\}$.

$$j_2(u) = \begin{pmatrix} k_3(u) & 0 \\ 0 & k_3(u)^* \end{pmatrix}, \quad j_2(\theta) = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}.$$

The magnetic character table[4] for G is given in Table III.

It is easy to verify that $k_3 \otimes k_3 \otimes k_3 = 20 \oplus \ldots$ and $k_3 \otimes k_3 \otimes k_3^* = 0 \oplus \ldots$ and that in the notation of II

$$m^{k_3}{}_{\{1^3\}} = m^{k_3}{}_{\{3\}} = 1, \quad m^{k_3}{}_{\{21\}} = 0$$

and

$$m^{k_3}{}_{\{1^2\}} = 0, \quad m^{k_3}{}_{\{2\}} = 1.$$

Hence the linear subgroup H is simple phase. Inducing upwards to G by Sec. 2 of II,

$$m^{j_2}{}_{\{1^3\}} = 2, \quad m^{j_2}{}_{\{3\}} = 4, \quad \text{and} \quad m^{j_2}{}_{\{21\}} = 2.$$

This group is not simple phase—the symmetrized 3 jm tensor for $(j_2j_2j_2)$ will have two antisymmetric rows, four symmetric ones, and two pairs of rows transforming as the IR$\{21\}$ of S_3. We remark that these results may be obtained more directly from the magnetic character table[4] but as illustration of the methods of II the ascent from H to G will be retained in this example.

To calculate the 3 jm tensor $(j_2j_2j_2)$, ascent requires that we first construct the tensors $(k_3\ k_3\ k_3)$ and $(k_3\ k_3\ k_3^*)$. This is easily done and the nonzero entries are given in Tables IV and V. From these the methods of Sec. 5 in II may be used to construct the 3 jm tensor P as

$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} k_3\,k_3\,k_3 & k_3\,k_3\,k_3^* & k_3\,k_3^*\,k_3 & k_3\,k_3^*\,k_3^* & k_3^*\,k_3\,k_3 & k_3^*\,k_3\,k_3^* & k_3^*\,k_3^*\,k_3 & k_3^*k_3^*k_3^* \\ Q & 0 & 0 & 0 & 0 & 0 & 0 & Q^* \\ iQ & 0 & 0 & 0 & 0 & 0 & 0 & -iQ^* \\ 0 & Q' & 0 & 0 & 0 & 0 & Q'^* & 0 \\ 0 & iQ' & 0 & 0 & 0 & 0 & -iQ'^* & 0 \\ 0 & 0 & Q'' & 0 & 0 & Q''^* & 0 & 0 \\ 0 & 0 & iQ'' & 0 & 0 & -iQ''^* & 0 & 0 \\ 0 & 0 & 0 & Q''' & Q'''^* & 0 & 0 & 0 \\ 0 & 0 & 0 & iQ''' & -iQ'''^* & 0 & 0 & 0 \end{pmatrix},$$

where Q reduces $(k_3\ k_3\ k_3)$, Q' reduces $(k_3\ k_3\ k_3^*)$, Q'' reduces $(k_3\ k_3^*\ k_3)$, and Q''' reduces $(k_3\ k_3^*k_3^*)$. The block structure is shown along the top. (A complete list of such constructions is given in Ref. 6.) This construction does not give a symmetrized 3 jm tensor but this is readily corrected for this simple case and the nonzero results are given in Table VI. In Tables III–VI the convention is that the m-values of k_3 are labeled 1, 2, and 3, the m-values of k_3^*, 4, 5, and 6, and by adjunction the m-values of j_2 by 1 through 6.

This tensor possesses many properties that are peculiar to corepresentation theory. For example, as an illustration of Eq. (5.4), the multiplicity row six may be obtained from the first row by setting the first two coupling coefficient tensors equal to the identity and the third to

TABLE IV. The 3km tensor $(k_3k_3k_3)^r{}_{m_1m_2m_3}$.[a]

Multiplicity label	Symmetry	$m_1m_2m_3$ 123	132	213	231	312	321
1	$\{2\}$	$1/\sqrt{6}$	$1/\sqrt{6}$	$1/\sqrt{6}$	$1/\sqrt{6}$	$1/\sqrt{6}$	$1/\sqrt{6}$
2	$\{1^2\}$	$1/\sqrt{6}$	$-1/\sqrt{6}$	$-1/\sqrt{6}$	$1/\sqrt{6}$	$1/\sqrt{6}$	$-1/\sqrt{6}$

[a] The m values for k_3 are labeled one, two, and three. All triples $(m_1m_2m_3)$ not given here have value zero.

TABLE V. The $3km$ tensor $(k_3k_3k\overset{*}{_3})^r{}_{m_1m_2m_3}$.[a]

Multiplicity label	Symmetry	$m_1m_2m_3$ 115	226	334
1	{2}	$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{3}$

[a]The m values of k_3 are labeled one, two, three, and of $k\overset{*}{_3}$ four, five and six. All triples $(m_1m_2m_3)$ not given here have value zero.

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

Similarly row eight follows from row three (to a common factor of -1) by setting all three equal to this commuting tensor.

Since the $3jm$ tensor has multiplicity of ten and j_3 is of type (c), the multiplicity of the coupling coefficient is five. In construction the $3jm$ naturally fell into two blocks of five rows each, the first five being real and the second five pure imaginary. On transforming each of these blocks by Eq. (7.9) of I with

$$j(\theta) = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}.$$

It was found that each block gave a coupling coefficient tensor. That these two are not orthogonal may be established directly from the $3jm$ tensor

$$(j_2j_2j_2)^3{}_{m_1m_25}(j_2j_2j_2)^{m_1m_25}{}_8 \neq 0.$$

On permutations of the $3jm$ tensor the first five rows map into each other and the second five also map into each other. Thus there is no "crossing-over" between these blocks and the two coupling coefficient tensors are symmetrized:

TABLE VI. The $3jm$ tensor $(j_2j_2j_2)_{m_1m_2m_3}$.[a]

Multiplicity label	Symmetry	$m_1m_2m_3$ 123	231	312	213	132	321	456	564	645	465	546	654
1	{3}	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$
6	{3}	$i/2\sqrt{3}$	$i/2\sqrt{3}$	$i/2\sqrt{3}$	$i/2\sqrt{3}$	$i/2\sqrt{3}$	$i/2\sqrt{3}$	$-i/2\sqrt{3}$	$-i/2\sqrt{3}$	$-i/2\sqrt{3}$	$-i/2\sqrt{3}$	$-i/2\sqrt{3}$	$-i/2\sqrt{3}$
2	$\{1^3\}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$-1/2\sqrt{3}$	$-1/2\sqrt{3}$	$-1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$1/2\sqrt{3}$	$-1/2\sqrt{3}$	$-1/2\sqrt{3}$	$-1/2\sqrt{3}$
7	$\{1^3\}$	$i/2\sqrt{3}$	$i/2\sqrt{3}$	$i/2\sqrt{3}$	$-i/2\sqrt{3}$	$-i/2\sqrt{3}$	$-i/2\sqrt{3}$	$-i/2\sqrt{3}$	$-i/2\sqrt{3}$	$-i/2\sqrt{3}$	$i/2\sqrt{3}$	$i/2\sqrt{3}$	$i/2\sqrt{3}$

Multiplicity label	Symmetry	$m_1m_2m_3$ 115	151	511	166	616	661	226	262
3	{3}	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$
8	{3}	$\frac{i}{3\sqrt{2}}$	$\frac{i}{3\sqrt{2}}$	$\frac{i}{3\sqrt{2}}$	$\frac{-i}{3\sqrt{2}}$	$\frac{-i}{3\sqrt{2}}$	$\frac{-i}{3\sqrt{2}}$	$\frac{i}{3\sqrt{2}}$	$\frac{i}{3\sqrt{2}}$
4	{21}	$\frac{1}{3}$	$\frac{-1}{6}$	$\frac{-1}{6}$	$\frac{-1}{6}$	$\frac{-1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{-1}{6}$
5	{21}	0	$\frac{1}{2\sqrt{3}}$	$\frac{-1}{2\sqrt{3}}$	$\frac{-1}{2\sqrt{3}}$	$\frac{1}{2\sqrt{3}}$	0	0	$\frac{1}{2\sqrt{3}}$
9	{21}	$\frac{i}{3}$	$\frac{-i}{6}$	$\frac{-i}{6}$	$\frac{i}{6}$	$\frac{i}{6}$	$\frac{-i}{3}$	$\frac{i}{3}$	$\frac{-i}{6}$
10	{21}	0	$\frac{i}{2\sqrt{3}}$	$\frac{-i}{2\sqrt{3}}$	$\frac{i}{2\sqrt{3}}$	$\frac{-i}{2\sqrt{3}}$	0	0	$\frac{i}{2\sqrt{3}}$

$m_1m_2m_3$ 622	334	343	433	442	424	244	553	535	355
$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$	$\frac{1}{3\sqrt{2}}$
$\frac{i}{3\sqrt{2}}$	$\frac{i}{3\sqrt{2}}$	$\frac{i}{3\sqrt{2}}$	$\frac{i}{3\sqrt{2}}$	$\frac{-i}{3\sqrt{2}}$	$\frac{-i}{3\sqrt{2}}$	$\frac{-i}{3\sqrt{2}}$	$\frac{-i}{3\sqrt{2}}$	$\frac{-i}{3\sqrt{2}}$	$\frac{-i}{3\sqrt{2}}$
$\frac{-1}{6}$	$\frac{1}{3}$	$\frac{-1}{6}$	$\frac{-1}{6}$	$\frac{1}{3}$	$\frac{-1}{6}$	$\frac{-1}{6}$	$\frac{1}{3}$	$\frac{-1}{6}$	$\frac{-1}{6}$
$\frac{-1}{2\sqrt{3}}$	0	$\frac{1}{2\sqrt{3}}$	$\frac{-1}{2\sqrt{3}}$	0	$\frac{1}{2\sqrt{3}}$	$\frac{-1}{2\sqrt{3}}$	0	$\frac{1}{2\sqrt{3}}$	$\frac{-1}{2\sqrt{3}}$
$\frac{-i}{6}$	$\frac{i}{3}$	$\frac{-i}{6}$	$\frac{-i}{6}$	$\frac{-i}{3}$	$\frac{i}{6}$	$\frac{i}{6}$	$\frac{-i}{3}$	$\frac{i}{6}$	$\frac{i}{6}$
$\frac{-i}{2\sqrt{3}}$	0	$\frac{i}{2\sqrt{3}}$	$\frac{-i}{2\sqrt{3}}$	0	$\frac{-i}{2\sqrt{3}}$	$\frac{i}{2\sqrt{3}}$	0	$\frac{-i}{2\sqrt{3}}$	$\frac{i}{2\sqrt{3}}$

[a]The m values of j_2 are labeled one through six. All triples $(m_1m_2m_3)$ not given here have value zero.

$$\delta^{m_3\dot{m}_4}\begin{pmatrix} m_5 & \\ & \dot{m}_4 \end{pmatrix}\langle j_2 j_2|\, j_2\rangle^{m_1 m_2}{}_{(1)r_1 m_3}$$
$$= M(23)^{r_2}{}_{r_1}\delta^{\dot{m}_6 m_7}\begin{pmatrix} m_2 & \\ & \dot{m}_6 \end{pmatrix}\langle j_2 j_2|\, j_2\rangle^{m_1 m_5}{}_{(1)r_2 m_7},$$

and so on, where the positive sign has been used in Eq. (7.9) of I.

Turning to the symbols of Sec. 6–8, consider the case when all ICRs are equal to j_2. There are 16 different recoupling coefficient tensors, each with a total multiplicity range of 625, four quasi-6*j* tensors with 2500 multiplicity values, and one 6*j* tensor with 10 000 multiplicity values. Due to the simplicity of this 3*jm* tensor the reality of the 6*jm* tensor sets many of these values to zero, namely whenever an odd number of r_1, r_2, r_3, and r_4 have the values six to ten. For the remaining values the properties of the 6*j* as in Table I can often be used to produce a large number of values from a given one.

For example, consider the case when one row of the 3*jm* tensor possesses {3} and the remaining three possess {21} symmetry. Then by use of the third entry in Table I and the permutation matrices of Butler[9] we find, e.g.,

$$\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1444}$$
$$= -\frac{1}{\sqrt{3}}\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1544} - \frac{1}{\sqrt{3}}\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1454} + \begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1554}.$$

Clearly, the use of this table allows one to produce a vast number of equations relating different components. They are not however in a very useful form as it is not even obvious how many independent components are required. Additional matrix work shows that for the components related by these symmetries, only two are needed:

$$\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1444} = -\frac{1}{2\sqrt{3}}\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1544} + \frac{1}{2\sqrt{3}}\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1555},$$

$$\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1445} = -\frac{1}{2}\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1544} + \frac{1}{2}\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1555},$$

$$\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1454} = -\frac{1}{2}\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1544} + \frac{1}{2}\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1555},$$

$$\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1455} = \frac{\sqrt{3}}{2}\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1544} + \frac{1}{2\sqrt{3}}\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1555},$$

$$\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1545} = \frac{1}{\sqrt{3}}\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1555},$$

$$\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1554} = \frac{1}{\sqrt{3}}\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1555},$$

with similar simplified relations when r_1 is also permuted. (A more elegant argument using the permutation group of the multiplicity labels is being studied, and will be reported in a context applicable to nonambivalent groups.)[12] A direct calculation using Eq. (8.1) gives

$$\begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1554} = \frac{1}{18} \quad \text{and} \quad \begin{Bmatrix} j_2 & j_2 & j_2 \\ j_2 & j_2 & j_2 \end{Bmatrix}_{1555} = 0,$$

from which all others can be calculated.

From the 6*j* tensor immediately follow the four quasi-6*j* tensors from Eq. (8.9). Some adjustments in the multiplicity labeling and additional dummy labels are required as there are two coupling coefficient tensors $\langle j_2 j_2|\, j_2\rangle$.

Of the 16 recoupling coefficient tensors only one is required if we fix on a single coupling coefficient tensor. (The example of Sec. 6 shows what happens if different tensors from the same coupling are used.) In this case we could choose the coupling coefficient from the first five rows of the 3*jm* tensor. The recoupling coefficient is then real as it is the product of four real coupling coefficients. From this it is diagonally constant in the *m*-values as for a type (c) ICR the commuting matrix is

$$\begin{pmatrix} zI & 0 \\ 0 & z^*I \end{pmatrix}.$$

Thus a satisfactory recoupling coefficient tensor may be found from the 6*j* tensor.

11. CONCLUSION

The weak form of Schur's Lemma for gray groups has been shown to have drastic consequences for the *n-j* symbols because the essentially unique 6*j* symbol of representation theory with excellent properties separates into a number of different tensors lacking many of these properties. This was foreshadowed on I by the difference between the double and triple direct products of ICRs. For every gray group we have examined, this situation can be retrieved by symmetrizing the coupling coefficient although care still needs to be exercised in using the various symbols. A proof that this can always be done is lacking but it is encouraging that no counterexamples have been found.

[1] J. D. Newmarch and R. M. Golding, J. Math Phys. **22**, 233 (1981).
[2] J. D. Newmarch and R. M. Golding, J. Math. Phys. **22**, 2113 (1981).
[3] E. P. Wigner, *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
[4] J. D. Newmarch and R. M. Golding "The Character Table for the Corepresentations of Magnetic Groups," J. Math. Phys. **23**, 695 (1982).
[5] A. Aviran and J. Zak, J. Math. Phys. **12**, 2138 (1968).
[6] J. D. Newmarch, Ph.D. thesis, University of New South Wales (1981).
[7] J. N. Kotzev and M. I. Aroyo, J. Phys. A **13**, 2275 (1980) [In their Eq. (16) the authors assume a certain matrix to be Hermitian. While such a *choice* can always be made when the branching multiplicity is only one, in general the assumption is not true.]
[8] J. R. Derome and W. T. Sharp, J. Math. Phys. **6**, 1584 (1965).
[9] P. H. Butler, Philos. Trans. R. Soc. London Ser. A **277**, 545 (1975).
[10] B. G. Basmaji, Am. Math. Monthly **86**, 47 (1979).
[11] C. W. Curtis and I. Riener, *Representation Theory of Finite Groups and Associative Algebras* (Wiley–Interscience, New York, 1962).
[12] J. D. Newmarch, J. Math Phys. **24**, 451 (1983).

On the symmetries of the 6j symbol

J. D. Newmarch
Department of Physics, Universiti Pertanian Malaysia, Serdang, Selangor, Malaysia[a)]

(Received 21 July 1981; accepted for publication 18 September 1981)

The 6j tensor for compact groups is shown to transform as a basis vector for the identity representation of the permutation group S_4. This allows character theory to be used to determine the minimum number of independent components and a projection operator to determine the relations between components—the symmetry properties.

PACS numbers: 02.20.Hj, 03.65.Fd

1. INTRODUCTION

Following the theoretical work of Derome and Sharp,[1] 3jm and 6j tensors may be defined for any compact linear[2] or compact grey[3,4] group. The 3jm tensor, which has been developed from the 3j symbol of Wigner, reduces a triple product of (co) representations to the identity 1 of the group. In addition to the generalized azimuthal quantum numbers m_1, m_2, m_3, the 3jm tensor also depends on a multiplicity index r which spans a range equal to the multiplicity of 1 in the triple product of (co) representations $j_1 \otimes j_2 \otimes j_3$. The symmetry properties of the 3jm tensor generalize[5] from those of the Wigner 3j symbol: the simple phase factors are replaced by permutation matrices in the multiplicity indices. When the three representations are equivalent these matrices form a representation of the permutation group S_3 and are completely determined (to within a unitary transformation) by the properties of the group. When only two are equivalent, two of the matrices generate a representation of S_2. These again are completely determined (understood: to within a unitary transformation) whereas the others are highly arbitrary. When none are equivalent the permutation symmetry is that of the trivial group S_1 and the matrices are highly arbitrary.

The 6j tensor[1,4] is defined as a certain invariant product of four 3jm tensors. It depends on the four multiplicity indices but not on the generalized azimuthal quantum numbers. The permutation properties of the 6j tensor depend on the permutation matrices of the four 3jm tensors and, with the addition of some $1-j$ phase factors, are completely determined by these matrices. This produces some rather extreme cases: when the 3j symmetries are arbitrary, so are the 6j symmetries; on the other hand, when all the representations are equivalent, these are completely determined by group properties.

In this paper a unified method of dealing with all cases is developed. The theory is straightforward and is given in Sec. 2. It shows that a set of permuted 6j tensors transforms as a basis vector for the totally symmetric representation [4] of the permutation group S_4. Character theory may then be used to determine the minimum number of independant components required and a projection operator to find relations between components. These may be used as replacements for the symmetries induced from the 3jm tensors, which were found to be quite unmanageable in the presence of [21] symmetry.[4] Sections 2 and 3 illustrate this theory for a number of cases.

A knowledge of the papers by Derome and Sharp[1,5] or of Butler[2] is assumed and their notations are used. For simplicity only representations of linear groups are used, although all results given here apply equally to the corepresentations of grey groups.

2. THEORY

The 6j tensor is defined by[1,2]

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}_{r_1 r_2 r_3 r_4} = (j_1 j_5 j_6)_{r_1 m_1}{}^{m_5}{}_{m_6} (j_4 j_2 j_6)_{r_2 m_4 m_2}{}^{m_6} \times (j_4 j_5 j_3)^{m_4}_{r_3 m_5 m_3} (j_1 j_2 j_3)^{m_1 m_2 m_3}_{r_1}.$$

The symmetry properties of the 6j tensor are found from the symmetry properties of the 3jm tensors, and are collected in Table I. Thus for example from the first entry,

$$\begin{Bmatrix} j_2 & j_1 & j_3 \\ j_5^* & j_4^* & j_6^* \end{Bmatrix}_{s_2 s_1 s_3 s_4} = \phi_{j_4}\phi_{j_5}\phi_{j6}\, m((12) j_1 j_5^* j_6)_{s_1 r_1}\; m((12) j_4 j_2 j_6^*)_{s_2 r_2} \times m((12) j_4^* j_5 j_3)_{s_3 r_3} m((12) j_1^* j_2^* j_3^*)_{s_4 r_4} \begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}_{r_1 r_2 r_3 r_4},$$

where $m((12) j_1 j_5^* j_6)_{s_1 r_1}$ is the permutation matrix given by

$$(j_5^* j_1 j_6)_{s_1 m_5 m_1 m_6} = m((12) j_1 j_5^* j_6)_{s_1 r_1} (j_1 j_5^* j_6)_{r_1 m_1 m_5 m_6}$$

and so on, and ϕ_{j_i} is the $1-j$ phase factor for j_i.

By examining the ordering of the multiplicity indices, it may be seen that each entry in the table corresponds uniquely to a permutation in S_4, and this is used to label the entries.

Given any 6j tensor

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}_{r_1 r_2 r_3 r_4},$$

it is related to other 6j tensors by Table I. The number of distinct tensors depends on the irreducible representations j_1 to j_6. Thus, if all are inequivalent there are 24 distinct tensors, whereas if they are all equivalent with real character there is only 1. Each component of each of these tensors may be considered as a basis vector for a vector space V of dimension equal to the product of the four 3j multiplicities (M) times the number of distinct 6j tensors (M'). We write

$$V = U_1 \oplus U_2 \oplus \cdots \oplus U_{M'},$$

where each U_i has as basis vectors the M components of the ith 6j tensor. A representation of the permutation group S_4 may now be defined over V by the entries of Table I: each permutation π maps the basis of U_i onto the basis of U_j to give the (j, i) block $d^{ji}(\pi)$ of the matrix $D(\pi)$. For example,

a) Present address: School of Electrical Engineering and Computer Science, University of New South Wales, P.O. Box 1, Kensington, NSW 2033, Australia.

 0022-2488/83/030451-06$02.50

when all the representations are inequivalent, the division into the subspaces U_i may be taken as

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}, \begin{Bmatrix} j_2 & j_1 & j_3 \\ j_5^* & j_4^* & j_6^* \end{Bmatrix}, \begin{Bmatrix} j_3 & j_1 & j_2 \\ j_6 & j_4 & j_5 \end{Bmatrix}, \begin{Bmatrix} j_1 & j_3 & j_2 \\ j_4^* & j_6^* & j_5^* \end{Bmatrix}, \cdots.$$

This gives for the block structure of $D((12))$

$$\mathrm{D}((12)) = \begin{pmatrix} 0 & d^{12}((12)) & 0 & 0 & \cdots \\ d^{21}((12)) & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & d^{34}((12)) & \cdots \\ 0 & 0 & d^{43}((12)) & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

Each block $d^{ji}(\pi)$ is then determined directly from the entry for π in Table I by relating U_j back to U_i:

$$\begin{aligned} d^{21}((12))_{s_2s_1s_3s_4,\, r_1r_2r_3r_4} &= \phi_{j_4}\phi_{j_5}\phi_{j_6} m((12)j_1j_5^*j_6)_{s_1r_1}\ m((12)j_4j_2j_6^*)_{s_2r_2} \\ &\times m((12)j_4^*j_5j_3)_{s_3r_3}\ m((12)j_1^*j_2^*j_3^*)_{s_4r_4}, \\ d^{12}((12))_{s_2s_1s_3s_4,\, r_1r_2r_3r_4} &= \phi_{j_5}^*\phi_{j_4}^*\phi_{j_6}^*\ m((12)j_2j_4j_6^*)_{s_1r_1}\ m((12)j_5^*j_1j_6)_{s_2r_2} \\ &\times m((12)j_5j_4^*j_3)_{s_3r_3}\ m((12)j_2^*j_1^*j_3^*)_{s_4r_4}, \end{aligned}$$

and so on. By the multiplicative properties of the $3j$ permutation matrices[1,2,5] it readily follows that the matrices $D = \{D(\pi) : \pi \epsilon S_4\}$ form a representation of S_4. Generally, of course, it is reducible.

We now write the basis vectors of V as a column vector

$$\mathbf{v} = \begin{pmatrix} \begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}_{r_1r_2r_3r_4} \\ \begin{Bmatrix} j_2 & j_1 & j_3 \\ j_5^* & j_4^* & j_6^* \end{Bmatrix}_{r_1r_2r_3r_4} \\ \vdots \end{pmatrix}.$$

It follows easily that

$$D(\pi)\mathbf{v} = \mathbf{v}.$$

Hence $\mathbf{v}$ transforms as a basis vector for any totally symmetric component [4] of D. The number of such basis vectors equals the multiplicity $n_{[4]}$ of the irreducible representation [4] in D, and so the character of D completely determines the minimum number of components of a $6j$ tensor required to determine all others. The actual relations between components requires a knowledge of these basis vectors, which can be found by use of the projection operator

$$P = \Sigma_{\pi \epsilon S_4} D(\pi),$$

and these may be used to replace the symmetry properties given in Table I.

This approach is rather long-winded when the symmetry properties are simple (e.g., in a simply reducible group). However, when the $3j$ permutation matrices possess a more complex symmetry this method has been found to drastically reduce the labor involved in using the symmetry properties. In Secs. 3 and 4 examples are given to show how the method is used.

3. EXAMPLE

The 24 S_4 permutations map one $6j$ tensor into itself only when all $6j$-values are equal to j, and when $j \equiv j^*$. Setting $\phi_j = 1$ as in the quasi-ambivalent case, the permutation matrices simplify and as representative of each class we have

$$\begin{aligned} D((12))_{s_2s_1s_3s_4,\, r_1r_2r_3r_4} &= m((12)jjj)_{s_1r_1}\ m((12)jjj)_{s_2r_2}\ m((12)jjj)_{s_3r_3}\ m((12)jjj)_{s_4r_4}, \\ D((123))_{s_2s_3s_1s_4,\, r_1r_2r_3r_4} &= m((123)jjj)_{s_1r_1}\ m((123)jjj)_{s_2r_2}\ m((123)jjj)_{s_3r_3}, m((123)jjj)_{s_4r_4}, \\ D((12)(34))_{s_2s_1s_4s_3,\, r_1r_2r_3r_4} &= m(I,jjj)_{s_1r_1}\ m(I,jjj)_{s_2r_2}\ m(I,jjj)_{s_3r_3}\ , m(I,jjj)_{s_4r_4}, \end{aligned}$$

and

$$D((1234))_{s_2s_3s_4s_1,\, r_1r_2r_3r_4} = m((13)jjj)_{s_1r_1}\ m((13)jjj)_{s_2r_2}\ m((13jjj)_{s_3r_3},\ m((13)jjj)_{s_4r_4}.$$

In this case the matrices $\{m(\sigma, jjj) : \sigma \epsilon S_3\}$ form a representation of S_3 and thus the trace $\chi_D(\pi)$ of each matrix $D(\pi)$ may be found in terms of the trace $\chi_{jjj}(\sigma)$ of $m(\sigma, jjj)$. Inspection of Table I shows that $\chi_D(\pi)$ is a class function of S_4 in terms of the class function $\chi_{jjj}(\sigma)$ of S_3. This gives for the character vector χ_D

	$\{1^4\}$	$\{1^22\}$	$\{13\}$	$\{2^2\}$	$\{4\}$
χ_D	$(\chi_{jjj}\{1^3\})^4$	$\chi_{jjj}\{1^3\}(\chi_{jjj}\{12\})^2)$	$\chi_{jjj}\{1^3\}\chi_{jjj}\{3\}$	$(\chi_{jjj}\{1^3\})^2$	$\chi_{jjj}\{1^3\}$

The multiplicity $n_{[4]}$ follows directly from this. It may be cast into different forms by letting

$$\chi_{jjj}(\sigma) = n_{[3]}\chi_{[3]}(\sigma) + n_{[21]}\chi_{[21]}(\sigma) + n_{[1^3]}\chi_{[1^3]}(\sigma), \quad (3.2)$$

which gives

$$n_{[4]} = (n_{[3]} + 2n_{[21]} + n_{[1^3]})\{(n_{[3]} + 2n_{[21]} + n_{[1^3]})^3 + 6(n_{[3]} - n_{[1^3]})^2 + 11n_{[3]} - 2n_{[21]} + 11n_{[1^3]} + 6\}/24 \quad (3.3)$$

or, by writing χ_{jjj} in terms of the character χ_j of the representation j,

$$n_{[4]} = \frac{1}{24|G|^4}\int_G [\chi_j(u)]^3\, du\left\{\left(\int_G \chi_j(u)^3\, du\right)^3 + 6|G| \times \left(\int_G \chi_j(u^2)\chi_j(u)\, du\right)^2 + |G|^2\int_G 3[\chi_j(u)]^3 + 8\chi_j(u)^3\, du + 6|G|^3\right\}. \quad (3.4)$$

The meaning of these equations may be illustrated by taking special cases. For example, if $n_{[21]} = n_{[1^3]} = 0$,

$$n_{[4]} = (n_{[3]}/24)(n_{[3]} + 1)(n_{[3]} + 2)(n_{[3]} + 3).$$

Thus if the $3j$ multiplicity is 2, the $6j$ tensor has 16 components, of which only 5 are independent. These are of course

$$\begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{1111}, \begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{1112}, \begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{1122}, \begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{1222}, \text{ and } \begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{2222}.$$

TABLE I. Permutations of the 6*j* tensor.[a]

S_4 permutation	$a\ b\ c\ d$	$j_1'\ j_2'\ j_3'\ j_4'\ j_5'\ j_6'$	Φ	σ
(12)	2 1 3 4	$j_2\ j_1\ j_3\ j_5^*\ j_4^*\ j_6^*$	$\phi_{j_4}\phi_{j_5}\phi_{j_6}$	(12)
(13)	3 2 1 4	$j_3\ j_2\ j_1\ j_6^*\ j_5^*\ j_4^*$	$\phi_{j_4}\phi_{j_5}\phi_{j_6}$	(13)
(14)	4 2 3 1	$j_1^*\ j_6^*\ j_5\ j_4\ j_3\ j_2^*$	$\phi_{j_1}\phi_{j_2}\phi_{j_6}$	(23)
(23)	1 3 2 4	$j_1\ j_3\ j_2\ j_4^*\ j_6^*\ j_5^*$	$\phi_{j_4}\phi_{j_5}\phi_{j_6}$	(23)
(24)	1 4 3 2	$j_6\ j_2^*\ j_4^*\ j_3^*\ j_5\ j_1$	$\phi_{j_2}\phi_{j_3}\phi_{j_4}$	(13)
(34)	1 2 4 3	$j_5^*\ j_4\ j_3^*\ j_2\ j_1^*\ j_6$	$\phi_{j_1}\phi_{j_3}\phi_{j_5}$	(12)
(123)	2 3 1 4	$j_2\ j_3\ j_1\ j_5\ j_6\ j_4$	1	(123)
(124)	2 4 3 1	$j_6^*\ j_1^*\ j_5\ j_3^*\ j_4^*\ j_2$	$\phi_{j_1}\phi_{j_3}\phi_{j_4}\phi_{j_5}$	(132)
(134)	3 2 4 1	$j_5\ j_6^*\ j_1^*\ j_2\ j_3^*\ j_4^*$	$\phi_{j_1}\phi_{j_3}\phi_{j_4}\phi_{j_5}$	(123)
(132)	3 1 2 4	$j_3\ j_1\ j_2\ j_6\ j_4\ j_5$	1	(132)
(142)	4 1 3 2	$j_2^*\ j_6\ j_4^*\ j_5^*\ j_3\ j_1^*$	$\phi_{j_1}\phi_{j_2}\phi_{j_4}\phi_{j_5}$	(123)
(143)	4 2 1 3	$j_3^*\ j_4\ j_5^*\ j_6^*\ j_1\ j_2^*$	$\phi_{j_2}\phi_{j_3}\phi_{j_5}\phi_{j_6}$	(132)
(234)	1 3 4 2	$j_6\ j_4^*\ j_2^*\ j_3\ j_1^*\ j_5^*$	$\phi_{j_1}\phi_{j_2}\phi_{j_4}\phi_{j_5}$	(132)
(243)	1 4 2 3	$j_5^*\ j_3^*\ j_4\ j_2^*\ j_6^*\ j_1$	$\phi_{j_2}\phi_{j_3}\phi_{j_5}\phi_{j_6}$	(123)
(12)(34)	2 1 4 3	$j_4^*\ j_5^*\ j_3\ j_1^*\ j_2^*\ j_6$	$\phi_{j_1}\phi_{j_2}\phi_{j_4}\phi_{j_5}$	I
(13)(24)	3 4 1 2	$j_4^*\ j_2^*\ j_6\ j_1^*\ j_5^*\ j_3$	$\phi_{j_1}\phi_{j_4}\phi_{j_4}\phi_{j_2}$	I
(14)(23)	4 3 2 1	$j_1^*\ j_5^*\ j_6^*\ j_4^*\ j_2^*\ j_3^*$	$\phi_{j_1}\phi_{j_3}\phi_{j_4}\phi_{j_6}$	I
(1234)	2 3 4 1	$j_6^*\ j_5^*\ j_1^*\ j_3\ j_2^*\ j_4$	$\phi_{j_1}\phi_{j_2}\phi_{j_6}$	(13)
(1243)	2 4 1 3	$j_4\ j_3^*\ j_5^*\ j_1^*\ j_6\ j_2$	$\phi_{j_1}\phi_{j_3}\phi_{j_5}$	(23)
(1324)	3 4 2 1	$j_5\ j_1^*\ j_6^*\ j_2^*\ j_4\ j_3$	$\phi_{j_1}\phi_{j_2}\phi_{j_6}$	(12)
(1342)	3 1 4 2	$j_4^*\ j_6\ j_2^*\ j_1\ j_3^*\ j_5$	$\phi_{j_2}\phi_{j_3}\phi_{j_4}$	(23)
(1423)	4 3 1 2	$j_2^*\ j_4^*\ j_6\ j_5\ j_1\ j_3^*$	$\phi_{j_2}\phi_{j_3}\phi_{j_4}$	(12)
(1432)	4 1 2 3	$j_3^*\ j_5^*\ j_4\ j_6\ j_2\ j_1^*$	$\phi_{j_1}\phi_{j_3}\phi_{j_5}$	(13)

[a] $\begin{Bmatrix} j_1' & j_2' & j_3' \\ j_4' & j_5' & j_6' \end{Bmatrix}_{s_\sigma t_\sigma u_\sigma v_\sigma} = \Phi m(\sigma, j_1 j_5^* j_6)_{s_1 r_1} m_2(\sigma, j_4 j_2 j_6^*)_{s_2 r_2} m_3(\sigma, j_4^* j_5 j_3)_{s_3 r_3} m_4(\sigma, j_1^* j_2^* j_3^*) \begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}_{r_1 r_2 r_3 r_4}$.

Similarly, if $n_{[3]} = n_{[1^3]} = 0$ and $n_{[21]} = 1$, there are 16 components, of which only 1 is required. The representation $j = [321]$ of the symmetric group S_6 is of interest in that it was the first nonsimple phase representation discovered.[5] Here $n_{[3]} = 2$, $n_{[21]} = 1$, and $n_{[1^3]} = 1$. Hence there are 35 independent 6*j* components out of a total of 625.

The equations above hold whatever the choice of basis in the 3*j* multiplicity space. However, in most practical cases it will almost certainly be chosen to block diagonalize the permutation matrices according to symmetry type. For example we may write

$$m(\sigma_{jjj}) = \begin{pmatrix} m_1^{[\lambda_1]}(\sigma) & & 0 \\ & m_2^{[\lambda_2]}(\sigma) & \\ 0 & & \ddots \end{pmatrix}$$

or

$$\begin{pmatrix} m^{[3]}(\sigma)\, I_1 & & 0 \\ & m^{[1^3]}(\sigma)\, I_2 & \\ 0 & & m^{[21]}(\sigma) I_3 \end{pmatrix},$$

where $m_i^{[\lambda]}(\sigma)$ is irreducible. Since we are about to introduce a large number of different multiplicity values we shall adopt the first form as it easily allows identification of one or a pair of multiplicity values with i in $\mathrm{m}_i^{[\lambda]}$. The argument is independent of the form, which is primarily notational.

The block diagonalization of these matrices allows an easy partial reduction of the representation D. Given four sets of multiplicity values corresponding to the ordered set i_1, i_2, i_3, i_4, the S_4 permutations transform these sets among themselves and form a closed subspace under S_4. The dimension of this subspace equals the product of the orders of each multiplicity set times the number of distinct permutations of i_1, i_2, i_3, and i_4. (The method in miniature is the same as the method applied to full 6*j* tensors.) The character depends on the matrices m_i and the equality or inequality of i_1, i_2, i_3, and i_4. Consider one case: $i_1 = i_2 = i_3 \neq i_4$. There are four permuted sets of multiplicity values corresponding to $i_1 i_1 i_1 i_4$, $i_1 i_1 i_4 i_1$, $i_1 i_4 i_1 i_1$, and $i_4 i_1 i_1 i_1$. A typical matrix over this subspace is

$$D'((12))_{js_2s_1s_3s_4,\ ir_1r_2r_3r_4} = \begin{pmatrix} m_1^{[\lambda_1]}((12))_{s_1r_1}\ m_1^{[\lambda_1]}((12))_{s_2r_2} & 0 & 0 & 0 \\ \times m_1^{[\lambda_1]}((12))_{s_3r_3}\ m_4^{[\lambda_4]}((12))_{s_4r_4} & m_1^{[\lambda_1]}((12))_{s_1r_1}\ m_1^{[\lambda_1]}((12))_{s_2r_2} & 0 & 0 \\ 0 & \times m_4^{[\lambda_4]}((12))_{s_3r_3}\ m_1^{[\lambda_1]}((12))_{s_4r_4} & 0 & m_1^{[\lambda_1]}((12))_{s_1r_1}\ m_4^{[\lambda_4]}((12))_{s_2r_2} \\ 0 & 0 & m_4^{[\lambda_4]}((12))_{s_1r_1}\ m_1^{[\lambda_1]}((12))_{s_2r_2} & \times m_1^{[\lambda_1]}((12))_{s_3r_3}\ m_1^{[\lambda_1]}((12))_{s_4r_4} \\ 0 & 0 & \times m_1^{[\lambda_1]}((12))_{s_1r_1}\ m_1^{[\lambda_1]}((12))_{s_4r_4} & 0 \end{pmatrix}.$$

TABLE II. The character vector χ' in terms of the character vectors $\chi_I^{[\lambda_4]}$ of the reduced matrices $m_I^{[\lambda_4]}$ when all j values are equal.

	$\{1^4\}$	$\{1^2 2\}$	$\{13\}$	$\{2^2\}$	$\{4\}$
$i_1=i_2=i_3=i_4$	$(\chi_1^{[\lambda_4]}\{1^3\})^4$	$\chi_1^{[\lambda_4]}\{1^3\}(\chi_1^{[\lambda_4]}\{12\})^2$	$\chi_1^{[\lambda_4]}\{1^3\}\chi_1^{[\lambda_4]}\{3\}$	$(\chi_1^{[\lambda_4]}\{1^3\})^2$	$\chi_1^{[\lambda_4]}\{1^3\}$
$i_1=i_2=i_3\neq i_4$	$4(\chi_1^{[\lambda_4]}\{1^3\})^3\chi_4^{[\lambda_4]}\{1^3\}$	$2\chi_1^{[\lambda_4]}\{1^3\}\chi_1^{[\lambda_4]}\{12\}\chi_4^{[\lambda_4]}\{12\}$	$\chi_1^{[\lambda_4]}\{1^3\}\chi_4^{[\lambda_4]}\{3\}$	0	0
$i_1=i_2\neq i_3=i_4$	$6(\chi_1^{[\lambda_4]}\{1^3\})^2\chi_3^{[\lambda_3]}\{1^3\})^2$	$\chi_1^{[\lambda_4]}\{1^3\}(\chi_3^{[\lambda_3]}\{12\})^2$ $+\chi_3^{[\lambda_3]}\{1^3\}(\chi_1^{[\lambda_4]}\{12\})^2$	0	$2\chi_1^{[\lambda_4]}\{1^3\}\chi_3^{[\lambda_3]}\{1^3\}$	0
$i_1=i_2\neq i_3\neq i_4\neq i_1$	$12(\chi_1^{[\lambda_4]}\{1^3\})^2\chi_3^{[\lambda_3]}\{1^3\}$ $\times\chi_4^{[\lambda_4]}\{1^3\}$	$2\chi_1^{[\lambda_4]}\{1^3\}\chi_3^{[\lambda_3]}\{12\}\chi_4^{[\lambda_4]}\{12\}$	0	0	0
none equal	$24\chi_1^{[\lambda_4]}\{1^3\}\chi_2^{[\lambda_2]}\{1^3\}$ $\times\chi_3^{[\lambda_3]}\{1^3\}\chi_4^{[\lambda_4]}\{1^3\}$	0	0	0	0

from which the character is

$$\chi'((12)) = 2\chi_{jjj}^{[\lambda_4]}(\mathrm{I})\,\chi_{jjj}^{[\lambda_4]}((12))\,\chi_{jjj}^{[\lambda_4]}((12))$$

or

$$\chi'\{1^2 2\} = 2\chi_{jjj}^{[\lambda_4]}\{1^3\}\,\chi_{jjj}^{[\lambda_4]}\{12\}\,\chi_{jjj}^{[\lambda_4]}\{12\}.$$

A complete list of character vectors is given in Table II. By substitution of the various S_3 characters into this the $n_{[4]}$ multiplicities may be readily calculated. When none of i_1, i_2, i_3, or i_4 are equal, $n_{[4]}$ is the product of the multiplicity ranges corresponding to the labels i_1, i_2, i_3, and i_4. When all are equal $n_{[4]}$ is easily shown to be one for all of $[\lambda_4] = [3]$, $[21]$, and $[1^3]$. The remaining cases are given in Tables III–V. Some of these may be compared to results in the literature. The tables of Butler[6] which extend the earlier work of Butler and Wybourne,[7] give all components of the $6j$ tensor

$$\begin{Bmatrix} T & T & T \\ T & T & T \end{Bmatrix}_{r_1 r_2 r_3 r_4},$$

where T is the three-dimensional simple phase representation of the tetrahedral group. The $3j$ permutation matrix is equivalent to $[3] \oplus [1^3]$, and Tables III and IV and equation (3.3) are in agreement with their results.

One which has not appeared before is when $i_1 = i_2 = i_3 = i_4$ and the symmetry $[\lambda] = [21]$. Using the ordering of components 1111, 1112, 1121, 1122, 1211, 1212, 1221, 1222, 2111, 2112, 2121, 2122, 2211, 2212, 2221, 2222, the projection operator is easily found from which the single [4] basis vector is

$$(9\,0\,0\,3\,0\,3\,3\,0\,0\,3\,3\,0\,3\,0\,0\,9)^T,$$

giving

$$\begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{1111} = 3\begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{1122} = 3\begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{1212} = 3\begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{1221} = 3\begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{2112} = 3\begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{2121} = 3\begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{2211} = \begin{Bmatrix} j & j & j \\ j & j & j \end{Bmatrix}_{2222},$$

with all other components zero.

4. FURTHER EXAMPLES

A closed set of three $6j$ tensors under S_4 permutations is formed when one of the $6j$ tensors is

$$\begin{Bmatrix} j_1 & j_1 & j_2 \\ j_1 & j_1 & j_2 \end{Bmatrix}_{r_1 r_2 r_3 r_4}$$

and $j_1 \equiv j_1^*, j_2 \equiv j_2^*$. The $1-j$ phase factors are $\phi_{j_1} = \pm 1, \phi_{j_2} = \pm 1$. Each $3jm$ tensor is formed of the triple $(j_1 j_1 j_2)$ and hence only those S_2 permutations of the two equal j values are determined by group theory. By the same techniques as before, typical matrices over the basis

$$\begin{Bmatrix} j_1 & j_1 & j_2 \\ j_1 & j_1 & j_2 \end{Bmatrix}, \quad \begin{Bmatrix} j_1 & j_2 & j_1 \\ j_1 & j_2 & j_1 \end{Bmatrix}, \quad \begin{Bmatrix} j_2 & j_1 & j_1 \\ j_2 & j_1 & j_1 \end{Bmatrix}$$

are

TABLE III. The multiplicity $n_{[4]}$ using reduced matrices when $i_1 = i_2 = i_3 \neq i_4$ for all j values equal.

$[\lambda_1]$	$[\lambda_4]$	Total number of components	Number $n_{[4]}$ of independent components
[3]	[3]	4	1
[3]	[21]	8	0
[3]	$[1^3]$	4	0
[21]	[3]	32	2
[21]	[21]	64	2
[21]	$[1^3]$	32	2
$[1^3]$	[3]	4	0
$[1^3]$	[21]	8	0
$[1^3]$	$[1^3]$	4	1

TABLE IV. The multiplicity $n_{[4]}$ using reduced matrices when $i_1 = i_2 \neq i_3 = i_4$ for all j values equal.

$[\lambda_1]$	$[\lambda_3]$	Total number of components	Number $n_{[4]}$ of independent components
[3]	[3]	6	1
[3]	[21]	24	2
[3]	$[1^3]$	6	1
[21]	[21]	96	5
[21]	$[1^3]$	24	2
$[1^3]$	$[1^3]$	6	1

$$D((12)_{js_2s_1s_3s_4,\ ir_1r_2r_3r_4}$$

$$= \begin{pmatrix} m((12)j_1j_1j_2)_{s_1r_1}\, m((12)j_1j_1j_2)_{s_2r_2} & 0 & 0 \\ \times m((12)j_1j_1j_2)_{s_3r_3}\, m((12)j_1j_1j_2)_{s_3r_3} & & \\ 0 & 0 & m((12)j_2j_1j_1)_{s_1r_1}\, m((12)j_2j_1j_1)_{s_2r_2} \\ & & \times m((12)j_2j_1j_1)_{s_3r_3}\, m((12)j_2j_1j_1)_{s_4r_4} \\ 0 & m((12)j_1j_2j_1)_{s_1r_1}\, m((12)j_1j_2j_1)_{s_2r_2} & 0 \\ & \times m((12)j_1j_2j_1)_{s_3r_3}\, m((12)j_1j_2j_1)_{s_4r_4} & \end{pmatrix}$$

and

$$D((12)(34))_{js_2s_1s_4s_3,\ ir_1r_2r_3r_4}$$

$$= \begin{pmatrix} m(I,j_1j_1j_2)_{s_1r_1}\, m(I,j_1j_1j_2)_{s_2r_2} & 0 & 0 \\ \times m(I,j_1j_1j_2)_{s_3r_3}\, m(I,j_1j_1j_2)_{s_4r_4} & & \\ 0 & m(I,j_1j_2j_1)_{s_1r_1}\, m(I,j_1j_2j_1)_{s_2r_2} & 0 \\ & \times m(I,j_1j_2j_1)_{s_3r_3}\, mI,j_1j_2j_1)_{s_4r_4} & \\ 0 & 0 & m(I,j_2j_1j_1)_{s_1r_1}\, m(I,j_2j_1j_1)_{s_2r_2} \\ & & \times m(I,j_2j_1j_1)_{s_3r_3}\, m(I,j_2j_1j_1)_{s_4r_4} \end{pmatrix}.$$

These combine to give the character vector

	$\{1^4\}$	$\{1^22\}$	$\{13\}$	$\{2^2\}$	$\{4\}$
χ_D	$3(\chi_{j_1j_1j_2}\{1^2\})^4$	$\chi_{j_1j_1j_2}\{1^2\}(\chi_{j_1j_1j_2}\{2\})^2$	0	$3(\chi_{j_1j_1j_2}\{1^2\}$	$\chi_{j_1j_1j_2}\{1^2\}(\chi_{j_1j_1j_2}\{2\})^2$

As in Sec. 3 this may be broken down into subrepresentations according to the block diagonalization of the $3jm$ permutation matrices. No new ideas are involved.

An interesting case occurs for the $6j$ tensor

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_1 & j_2 & j_3 \end{Bmatrix},$$

with $j_1 \equiv j_1^*, j_2 \equiv j_2^*, j_3 \equiv j_3^*$. S_4 permutations generate only six tensors, while the $3jm$ permutation matrices are essentially arbitrary. However, in this case it is easy to see that the only nontrivial elements of S_4 whose character does not vanish belong to the class $\{2^2\}$, i.e., those which interchange a pair on the top row with the corresponding pair on the bottom. The elements of this class all use the identity permutation of the $3jm$ tensors and hence the character is completely determined from a knowledge of the multiplicities alone:

	$\{1^4\}$	$\{1^22\}$	$\{13\}$	$\{2^2\}$	$\{3\}$
χ_D	$6(\chi_{j_1j_2j_3}\{1\})^4$	0	0	$6(\chi_{j_1j_2j_3}\{1\})^2$	0

For example in the gray tetrahedral group the multiplicity of A in the product $U' \otimes E' \otimes E$ is two.[3] This gives 5 independent components from a total of 16, no matter what form the $3j$ permutation matrices are in.

By way of contrast, the $6j$ tensor

$$\begin{Bmatrix} j_1 & j_1 & j_2 \\ j_1 & j_1^* & j_1 \end{Bmatrix}_{s_1s_2s_3s_4}$$

with $j_1 \not\equiv j_1^*$ and $j_2 \not\equiv j_2^*$, contains a wealth of information as each $3jm$ permutation matrix is determined by either S_2 or S_3 symmetries. However, none of this is used in the character (apart from multiplicities) as the S_4 permutations generate twenty-four different $6j$ tensors. The only non-vanishing character is in the class $\{1^4\}$ where it is

$$24\ \chi_{j_1j_1j_1}\{1^3\}\ \chi_{j_1j_1j_1^*}\{1^2\}\ \chi_{j_1^*j_1^*j_2}\{1^2\}\ \chi_{j_1^*j_1^*j_2^*}\{1^2\}.$$

The S_2 and S_3 information is of course used in the projection operator.

This is merely a representative sample of possible $6j$ tensors. All others may be dealt with in the same manner and it is not difficult to calculate projection operators for each.

TABLE V. The multiplicity $n_{[4]}$ using reduced matrices when $i_1 = i_2 \neq i_3 \neq i_4 \neq i_1$ for all j values equal.

$[\lambda_1]$	$[\lambda_3]$	$[\lambda_4]$	Total number of components	Number $n_{[4]}$ of independent components
[3]	[3]	[3]	12	1
[3]	[3]	[21]	24	1
[3]	[3]	$[1^3]$	12	0
[3]	[21]	[21]	48	2
[3]	[21]	$[1^3]$	24	1
[3]	$[1^3]$	$[1^3]$	12	1
[21]	[3]	[3]	48	3
[21]	[3]	[21]	96	4
[21]	[3]	$[1^3]$	48	1
[21]	[21]	[21]	192	8
[21]	[21]	$[1^3]$	96	4
[21]	$[1^3]$	$[1^3]$	48	3
$[1^3]$	[3]	[3]	12	1
$[1^3]$	[3]	[21]	24	1
$[1^3]$	[3]	$[1^3]$	12	0
$[1^3]$	[21]	[21]	48	2
$[1^3]$	[21]	$[1^3]$	24	1
$[1^3]$	$[1^3]$	$[1^3]$	12	1

5. CONCLUSION

With the basic theory of $3jm$ and $6j$ tensors for compact groups given by Derome and Sharp,[1,5] the major problem in applying Racah algebra methods appears to be the computational one of actually finding these tensors for specific

groups. Butler and Wybourne[8] have developed a recursive technique which has been successfully applied to a number of simple phase representations.[6,7,9–13] Many groups, however, have nonsimple phase representations[14,15] and if $6j$ symbols are to be calculated efficiently for these cases it will prove important to know how many components need to be calculated. The method developed here will allow an easy determination of this number. A simple projection operator can be used to give other components.

[1]J.-R. Derome and W. T. Sharp, J. Math. Phys. **6**, 1584 (1965).
[2]P. H. Butler, Philos. Trans. R. Soc. London Ser. A **277**, 545 (1975).
[3]J. D. Newmarch and R. M. Golding, J. Math. Phys. **22**, 233 (1981).
[4]J. D. Newmarch and R. M. Golding, J. Math. Phys. **24**, 441 (1983).
[5]J.-R. Derome, J. Math. Phys. **7**, 612 (1966).
[6]P. H. Butler, in *Recent Advances in Group Theory*, previous paper edited by J. C. Donini (Plenum, New York, 1979).
[7]P. H. Butler and B. G. Wybourne, Int. J. Quantum Chem. **10**, 615 (1976).
[8]P. H. Butler and B. G. Wybourne, Int. J. Quantum Chem. **10**, 581 (1976).
[9]P. H. Butler, Int. J. Quantum Chem. **10**, 599 (1976).
[10]P. H. Butler, R. W. Haase, and B. G. Wybourne, Aust. J. Phys. **31**, 131 (1978).
[11]L. S. R. K. Prasad and K. Bharathi, J. Phys. A **13**, 781 (1980).
[12]D. R. Pooler, J. Phys. A **13**, 1197 (1980).
[13]P. H. Butler and M. F. Reid, J. Phys. A **10**, 1655 (1979).
[14]P. H. Butler and R. C. King, Can. J. Math. **26**, 328 (1974).
[15]A. J. Van Zanten and E. de Vries, Can. J. Math. **27**, 528 (1975).

Models of nonlinear representations and examples of linearization techniques

Georges Pinczon and Jacques C. H. Simon
Faculté des Sciences Mirande, Université de Dijon, B. P. 138, 21004-Dijon Cedex, France

(Received 8 April 1981; accepted for publication 12 February 1982)

Examples of classes of nonlinear representations of Lie groups are given. Nonlinear representations which are a perturbation of a unitary representation of the discrete series of SU(1,1) are then proved to be formally linearizable.

PACS numbers: 02.20.Qs

I. INTRODUCTION

Nonlinear actions of Lie groups in a vector space appear in the study of dynamical systems with symmetries. In most of the examples these actions have a fixed point (say the origin) in which case they are called nonlinear representations (Ref. 1). More precisely, we have the following definitions:

1.1. Definition: A formal representation (S,H) of a real Lie group G, in a Fréchet space H, is a morphism S from G to the group of invertible elements in the space of formal power series of the form $f=\Sigma_{n\geq 1}f_n, f_n$ belonging to the space $\mathscr{L}_n(H)$ of continuous, symmetric n-linear mappings from H to H. Writing $g\to S_g=\Sigma_{n\geq 1}S_g^n, S_g^n\in\mathscr{L}_n(H)$, this morphism, we demand in addition that the mappings $g\to S_g^n(\varphi_1,\ldots,\varphi_n), \varphi_i\in H$, be measurable for any $n\geq 1$. The linear part S^1 is a linear representation of G. If, in addition, H being a Banach space, there exists a neighborhood V of the identity in G such that S_g is an analytic mapping in a neighborhood of the origin U_g in H for every $g\in V$, one says that (S,H) is an analytic representation of G in H.

The interest of introducing larger groups than the real line (i.e., time parameter) has been stressed in Refs. 1–4 for relativistic wave equations, where the Poincaré group generates strong constraints.

Nonlinear representations of Lie groups are studied up to the following equivalences:

1.2. Definition: Two formal (resp. analytic) representations (S,H) and (S',H) of G are formally (resp. analytically) equivalent, if there exists a formal power series (resp. an analytic mapping around the origin) $A=\Sigma_{n\geq 1}A^nA^n\in\mathscr{L}_n(H)$, with A^1 invertible, such that $S'_g\circ A=A\circ S_g$. (S,H) is said to be formally (resp. analytically) linearizable if it is formally (resp. analytically) equivalent to its linear part.

In this article we are mainly interested in properties of some nonlinear representations of low-dimension, noncompact, semisimple Lie groups. The choice of these groups is not arbitrary: They are simple groups which appear as subgroups of most groups presenting an interest in physics.

Starting calculations with these groups, it appeared that the computation of some representations applies to any Lie group. These nonlinear representations are studied in Sec. III: They are denoted by $S^{(p,\xi)}$, $T^{(p,\xi)}$, and $U^{(p,\xi)}$, and their definition is given in 3.4 and 3.9. What is suprising is the universality of the formulae defining these representations: The Lie group structure does not appear explicitly. In Sec. IV, we study formal representations of SU(1,1), the linear part of which belongs to the discrete series.

A representation (L,H) belonging to the discrete series of SU(1,1) extends to the space $\mathscr{A}$ of analytic functions on the open unit disc. It also has a restriction to its space $\mathscr{S}_1$ of differentiable vectors. One proves that any formal representation of SU(1,1) on $\mathscr{A}$ (resp. $\mathscr{S}_1$), the linear part of which is the extension of L to $\mathscr{A}$ (resp. restriction of L to $\mathscr{S}_1$), is formally linearizable. In the Hilbert space H this is not necessarily the case: A counterexample is given by the series $S^{(p,\xi)}$ introduced in Sec. III. Since these formal representations extend to $\mathscr{A}$ and leave $\mathscr{S}_1$ invariant, this stresses the important role played by the topology of the representation space concerning linearization. Finally, let us mention that when the linear part of a formal representation of SU(1,1) belongs to the continuous series, some linearization properties remain: Algebraic spaces of 1-cohomology vanish. We hope to come back later to this question.

II. NOTATIONS AND GENERAL DEFINITIONS

In the following, G is a connected real Lie group, $\mathscr{G}_0$ its Lie algebra, $\mathscr{G}$ the complexified of $\mathscr{G}_0$, $\mathscr{U}(\mathscr{G})$ the enveloping algebra of $\mathscr{G}$.

Given two Fréchet spaces H_1 and H_2, one denotes by $\mathscr{L}_n(H_1,H_2)$ the space of continuous, symmetric n-linear maps from H_1 to H_2. We denote by $\mathscr{L}_n(H)=\mathscr{L}_n(H,H)$ and $\mathscr{F}_n(H)=\mathscr{L}_n(H,\mathbb{C})$. Given a continuous linear representation (V,H) of G on a Fréchet space H, we denote by H^∞ its space of C^∞ vectors. The differential representation dV of $\mathscr{G}$ acts on H^∞.

On the space $\mathscr{L}_n(H)$ a linear representation $V^{(n)}$ of G is defined by $(V_g^{(n)})T(\varphi_1,\ldots,\varphi_n)=V_gT(V_{g^{-1}}\varphi_1,\ldots,V_{g^{-1}}\varphi_n)$, $g\in G, \varphi_i\in H, T\in\mathscr{L}_n(H)$. On the space $\mathscr{F}_n(H)$ one defines a
linear representation, denoted also by $V^{(n)}$, by $V_g^n(T)(\varphi_1,\ldots,\varphi_n)=T(V_{g^{-1}}\varphi_1,\ldots,V_{g^{-1}}\varphi_n), g\in G, \varphi_i\in H$, $T\in\mathscr{F}_n(H)$. The representation $V^{(n)}$ is in general not continuous when the topology on $\mathscr{L}_n(H)$ [resp. $\mathscr{F}_n(H)$] is defined by uniform convergence on bounded sets. One denotes by $\hat{\otimes}_\pi^n H$ the completed projective tensor product $H\hat{\otimes}_\pi\cdots\hat{\otimes}_\pi H$ (n times). One defines the symmetrization operator σ_n from $\hat{\otimes}_\pi^n H$ to itself by

$$\sigma_n(\varphi_1\otimes\cdots\otimes\varphi_n)=\frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n}\varphi_{\sigma(1)}\otimes\cdots\otimes\varphi_{\sigma(n)},$$

 0022-2488/83/030457-06$02.50

$\mathfrak{S}_n$ being the group of permutations of n elements. We identify as usual $\mathscr{L}_n(H_1,H_2)$ to a subspace of $\mathscr{L}_1(\hat{\otimes}^n_\pi H_1,H_2)$.

Denote by $(L,\mathscr{H})$ one of the previous linear representations. The spaces $Z^1(G,\mathscr{H})$ of 1-cocycles, $B^1(G,\mathscr{H})$ of 1-coboundaries, and $H^1(G,\mathscr{H})$ of classes of 1-cohomology with coefficients in $\mathscr{H}$ have been defined in Refs. 5 and 6. We recall that it is equivalent to consider continuous or differentiable cohomology.

Differentiating on G at the identity, one gets the spaces $Z^1(\mathscr{G},\mathscr{H}^\infty)$, $B^1(\mathscr{G},\mathscr{H}^\infty)$, and $H^1(\mathscr{G},\mathscr{H}^\infty)$. The inclusion $H^1(G,\mathscr{H})\subset H^1(\mathscr{G},\mathscr{H}^\infty)$ becomes an equality when G is simply connected or when G is semisimple and $(L,\mathscr{H})$ is irreducible.[5,6]

III. CONSTRUCTION OF SERIES OF ANALYTIC REPRESENTATIONS

In the following (V,H) is a continuous linear representation of G in a reflexive Banach space H. We shall build cocycles in $Z^1(G,\ \mathscr{L}_n(H))$ from 1-cocycles with coefficients in simpler representations.

3.1. Guillemin–Sternberg injection: Define a G morphism τ_n from the space $\mathscr{F}_n(H)$ to $\mathscr{L}_{n+1}(H)$ by

$$\tau_n(T)=(T\otimes \mathrm{Id}_H)\circ\sigma_{n+1}.$$

τ_n induces an application from $Z^1(G,\ \mathscr{F}_n(H))$ to $Z^1(G,\ \mathscr{L}_{n+1}(H))$, which we denote by the same letter.

3.2. Lemma: Suppose that (V, H) is either of finite dimension or equicontinuous. τ_n induces a one-to-one mapping from $H^1(G,\ \mathscr{F}_n(H))$ to $H^1(G,\ \mathscr{L}_{n+1}(H))$.

Proof: (a) Suppose that $\dim H<+\infty$. Every element of $\mathscr{L}_{n+1}(H)$ can be written as a linear combination of elements of the type $(T\otimes T')\circ\sigma_{n+1}$, $T\in\mathscr{F}_n(H)$, $T'\in\mathscr{L}_1(\mathrm{H})$. We define a G-morphism $\mathscr{O}_n:\mathscr{L}_{n+1}(H)\to\mathscr{F}_n(H)$ by $\mathscr{O}_n((T\otimes T')\circ\sigma_n)=\mathrm{Tr}(T')\cdot T$. Since $\mathscr{O}_n\circ\tau_n$ $=\{[(\dim H)+n]/(n+1)\}\mathrm{Id}/\mathscr{F}_n(H)$, τ_n is one-to-one.

(b) Suppose now that (V,H) is equicontinuous and $\dim H=+\infty$. Given $T\in\mathscr{F}_n(H)$ and $\varphi_1,\dots,\varphi_{n+1}\in H$, we have

$$\begin{aligned}\tau_n(T)(\varphi_1,\dots,\varphi_{n+1})&=[1/(n+1)](T(\varphi_1,\dots,\varphi_n)\varphi_{n+1}\\&\quad+T(\varphi_{n+1},\dots,\varphi_{n-1})\varphi_n+\cdots\\&\quad+T(\varphi_2,\dots,\varphi_n,\varphi_{n+1})\varphi_1),\varphi_1,\dots,\varphi_{n+1}\in H.\end{aligned}$$

Therefore, $\|\tau_n(T)\|\leqslant\|T\|$. Now, given $\varphi_1,\dots,\varphi_n$, there exists $\varphi_{n+1}\neq0$ such that $T(\varphi_{n+1},\varphi_1,\dots,\varphi_{n-1})$ $=\cdots=T(\varphi_2,\dots,\varphi_n,\varphi_{n+1})=0$. We then have $[1/(n+1)]\|T(\varphi_1,\dots,\varphi_n)\|\leqslant\|\tau_n(T)\|\ \|\varphi_1\|\cdots\|\varphi_n\|$, which means that $\|T\|\leqslant(n+1)\|\tau_n(T)\|$. We now choose $\xi\in Z^1(G,\mathscr{F}_n(H))$ such that $\tau_n(\xi)\in B^1(G,\mathscr{L}_{n+1}(H))$. We can write, for any $g\in G$, $\tau_n(\xi)_g=V_g^{(n+1)}(T)-T$ for some $T\in\mathscr{L}_{n+1}(H)$. But, since there exists $K>0$ such that $\|V_g\varphi\|\leqslant K\|\varphi\|$, $g\in G$, $\varphi\in H$, we have $\|\tau_n(\xi)_g\|=\|\tau_n(\xi_g)\|$ $\leqslant(K^{n+1}+1)\|T\|$. Consequently, $\|\xi_g\|\leqslant(n+1)(K^{n+1}+1)$ $\|T\|$. The cocycle ξ being uniformly bounded, it results from a theorem of Johnson (Ref. 7, Theorem 3.4) that it is a coboundary.

3.3. Remark: Lemma 3.2 for $\dim H<+\infty$ and $n=1$ has been proved in Ref. 8.

3.4. Proposition:

3.4.1: Given $\xi\in Z^1(G,\mathscr{F}_p(H))$, formulae

$$S_g^{(p,\xi)}\varphi=[1-\xi_g(V_g\varphi,\dots,V_g\varphi)]^{-1/p}V_g\varphi,\quad g\in G,$$

and φ in the open ball of radius $\|\xi_g\|^{-1}$ in H, define analytic representations of G.

3.4.2: $S^{(p,\xi)}$ is formally linearizable if and only if $\xi\in B^1(G,\mathscr{F}_p(H))$, and in which case it is also analytically linearizable.

3.4.3: If $S^{(p,\xi)}$ and $S^{(p',\xi')}$ are not linearizable, they are formally equivalent if and only if $p=p'$ and if, moreover, there exists a G-morphism u of (V,H) such that $\xi^u-\xi'\in B^1(G,\mathscr{F}_p(H))[\xi^u_g(\varphi_1,\dots,\varphi_n)$ $\equiv\xi_g(u^{-1}\varphi_1,\dots,u^{-1}\varphi_n)]$, in which case they are also analytically equivalent.

Proof: 3.4.1 is straightforward. Let us pass to 3.4.2. Suppose that $S=S^{(p,\xi)}$ is formally linearizable. There exists $A=\Sigma_{i\geqslant1}A_i$, $A_i\in\mathscr{L}_i(H)$, A_1 invertible, such that $AV_g=S_gA$. Writing this equality degree by degree, one gets first that A_1 is a G-morphism of (V,H), and then that

$$\begin{aligned}\tau_n(\xi_g)(\varphi,\dots,\varphi)&=\xi_g(\varphi,\dots,\varphi)\varphi\\&=A_{p+1}(A_1^{-1}\varphi,\dots,A_1^{-1}\varphi)\\&\quad-V_gA_{p+1}(A_1^{-1}V_g^{-1}\varphi,\dots,A_1^{-1}V_g^{-1}\varphi).\end{aligned}$$

This proves that $\tau_p(\xi)\in B^1(G,\mathscr{L}_{p+1}(H))$. It results from 3.2 that $\xi\in B^1(G,\mathscr{F}_p(H))$. Let us write $\xi_g=V_g^{(p)}T-T$, where $T\in\mathscr{F}_p(H)$.

We introduce the following analytic mapping

$$\hat{T}(\varphi)=[1+T(\varphi,\dots,\varphi)]^{-1/p}\cdot\varphi.$$

We have

$$\begin{aligned}S_g\hat{T}\varphi&=[1+T(\varphi,\dots,\varphi)-(\xi_g)(V_g\varphi,\dots,V_g\varphi)]^{-1/p}\cdot V_g\varphi\\&=[1+T(V_g\varphi,\dots,V_g\varphi)]^{-1/p}V_g\varphi=\hat{T}V_g\varphi,\end{aligned}$$

which proves that (S,H) is analytically linearizable if and only if $\xi\in B^1(G,\mathscr{F}_p(H))$.

Passing to 3.4.3, suppose that $S=S^{(p,\xi)}$ and $S'=S^{(p',\xi')}$ are formally equivalent but not linearizable. There exists a formal power series $A=\Sigma_{n\geqslant1}A_n$, $A_n\in\mathscr{L}_n(H)$ such that $AS_g=S'_gA$, $g\in G$. Identification degree by degree shows that A_1 is a G-morphism of (V,H), then that $p=p'$ (because neither ξ nor ξ' is a coboundary), and then that $\tau_n(\xi^{A_1}-\xi')$ $\in B^1(G,\mathscr{L}_{p+1}(H))$. It results from 3.2 that ξ^{A_1} $-\xi'\in B^1(G,\mathscr{F}_p(H))$. One then remarks that:

(a) $A_1S_g^{(p,\xi)}=S_g^{(p,\xi^{A_1})}A_1$.

(b) Given $T\in\mathscr{F}_p(H)$, the analytic mapping $\hat{T}(\varphi)=[1+T(\varphi,\dots,)]^{-1/p}\cdot\varphi$ satisfies $\hat{T}S_g^{(p,\xi)}=S_g^{(p,\xi')}\hat{T}$, with $\xi'_g=\xi_g+V_g^{(p)}T-T$. Combining these two points, one gets the result.

Q.E.D.

3.5. Examples:

3.5.1: Take $G=\mathbf{R}$ and $H=L^2(\mathbf{R})$ on which acts the regular representation $(V_tf)(x)=f(x+t)$. Given $\xi\in H^\infty$, we define $\tilde{\xi}_t(f)=\int_0^t(\xi*\check{f})(s)ds$, where $\check{f}(x)=f(-x)$. $\tilde{\xi}$ belongs to $Z^1_\infty(G,\ H^*)$ and every element of $Z^1_\infty(G,\ H^*)$ can

be written in this way. Define now an analytic representation of G in H by

$$S_i^{\xi}(f) = [1 - \xi_i(V_i f)]^{-1} V_i(f).$$

It results from 3.4.2 that (S^{ξ}, H) is linearizable if and only if $\xi = dF/dx$ with $F \in H^{\infty}$, in which case the linearization operator is

$$T(f) = \left[1 + \int_{\mathbf{R}} F(s) f(s)\, ds\right]^{-1} f.$$

3.5.2: G is semisimple, of real rank 1, and has a finite center. Then, there exists[9] a unitary irreducible representation (V, H^*) such that $H^1(G,H) \neq \{0\}$ [for instance, when $G = \mathrm{SL}(2,\mathbf{R})$, the representations $D_i^{\pm}$ when $G = \mathrm{SL}(2,\mathbf{C})$ the representation $D_{-2,0}$.[10] In all of these cases we can build analytic representations which are not linearizable. We shall come back in Sec. IV to the case of $\mathrm{SL}(2,\mathbf{R})$.

3.6. Construction of the mapping η_p. Now, we suppose given on the reflexive Banach space H a bilinear symmetric form, denoted $\langle \cdot, \cdot \rangle$, which is continuous and invariant (i.e., $\langle V_g\varphi, V_g\psi, \rangle = \langle \varphi, \psi \rangle$, $\varphi, \psi \in H$). A representation (V,H) of G in H is said *pseudo-orthogonal* if there exists a real subvector space H_0 in H such that $H_0^{\mathbf{C}} = H$ and such that the restriction of the bilinear form to H is positive definite. This last property is always satisfied when $\dim H < +\infty$. We define a G-morphism $\eta_p : H^* \times H \to \mathscr{L}_{2p+2}(H)$ by

$$\eta_p(f,\psi)(\varphi,\ldots,\varphi) = \langle \varphi,\varphi \rangle^p f(\varphi)\cdot\varphi + \langle \varphi,\varphi \rangle^{p+1}\psi, \quad f \in H^*, \varphi, \psi \in H.$$

3.7. Lemma: Given a pseudo-orthogonal representation (V,H) of G, η_p induces a one-to-one mapping from $H^1(G, H^* \times H)$ to $H^1(G, \mathscr{L}_{2p+2}(H))$ when $2 < \dim H < +\infty$ *or when* $\dim H = +\infty$ *and* (V, H) *is* equicontinuous.

Proof: Suppose that $2 < \dim H < +\infty$. It is easily seen that η_p is an $\mathrm{SO}(n)$-morphism ($n = \dim H$). If $n = 2$, $\eta_p(H^* \times H)$ is clearly isomorphic to $H^* \times H$, as a representation of $\mathrm{SO}(2)$. If $n > 3$, $\eta_p(H^* \times H)$ is (*a priori*) isomorphic to $\{0\}, H, H^*$, or $H^* \times H$. But since $\eta_p(H^* \times H)$ contains two vectors of dominant weight for $\mathrm{so}(n)$, which are linearly independent, one has $\eta_p(H^* \times H) \simeq H^* \times H$. This implies that η_p is one-to-one. Moreover, for any value of n, the representation of $\mathrm{SO}(n)$ in $\mathscr{L}_{2p+2}(H)$ is completely reducible. Therefore, $\eta_p(H^* \times H)$ has an $\mathrm{SO}(n)$-invariant (and indeed G-invariant) direct factor in $\mathscr{L}_{2p+2}(H)$. This way, η_p induces a one-to-one mapping from $H^1(G, H^* \times H)$ into $H^1(G, \mathscr{L}_{2p+2}(H))$. Suppose now that $\dim H = +\infty$ and that (V,H) is equicontinuous. One can replace the initial norm on H by the equivalent Hilbert norm defined by extension of the scalar product from H_0 to H. Choose now $\xi^1 \in Z^1(G, H^*)$ and $\xi^2 \in Z^1(G, H)$, and suppose that $\eta_p(\xi^1, \xi^2) \in B^1(G, \mathscr{L}_{2p+2}(H))$. Since V is equicontinuous, there exists $M > 0$ such that $\|\eta_p(\xi^1_g, \xi^2_g)\| < M$ for any $g \in G$.

Now, given $g \in G$, there exists $\varphi \in H$ such that $\|\varphi\| = 1$ and $\xi^1_g(\varphi) = 0$. Going back to the definition of η_p, one gets $\|\xi^2_g\| < M$. Using Ref. 7, Theorem 3.4, we get that $\xi^2 \in B^1(G,H)$. Adding to $\eta_p(\xi^1, \xi^2)$ a coboundary, we can suppose that $\xi^2 = 0$ without changing ξ^1 and $\|\eta_p(\xi^1_g, 0)\| < M'$ for any $g \in G$, for some $M' > 0$. We then have $\|\xi^1_g(\varphi)\| < M'\|\varphi\|$, $\varphi \in H_0$. Any $\varphi \in H$ writes $\varphi = \varphi_1 + i\varphi_2$, $\varphi_1, \varphi_2 \in H_0$ and $\|\varphi\|^2 = \|\varphi_1\|^2 + \|\varphi_2\|^2$. Consequently, $\|\xi^1_g \varphi\| < 2M'\|\varphi\|$, for any $\varphi \in H$. Using again Ref. 7, Theorem 3.4, ξ^1 is a coboundary.

Q.E.D.

3.8. Remark: Consider the linear representation (V, H) of the Euclidean group $G = \mathrm{SO}(n)\cdot\mathbf{P}^n$ $(n > 3)$ on $H = \mathbf{C}^n$ such that the action of $\mathrm{SO}(n)$ is the natural one and the action of $\mathbf{R}^n$ is trivial. One has

$$H^1(G, \mathscr{L}_{2p+2}(H)) = \eta_p(H^1(G, H^* \times H)) \quad \text{(see Ref. 11)}.$$

3.9. Proposition:

3.9.1: Given $\xi \in Z^1(G,H)$, formulae

$$T_g^{(p,\xi)}(\varphi) = (1 - \langle \xi_g, V_g\varphi \rangle)(\langle \varphi,\varphi \rangle^{p-1})^{-1/(2p-1)} V_g\varphi$$

and

$$U_g^{(p,\xi)} = (1 - 2\langle \varphi,\varphi \rangle^{p-1}\langle \xi_g, V_g\varphi \rangle + \langle \xi_g\xi_g \rangle \langle \varphi,\varphi \rangle^{2p-1})^{-p/(2p-1)} \times (V_g\varphi - \langle \varphi,\varphi \rangle^p \xi_g),$$

$g \in G$ and φ small enough in H, define analytic representations of G.

3.9.2: Suppose that $2 < \dim H < +\infty$, or that $\dim H = +\infty$ and (V, H) is equicontinuous, $T^{(p,\xi)}$ (resp. $U^{(p,\xi)}$) is formally linearizable if and only if $\xi \in B^1(G,H)$, in which case it is also analytically linearizable.

3.9.3: Suppose that (V,H) is Schur-irreducible and satisfies the conditions of 3.9.2:

(a) when $T^{(p,\xi)}$ and $U^{(p',\xi')}$ are not both linearizable, they are not formally equivalent.

(b) $T^{(p,\xi)}$ and $T^{(p',\xi')}$ (resp. $U^{(p,\xi)}$ and $U^{(p',\xi')}$) are formally equivalent if and only if $p = p'$ and $\xi - \lambda\xi' \in B^1(G,H)$ for some $\lambda \in \mathbf{C}$, in which case they are also analytically equivalent.

Proof: 3.9.1 is straightforward. Suppose that $T = T^{(p,\xi)}$ is formally linearizable. Write $\xi^*_g = \langle \xi_g, \cdot \rangle$. A look at the first nonlinear term in the series development of equality $A \circ V_g = T_g \circ A$ $[A = \Sigma_{n>1} A_n, A_n \mathscr{L}_n(H)]$ shows that $\eta_p(\xi^*, 0) \in B^1(G, \mathscr{L}_{2p+2}(H))$. 3.7 implies then that $\xi^* \in B^1(G, H^*)$. Write $\xi_g V_g f - f, f \in H$. We then define the analytic mapping $A(\varphi) = (1 - \langle f,\varphi \rangle \langle \varphi,\varphi \rangle^{p-1})^{-1/(2p-1)}\cdot\varphi$. We have $V_g A = A T_g$. Suppose that $U = U^{(p,\xi)}$ is formally linearizable. As before, one sees that $\eta_p(\xi^*, \xi) \in B^1(G, \mathscr{L}_{2p+2}(H))$ and therefore $\xi \in B^1(G,H)$. Write $\xi_g = V_g f - f$, $f \in H$, and introduce the analytic mapping

$$A(\varphi) = (1 - 2\langle \varphi,\varphi \rangle^{p-1}\langle f,\varphi \rangle + \langle ff \rangle \langle \varphi,\varphi \rangle^{2p-1})^{-p/(2p-1)}(\varphi - \langle \varphi,\varphi \rangle^p f).$$

We verify that $V_g A = A U_g$. The same type of calculation proves 3.9.3. Q.E.D.

3.10. Examples:

3.10.1: Consider the linear representation (V,H) of 3.8; the authors proved in Ref. 11 that $T^{(p,\xi)}$ and $U^{(p,\xi)}$ classify all the analytic representations of G, the linear part of which is (V,H).

3.10.2: Take $G = SL(2,\mathbb{C})$ and (V,H) the unitary representation $D_{-2.0}$.[10] Conditions of 3.9 are satisfied, and $\dim H^1(G,H) = 1$. We can therefore build two series of nonequivalent analytic representations of $SL(2,\mathbb{C})$.

IV. PERTURBATION OF THE DISCRETE SERIES OF SU(1,1)

In the following, $G = SU(1,1)$. Given

$$g = \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix} \in G,$$

and z in the open unit disc in $\mathbb{C}$, we define $g\cdot z = (\bar{\alpha}z + \bar{\beta})/(\beta z + \alpha)$. Consider the representation $(V^{+,\rho}, \mathscr{A})$ of G on the Fréchet space $\mathscr{A}$ of holomorphic functions on the unit disc, defined by $(V_g^{+,\rho} f)(z) = \mu(g,z)^p \times f(g^{-1}\cdot z)$, where $\mu(g,z) = (\bar{\alpha} - \beta z)^{-2}$, $p \in \mathbb{N}^*$, and $f \in \mathscr{A}$. We denote by $H^{+,\rho}$ the Hilbert space of the functions in $\mathscr{A}$ such that

$$\|f\| = \{[(2p-1)/\pi] \times \int (1-|z|^2)^{2p-2} |f(z)|^2 \, dxdy\}^{1/2} < +\infty.$$

$H^{+,\rho}$ is invariant under $V^{+,\rho}$ and $(V^{+,\rho}, H^{+,\rho})$ is a unitary representation of G which belongs to the class D_p^+ (in Bargmann's classification). One defines in the same way the representation $(V^{-,\rho}, \mathscr{A})$ by

$$(V_g^{-,\rho} f)(z) = \mu(\bar{g},z) f(\bar{g}^{-1}, z) \quad \bar{g} = \begin{pmatrix} \bar{\alpha} & \bar{\beta} \\ \beta & \alpha \end{pmatrix}.$$

On the invariant Hilbert space $H^{-,\rho} \equiv H^{+,\rho}$, $V_g^{-,\rho}$ is unitary. The representation $(V^{-,\rho}, H^{-,\rho})$ belongs to the class D_p^-. We introduce on $\mathscr{G}$ a basis:

$$Y = \begin{pmatrix} -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 0 \\ -i & 0 \end{pmatrix}, \quad G = \begin{pmatrix} 0 & i \\ 0 & 0 \end{pmatrix}.$$

The Casimir element is $Q = GF + Y + Y^2$. It takes the value $p(p-1)$ in the class $D_p^{\pm}$. $k = i\mathbb{R}Y$ is a compact subalgebra of $\mathscr{G}_0$. The polynomials

$$\varphi_m = \binom{2p+m-1}{m}^{1/2} z^m \quad (m \geqslant 1)$$

form an orthonormal basis of $H^{+,\rho}$; the differential $dV^{+,\rho}$ of $V^{+,\rho}$ writes on this basis:

$$dV_Y^{+,\rho}\varphi_m = -(p+m)\varphi_m,$$
$$dV_F^{+,\rho}\varphi_m = i[m(2p+m-1)]^{1/2}\varphi_{m-1} \quad (\text{where } \varphi_{-1} = 0),$$
$$dV_G^{+,\rho}\varphi_m = i[(m+1)(2p+m)]^{1/2}\varphi_{m+1}.$$

It is known[9] that $H^1(G, H^{\pm,\rho}) = \{0\}$ if $p > 1$ and that $\dim H^1(G,H^{\pm,1}) = 1$. An example of a nontrivial cocycle in $Z^1(G,H^{+,1})$ [resp. $Z^1(G,H^{-,1})$] is $\xi_g(z) = \beta(\bar{\alpha} - \beta z)^{-1}$ $\left[\text{resp. } \xi_g^*(f) = \int_0^{g\cdot 0} f(z)\,dz, f \in H^{+,1}, g\cdot 0 = \alpha^{-1}\bar{\beta}\right]$.

4.1. Proposition: The formula

$$S_g(f) = \left[1 - \int_0^{g(0)} (V_g^{+,1} f)(z)\,dz\right]^{-1} V_g f, \quad g \in G, f \in H^{+,1},$$

defines an analytic representation of G, which is not linearizable, on $H^{+,1}$.

Proof: This is a reformulation of 3.4 in this particular situation. Q.E.D.

$(S,H^{+,1})$ extends to a formal representation $(S,\mathscr{A})$ defined by the same formula on $\mathscr{A}$. Consider the transformations $(Tf)(z) = (1 - \int_0^z f(\xi)\,d\xi)^{-1} f(z)$, defined for f small enough (in an obvious sense) in $\mathscr{A}$, and $(Ef)(z) = f(z)\exp\left[-\int_0^z f(\xi)\,d\xi\right]$. The transformations $T\circ E$ and $E\circ T$ are respectively the identity on $\mathscr{A}$, and the identity on the domain of T. We have $S_g f = E\cdot V_g^{+,1}\cdot T$ for small enough in $\mathscr{A}$ (i.e., S_g is linearizable in $\mathscr{A}$). Obviously T does not leave $H^{+,1}$ invariant. A similar treatment can be applied to the representation $(V^{-,1}, \mathscr{A})$.

Denote by $H_k^{+,\rho}$ the subspace of k-finite vectors in $H^{+,\rho}$, by E_t the eigenspace for the eigenvalue $-(p+t)$ of the operator $dV_Y^{+,\rho}$, by $\mathscr{H}^n$ the Hilbert tensor product of $H^{+,\rho}$ by itself n times, by $\mathscr{H}_k^n$ the algebraic tensor product of $H_k^{+,\rho}$ by itself n times, and by $\mathscr{P}_t = \oplus_{i_1+\cdots+i_n=t} E_{i_1} \otimes \cdots \otimes E_{i_n}$.

4.2. Lemma: The representation $(d(\otimes^n V^{+,\rho}), \mathscr{H}^n)$ reduces as $\oplus_{t\geqslant 0}\binom{n+t-2}{t} D_{np+t}^+$. $R = d(\otimes^n V^{+,\rho})_Q - p(p-1)$ is diagonal on $\mathscr{P}_t$; its eigenvalues $\lambda_{s_1,\ldots,s_n}$ $(s_1 + \ldots + s_n = t)$ on $\mathscr{P}_t$ satisfy:

$$\lambda_{s_1,\ldots,s_n} \geqslant np(np-1)$$

and

$$\sum_{s_1+\ldots+s_n=t} (\lambda_{s_1,\ldots,s_n} - p[(p-1)]^{-1} \leqslant \binom{n+t-1}{t}.$$

The operator R is invertible in $\mathscr{H}_k^n$ and has a bounded inverse in $\mathscr{H}^n$.

Proof: We have $d(\otimes^n V^{+,\rho})_Y \psi = -(np+t)\psi$ for any $\psi \in \mathscr{P}_t$, $d(\otimes^n V^{+,\rho})_F \mathscr{P}_t \subset \mathscr{P}_{t-1}$, and $d(\otimes^n V^{+,\rho})_G \mathscr{P}_t \subset \mathscr{P}_{t+1}$. Denote by K_t the subspace of $\mathscr{P}_t$ of the solutions of $d(\otimes^n V_F^{+,\rho})\psi = 0$. We have $\dim K_t = \binom{n+t-2}{t}$. Choose $\psi \neq 0$ in K_t. One sees that the vector space generated by the vectors $\{d(\otimes^n V^{+,\rho})_G^\alpha \psi, \alpha \geqslant 0\}$ is a representation of $\mathscr{G}$ in the class D_{np+t}^+. The subvector space generated by $\{d(\otimes^n V^{+,\rho})_G^\alpha K_t, \alpha \geqslant 0\}$ reduces into a direct sum of $\binom{n+t-2}{t}$ representations in the class D_{np+t}^+; this is what one denotes by $\binom{n+t-2}{t} D_{np+t}^+$.

Suppose that $\mathscr{P}^t = \oplus_{\alpha=0}^t d(\otimes^n V^{+,\rho})_G^\alpha K_{t-\alpha}$. The operator $d(\otimes^n V^{+,\rho})_Q$ is scalar on $d(\otimes^n V^{+,\rho})_G^{\alpha+1} K_{t-\alpha}$ with the eigenvalue $(np+t-\alpha)(np+t-\alpha-1)$, while on K_{t+1} it is scalar with the eigenvalue $(np+t+1)(np+t)$. As a consequence, $d(\otimes^n V^{+,\rho})_G \mathscr{P}_t \cap K_{t+1} = \{0\}$. Moreover, $d(\otimes^n V^{+,\rho})_G$ being one-to-one on $d(\otimes^n V^{+,\rho})_G^\alpha K_{t-\alpha}$, we have $\dim[d(\otimes^n V^{+,\rho})_G]\mathscr{P}_t = \dim \mathscr{P}_t = \binom{n+t-1}{t}$. Consequently,

$$\dim \mathscr{P}_{t+1} = \binom{n+t}{t+1} = \binom{n+t-1}{t+1} + \binom{n+t-1}{t}$$
$$= \dim K_{t+1} + \dim[d(\otimes^n V^{+,p})_G \mathscr{P}_t]$$

and $\mathscr{P}_{t+1} = \oplus_{\alpha=0}^{t+1} d(\otimes^n V^{+,p})_G^{\alpha} K_{t+1-\alpha}$. This proves that $\mathscr{H}_k^n = \oplus_{t\geq 0} \binom{n+t-2}{t} D^+_{np+t}$. On $\mathscr{P}_t$, $d(\otimes^n V^{+,p})_Q$ is diagonal with eigenvalues $(np+t-\alpha)(np+t-\alpha-1)$ of multiplicity

$$\binom{n+t-\alpha-2}{t-\alpha}, \alpha = 0,\ldots,t.$$

Consequently,

$$\sum_{s_1+\cdots+s_n=t} (\lambda_{s_1,\ldots,s_n} - p[p-1])^{-1}$$
$$= \sum_{\alpha=0}^{t} ((np+t-\alpha)(np+t-\alpha-1) - p[p-1])^{-1} \times \binom{n+t-\alpha-2}{t-\alpha}$$
$$\leq \sum_{u=0}^{t} \binom{n+u-2}{u} = \binom{n+t-1}{t} .$$

The operator R is self-adjoint in $\mathscr{H}^n$, diagonalizable, and its eigenvalues are minored by $(n-1)p(np+p-1) > 0$. It has therefore a bounded inverse. Q.E.D.

We denote by $\mathscr{A}_n$ the space of analytic mappings from the unit polydisc in $\mathbb{C}^n$ to $\mathbb{C}$. This is a Fréchet space with the seminorms $\|f\|_{r_1,\ldots,r_n} = \sup_{|z_i|<r_i} |f(z_1,\ldots,z_n)|$, $0 < r_i < 1$. Moreover, $\mathscr{A}_n \simeq \hat{\otimes}_\pi^n \mathscr{A}$ (see Ref. 12).

4.3. Lemma: The operator $R = d(\otimes^n V^{+,p})_Q - p(p-1)$ is continuous and has a continuous inverse on $\mathscr{A}_n$.

Proof: R is obviously continuous and, as we saw in 4.2, it has an inverse in $\mathscr{H}_k^n$. One has only to prove that this inverse is continuous when we induce on $\mathscr{H}_k^n$ the topology of $\mathscr{A}_n$.

Fix an orthonormal basis $P_{i_1,\ldots,i_n}, i_1+\cdots+i_n = t$, of eigenvectors of R in $\mathscr{P}_t$, $\lambda_{i_1,\ldots,i_n}$ being the eigenvalue on the vector $P_{i_1,\ldots,i_n}$. We write the change of basis

$$P_{i_1,\ldots,i_n} = \sum_{j_1+\cdots+j_n=t} A^{j_1,\ldots,j_n}_{i_1,\ldots,i_n} \varphi_{j_1} \otimes \cdots \otimes \varphi_{j_n}$$

and

$$\varphi_{j_1} \otimes \cdots \otimes \varphi_{j_n} = \sum_{i_1+\cdots+i_n=t} B^{i_1,\ldots,i_n}_{j_1,\ldots,j_n} P_{i_1,\ldots,i_n} .$$

We have $|A^{j_1,\ldots,j_n}_{i_1,\ldots,i_n}| \leq 1$ and $|B^{i_1,\ldots,i_n}_{j_1,\ldots,j_n}| \leq 1$.

Choose a function $f(z_1,\ldots,z_n) = \sum f_{i_1,\ldots,i_n} z_1^{i_1}\cdots z_n^{i_n}$ in $\mathscr{A}_n$ and introduce the function $K(r,r') = (1-r/r')^{-n}$ $(0<r<r'<1)$. We have

$$\sum_{i_1+\cdots+i_n=t} |f_{i_1,\ldots,i_n}| \leq K(r,r') r^{-t} \|f\|_{r',\ldots,r'} .$$

Given now a polynomial

$$h(z_1,\ldots,z_n) = \sum_{t\in T} \sum_{i_1+\cdots+i_n=t} h_{i_1,\ldots,i_n} z_1^{i_1}\cdots z_n^{i_n} \quad (T \text{ finite}),$$

we have

$$R^{-1}h = \sum_{t\in T} \sum_{i_1+\cdots+i_n=t} \left(\sum_{\substack{j_1+\cdots+j_n=t \\ s_1+\cdots+s_n=t}} \{\lambda_{s_1,\ldots,s_n} - p \times [p-1]\}^{-1} h_{j_1\cdots j_n} B^{s_1\cdots s_n}_{j_1\cdots j_n} A^{j_1\cdots j_n}_{s_1\cdots s_n} \right)$$

It results then from 4.2 that

$$|R^{-1}h(z_1,\ldots,z_n)| \leq \sum_{t\in T} \binom{n+t-1}{t} \sum_{\substack{i_1+\cdots+i_n=t \\ j_1+\cdots+j_n=t}} |\varphi_{i_1}(z_1)| \times \cdots |\varphi_{i_n}(z_n)| |h_{j_1,\ldots,j_n}|$$
$$\leq K(r,r') \|h\|_{r',\ldots,r'} \sum_{t\in T} r^{-t} \binom{n+t-1}{t} \times \sum_{i_1+\cdots+i_n=t} |\varphi_1(z_1)|\cdots|\varphi_n(z_n)| .$$

Consider now, in $\mathscr{A}_n$, the function

$$S(z_1,\ldots,z_n) = \sum_{t\geq 0} \sum_{i_1+\cdots+i_n=t} \varphi_{i_1}(z_1)\cdots\varphi_{i_n}(z_n)$$
$$= \prod_{j=1}^{n} \left[\sum_{i_j\geq 0} \binom{2p+i_j-1}{i_j}^{1/2} z_j^{i_j} \right].$$

We have

$$\frac{1}{(n-1)!} \frac{d^{n-1}}{dz^{n-1}} [z^{n-1} S(z,\ldots,z)]$$
$$= \sum_{t\geq 0} \binom{n+t-1}{t} z^t \sum_{i_1+\cdots+i_n=t} \left[\prod_{j=1}^{n} \binom{2p+i_j-1}{i_j}^{1/2} \right].$$

Suppose that $0 < \tau < r < r' < 1$. We have:

$$\|R^{-1}h\|_\tau \leq K(r,r') \|h\|_{r',\ldots,r'} \sum_{t\geq 0} \binom{n+t-1}{t} \times \left[\sum_{i_1+\cdots+i_n=t} \prod_{j=1}^{n} \binom{2p+i_j-1}{i_j}^{1/2} \right] \left(\frac{\tau}{r}\right)^t$$
$$\leq \frac{1}{(n-1)!} K(r,r') \|h\|_{r',\ldots,r'} \frac{d^{n-1}}{dz^{n-1}} \{z^{n-1} S(z,\ldots,z)\}_{\tau/r} .$$

The last inequality proves that R^{-1} is continuous from $\mathscr{H}_k^n$ to itself for the topology induced by that of $\mathscr{A}_n$. Q.E.D.

4.4. Proposition: A formal representation $(S, \mathscr{A})$ of G in $\mathscr{A}$, the linear part of which is in the class D_p^+ $(p \geq 1)$, is formally linearizable.

Proof: It is sufficient to prove that $H^1(G, \mathscr{L}_n(\mathscr{A})) = \{0\}$, since then there is no obstruction to formal linearization. Given $\xi \in Z^1(\mathscr{G}, \mathscr{L}_1(\hat{\otimes}_\pi^n \mathscr{A}, \mathscr{A}))$, where $n \geq 2$, we extend it to $\mathscr{U}(\mathscr{G})$ in such a way that $\xi_{uv} = dS^1_u \xi_v + \xi_u d(\hat{\otimes}_\pi^n S^1)_v$ for any $u, v \in \mathscr{U}(\mathscr{G})$ (see Ref. 5). Since $\xi_{xQ} = \xi_{Qx}$, we get $\xi_x = dS^1_x \xi_Q R^{-1} - \xi_Q R^{-1} d(\hat{\otimes}_\pi^n S_1)_x$. Therefore, $H^1(\mathscr{G}, \mathscr{L}_1(\hat{\otimes}_\pi^n \mathscr{A}, \mathscr{A})) = \{0\}$, and then $H^1(G, \mathscr{L}_n(\mathscr{A})) = H^1(G, \mathscr{L}(\hat{\otimes}_\pi^n \mathscr{A}, \mathscr{A})) = \{0\}$. Q.E.D.

We denote by $\mathscr{S}_n$ the Fréchet space of analytic functions in $\mathscr{A}_n$ of the form

$$f(z_1,\ldots,z_n) = \sum_{i_1,\ldots,i_n>0} f_{i_1,\ldots,i_n} z_1^{i_1}\cdots z_n^{i_n}$$

such that

$$|f|_j = \sum_{i_1,\ldots,i_n>0} (1+i_1+\cdots+i_n)^j |f_{i_1,\ldots,i_n}| < +\infty .$$

It is known[12] that $\mathscr{S}_n$ is isomorphic to $\hat{\otimes}_\pi^n \mathscr{S}_1$. Moreover, $\mathscr{S}_1$ is the space of C^∞-vectors of $(V^{+,p}, H^{+,p})$.

4.5. Lemma: $\mathscr{S}_n$ is the space of C^∞ vectors of the representation $(d(\otimes^n V^{+,p}), \mathscr{H}^n)$.

Proof: The action of G on $\mathscr{S}_n = \hat{\otimes}_\pi^n \mathscr{S}_1$ is C^∞, and we have the topological inclusion $\mathscr{S}_n \subset (\mathscr{H}^n)^\infty$. In addition, $\mathscr{S}_n$ is the space of C^∞ vectors of the operator $d(\otimes^n V^{+,p})_Y$. Consequently, $(\mathscr{H}^n)^\infty$ is an algebraic subvector space of $\mathscr{S}_n$; hence the algebraic equality $\mathscr{S}_n = (\mathscr{H}^n)^\infty$. Now, the Fréchet space topologies on $\mathscr{S}_n$ and $(\mathscr{H}^n)^\infty$ coincide since they are comparable.

Q.E.D.

4.6. Lemma: $R = d(\otimes^n V^{+,p})_Q - p(p-1)$ is continuous and has a continuous inverse on $\mathscr{S}_n$.

Proof: Obviously, R belongs to $\mathscr{L}_1(\mathscr{S}_n) = \mathscr{L}_1((\mathscr{H}^n)^\infty)$. Since R is a self-adjoint operator in $\mathscr{H}^n$, which is bigger than $(n-1)p(np+p-1) > 0$, in $\mathscr{L}_1(\mathscr{H}^n)$ we have $\|R^{-1}\| < [(n-1)p(np+p-1)]^{-1}$. Choose $h \in \mathscr{H}_k^n$ and $u \in \mathscr{U}(\mathscr{G})$; we have

$$\|d(\otimes^n V^{+,p})_u R^{-1} h\| = \|R^{-1} d(\otimes^n V^{+,p})_u h\|$$
$$< ((n-1)p(np+p-1))^{-1} \|d(\otimes^n V^{+,p})_u h\| .$$

Therefore, R^{-1} is continuous from $\mathscr{H}_k^n$ to itself for the topology induced by that of $\mathscr{S}_n$. This implies that R^{-1} is defined and continuous on $\mathscr{S}_n$.

Q.E.D.

4.7. Proposition: A formal representation $(S, \mathscr{S}_1)$ of G in $\mathscr{S}_1$, the linear part of which is in the class D_p^+, is formally linearizable.

Proof: Utilizing 4.6, it is proved the same way as 4.4.

Q.E.D.

4.8. Remark: It is clear that Propositions 4.4 and 4.7 remain true when we replace D_p^+ by D_p^-.

ACKNOWLEDGMENTS

We wish to thank Professor M. Flato and Professor J. P. Labesse for stimulating discussions.

[1] M. Flato, G. Pinczon, and J. Simon, "Nonlinear representations of Lie groups," Ann. Sci. Ec. Norm. Sup. **10**, 405–418 (1977).
[2] M. Flato and J. Simon, "Nonlinear equations and covariance," Lett. Math. Phys. **2**, 155–60 (1977).
[3] M. Flato and J. Simon, "Linearization of relativistic nonlinear wave equations," J. Math. Phys. **21**, 913–7 (1980).
[4] M. Flato and J. Simon, "On a linearization program of nonlinear field equations," Phys. Lett. **94**, 518–22 (1980).
[5] G. Pinczon and J. Simon, "On the 1-cohomology of Lie groups," Lett. Math. Phys. **1**, 83–91 (1975).
[6] G. Pinczon and J. Simon, "Extension of representations and cohomology," Rep. Math. Phys. **16**, 49–77 (1979).
[7] B. E. Johnson, "Cohomology in Banach algebras," Mem. Am. Math. Soc. **127** (1972).
[8] V. Guillemin and S. Sternberg, "Remarks on a paper of Hermann," Trans. Am. Math. Soc. **130**, 110–6 (1968).
[9] A. Guichardet, *Cohomologie des groupes topologiques et des algèbres de Lie* (Fernand Nathan, Paris, 1980).
[10] M. A. Naimark, *Linear Representations of the Lorentz Group* (Macmillan, New York, 1964).
[11] G. Pinczon and J. Simon, "Nonlinear representations of inhomogeneous groups," Lett. Math. Phys. **2**, 499–504 (1978).
[12] F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967).

Euler angle parametrization of the complex rotation group and its subgroups. I

Ansaruddin Syed
Department of Mathematics, University of Karachi, Karachi, Pakistan

(Received 22 October 1981; accepted for publication 15 January 1982)

A parametrization of the n-dimensional complex rotation group $O^+(n, \mathbb{C})$ by a set of $\frac{1}{2}n(n-1)$ complex Euler angles is obtained in exactly the same way as that for the n-dimensional real rotation group $O^+(n, \mathbb{R})$. Certain subgroups $\hat{L}(n, r)$ of $O^+(n, \mathbb{C})$ are then considered and found isomorphic with the "generalized Lorentz groups" $L(n, r)$, of which the special case $L(4, 1)$ is the usual Lorentz group L of special relativity. Distinguishing features, in the form of reality nature, of Euler angles of $\hat{L}(n, r)$ are obtained for $r = 1, 2$, and it is proved that these lead naturally to the definition of "real" Euler angles of the corresponding $L(n, r)$, which, of course, include the Lorentz group L.

PACS numbers: 02.20.Qs

I. INTRODUCTION

Euler angle parametrization of rotation and some related groups has attracted considerable attention in the recent past. Originally, of course, it was defined, in connection with rigid dynamics, for the three-dimensional real rotation group $O^+(3, \mathbb{R})$. With the advent of extensive use of some of the linear groups in quantum and relativity physics, the concept was generalized to the n-dimensional real rotation group by the author,[1] Hoffman *et al.*[2,3] and Wolf[4] and to the n-dimensional (restricted) Lorentz group $L(n, 1)$ by Wolf.[4] In a series of papers, Wolf,[4–6] Mackawa,[7–9] and Wong[10] used these generalizations for the explicit computation of matrix elements in irreducible representations of $O^+(0n, \mathbb{R})$ and $L(n, 1)$. Now both these groups are special cases of the "generalized Lorentz group" $L(n, r)$, $0 \leqslant r \leqslant n$, which is the collection of all linear transformations which keep

$$x_1^2 + x_2^2 + \cdots + x_{n-r}^2 - x_{n-r+1}^2 - \cdots - x_n^2$$

invariant. In addition to the above-mentioned cases $r = 0$ and $r = 1$, present day physics also uses some of the groups $L(n, r)$ with $r > 1$; thus one of the de Sitter groups in general relativity is $L(5, 2)$,[11] while twister theory[12] (also in general relativity) uses $L(6, 2)$. Of course, in the future, we may have to deal with groups corresponding to $r > 2$. It therefore appears very desirable to obtain Euler angle parametrization of $L(n, r)$ with arbitrary r. From the point of view of mathematics also, it is clearly far more satisfying to consider the general case of arbitrary r rather than that of some particular value of r, say $r = 0$ or $r = 1$. Actually this study of the general case proves to be quite rewarding because, as will be seen later, this case has, for $r > 1$, new and interesting features not possessed by $L(n, 1)$.

Keeping in view the above facts, we consider in this paper the slightly more general problem of defining "Euler angles" for the n-dimensional complex rotation group $O^+(n, \mathbb{C})$ and some of its subgroups; as $L(n, r)$ is found to be isomorphic to one of these subgroups, its "Euler angles" are naturally obtained. The essential idea behind the definition of Euler angles of $O^+(n, \mathbb{C})$ considered here is really quite simple; we just complexify the Euler angles of $O^+(n, \mathbb{R})$ defined in Refs. 1 and 3. It may be mentioned in this connection that a somewhat different set of "Euler angles" for $O^+(n, \mathbb{R})$ are given by Murnaghan[13]; these are; however, not very useful for our purpose as their complexification leads to parametrization of unitary groups with which the generalized Lorentz groups are not related in as simple a way as with the subgroups of $O^+(n, \mathbb{C})$ mentioned above.

II. THE COMPLEX ROTATION GROUP AND TRANSFORMATION OF COORDINATE SYSTEMS

As is well known, the complex orthogonal group $O(n, \mathbb{C})$ consists of all linear transformations in $\mathbb{C}^n$ which keep

$$z_1^2 + z_2^2 + \cdots + z_n^2 = \sum_{i=1}^{n} z_i^2$$

invariant. Here $z \equiv (z_1, z_2, \ldots, z_n)$ are the coordinates of a point $P \in \mathbb{C}^n$ relative to some Cartesian coordinate system I. If $A \in O(n, \mathbb{C})$, then

$$AA^T = \mathbf{1} \quad \text{the } n\text{-dimensional unit matrix,}$$

and

$$\det A = \pm 1.$$

We call an element $A \in O(n, \mathbb{C})$ with $\det A = 1$ a proper complex orthogonal transformation, or a complex rotation; the set $O^+(n, \mathbb{C})$ of these rotations is a subgroup of $O(n, \mathbb{C})$ and is called the complex rotation group. As the n^2 elements of $R \in O^+(n, \mathbb{C})$ satisfy the $\frac{1}{2}n(n+1)$ conditions

$$RR^T = \mathbf{1},$$

only $\frac{1}{2}n(n-1)$ of them will be independent, i.e., $O^+(n, \mathbb{C})$ is a $\frac{1}{2}n(n-1)$ complex parameter, and hence $n(n-1)$ real parameter, group.

It is easy to verify that a complex rotation R transforms a Cartesian coordinate system I in $\mathbb{C}^n$ into another such system I' with the same origin, and that the transformation from I to I' uniquely determines, and is determined by, the rotation. To parametrize the complex rotations, it is therefore sufficient to define $\frac{1}{2}n(n-1)$ complex Euler angles which completely and uniquely determine the transformation from I to I'. In order to do this, we start with a description of spherical polar coordinates in $\mathbb{C}^m$.

 0022-2488/83/030463-08$02.50

III. SPHERICAL POLAR COORDINATES IN $\mathbb{C}^m$

Let

$$z \equiv (z_1, z_2, ..., z_m)$$

be the coordinates of a general point $P \in \mathbb{C}^m$ referring to a Cartesian coordinate system I. Let

$$I(i), \quad i = 1, 2, ..., m,$$

be the m axes of I and

$$I(i_1, i_2, ..., i_r)$$

be the subspace of $\mathbb{C}^m$ spanned by

$$I(i_1), I(i_2), ..., I(i_r), \quad 1 \leqslant i_1, i_2, ..., i_r \leqslant m.$$

The coordinate system in $I(1, 2, ..., i)$ consisting of

$$I(1), I(2), ..., I(i)$$

will be denoted by I_i so that $I_m = I$. Similar notation for I', and, in fact, for any other coordinate system I^a which may be introduced later, is assumed. Note that

$$I(1, 2, ..., m) = I'(1, 2, ..., m) = \mathbb{C}^m.$$

Let $Z(i)$ be the tip of the unit vector along the axis $I'(i)$. We shall denote by

$$(z^a(i, 1), z^a(i, 2), ..., z^a(i, m)).$$

The coordinates of $Z(i)$ in the system I^a. Clearly, if

$$I'(i) \in I^a(1, 2, ..., j),$$

which is certainly the case if

$$I'(1, 2, ..., j) = I^a(1, 2, ..., j), \quad j \geqslant i,$$

then the coordinates of $Z(i)$ in I^a_j will be

$$(z^a(i, 1), z^a(i, 2), ..., z^a(i, j)).$$

Let us now introduce m (complex) variables $(t, x_m, x_{m-1}, ..., x_2)$ by

$$\begin{aligned}
&t = t_1 + it_2, \quad t_1 \geqslant 0,\\
&x_i = \theta_i + i\varphi_i, \quad i = 2, 3, ..., m,\\
&z_m = t \cos x_m, \quad 0 \leqslant \theta_m \leqslant \pi,\\
&z_{m-1} = t \sin x_m \sin x_{m-1}, \quad -\pi/2 \leqslant \theta_{m-1} \leqslant \pi/2,\\
&z_{m-2} = t \sin x_m \cos x_{m-1} \sin x_{m-2},\\
&\qquad -\pi/2 \leqslant \theta_{m-2} \leqslant \pi/2,\\
&\qquad \cdots\\
&z_3 = t \sin x_m \cos x_{m-1} \cdots \cos x_4 \sin x_3,\\
&\qquad -\pi/2 \leqslant \theta_3 \leqslant \pi/2,\\
&z_2 = t \sin x_m \cos x_{m-1} \cdots \cos x_3 \sin x_2, \quad 0 \leqslant \theta_2 \leqslant 2\pi,\\
&z_1 = t \sin x_m \cos x_{m-1} \cdots \cos x_3 \cos x_2.
\end{aligned} \tag{1}$$

If z is given, then

$$\begin{aligned}
&t^2 = z_1^2 + z_2^2 + \cdots + z_m^2\\
&\Rightarrow t = \pm (z_1^2 + z_2^2 + \cdots + z_m^2)^{1/2}.
\end{aligned}$$

The condition $t_1 \geqslant 0$ determines which sign is to be chosen so that t is uniquely determined by z.

Similarly, we have

$$\cos x_m = z_m / t.$$

As t has been uniquely determined, z_m/t is a known complex number so that, according to the Appendix, there exists a unique solution $x_m = \theta_m + i\varphi_m$ with $0 \leqslant \theta_m \leqslant \pi$, of the above equation. Thus x_m of (1) is uniquely determined by z.

Again as

$$\sin x_{m-1} = z_{m-1} / (t \sin x_m),$$

and t and $\sin x_m$ have already been determined, the rhs will be a known complex number so that, as before, there exists a unique solution

$$x_{m-1} = \theta_{m-1} + i\psi_{m-1}, \quad \text{with } -\pi/2 < \theta_{m-1} \leqslant \pi/2,$$

of the above equation. Hence, x_{m-1} of (1) is also uniquely determined by z.

We can similarly show that

$$x_{m-2}, x_{m-3}, ..., x_3, x_2$$

are all uniquely determined by z. Thus all the variables

$$(t, x_m, x_{m-1}, ..., x_2)$$

[with the restrictions mentioned in (1)] are uniquely determined by (1) for a given z, i.e., to each point $P \in \mathbb{C}$, there corresponds a unique set of values of

$$(t, x_m, x_{m-1}, ..., x_2),$$

which determine coordinates of P relative to I by Eq. (1). We therefore call

$$(t, x_m, x_{m-1}, ..., x_2)$$

the "spherical polar coordinates relative to I" of P. We shall also call

$$(x_m, x_{m-1}, ..., x_2)$$

the "set of (complex) polar angles (relative to I)" of P.

Note that

$$\begin{aligned}
\sin x_{m-j} &= z_{m-j} / t \sin x_m \cos x_{m-1} \cdots \cos x_{m-j+1}\\
&= \pm\, z_{m-j} / (z_m^2 + z_{m-1}^2 + \cdots + z_1^2)^{1/2}
\end{aligned} \tag{2a}$$

for $j = 1, 2, ..., m-2$, and

$$\cos x_m = z_m / t \equiv \pm \frac{z_m}{(z_m^2 + z_{m-1}^2 + \cdots + z_1^2)^{1/2}}. \tag{2b}$$

We now prove the following result which plays a crucial role in the definition of Euler angles.

Theorem: Let I be a Cartesian coordinate system in $\mathbb{C}^m$ and

$$(z_1, z_2, ..., z_m)$$

be the coordinates of a point $P \in \mathbb{C}^m$ relative to I. Also, let

$$(t, x_m, x_{m-1}, ..., x_2)$$

be the spherical polar coordinates relative to I of P. If I is given a sequence of rotations

$$\mathscr{J}_2, \mathscr{J}_3, ..., \mathscr{J}_m$$

(in this order), where

$$\begin{aligned}
&\mathscr{J}_i = \text{rotation by an angle } \chi_i \text{ in the } (1, i) \text{ plane}, \quad i = 2, 3,\\
&\qquad ..., m-1,\\
&\mathscr{J}_m = \text{rotation by an angle } -x_m \text{ in the } (1, m) \text{ plane},
\end{aligned}$$

and I' is the system so obtained, then $P \in I'(m)$.

First of all, we recall the following rule:

(A) If a system I' is obtained from I by rotation $\mathcal{J}$ represented by the matrix M and $\mathbf{t}, \mathbf{t}'$ are the column vectors representing the coordinates of the same point in I and I', respectively, then

$$\mathbf{t}' = M^T\mathbf{t},$$

where M^T is the transposed of M.

Next, notice that if M_i is the matrix representing the rotation $\mathcal{J}_i$, $2 \leqslant i \leqslant m$, then

$$M_i = \begin{bmatrix} \cos x_i & \cdots & -\sin x_i & \cdots & 0 \\ \vdots & & & & \\ \sin x_i & \cdots & \cos x_i & \cdots & 0 \\ \vdots & & & & \\ 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \rightarrow i\text{th row}$$

for $i = 2, 3, \ldots, m-1$, and

$$M_m = \begin{bmatrix} \cos x_m & \cdots & \sin x_m \\ \vdots & & \\ -\sin x_m & \cdots & \cos x_m \end{bmatrix}.$$

Here all the rows and columns which have not been explicitly written are assumed to be those of the unit $m \times m$ matrix.

Now, let $\mathbf{t}, \mathbf{t}'$ be the column vectors representing the coordinates of a point Y in I and I', respectively. Then

$$\mathbf{t}' = M_m^T M_{m-1}^T \cdots M_2^T \mathbf{t}$$
$$\Rightarrow \mathbf{t} = M_2 M_3 \cdots M_m \mathbf{t}'. \tag{3}$$

Take now

$$\mathbf{t}' = [0, 0, \ldots, t]^T$$

so that $Y \in I'(m)$. It is easy to check that

$$M_m \mathbf{t}' = t \begin{bmatrix} \sin x_m \\ 0 \\ \vdots \\ 0 \\ 0 \\ \cos x_m \end{bmatrix}$$

$$\Rightarrow M_{m-1} M_m \mathbf{t}' = t \begin{bmatrix} \sin x_m \cos x_{m-1} \\ 0 \\ \vdots \\ 0 \\ \sin x_m \sin x_{m-1} \\ \cos x_m \end{bmatrix}. \tag{4}$$

Now suppose that

$$M_i M_{i+1} \cdots M_m \mathbf{t}' = t \begin{bmatrix} \sin x_m \cos x_{m-1} \cdots \cos x_{i+2} \cos x_{i+1} \cos x_i \\ 0 \\ \vdots \\ 0 \\ \sin x_m \cos x_{m-1} \cdots \cos x_{i+2} \cos x_{i+1} \sin x_i \\ \sin x_m \cos x_{m-1} \cdots \cos x_{i+2} \sin x_{i+1} \\ \vdots \\ \sin x_m \sin x_{m-1} \\ \cos x_m \end{bmatrix} \rightarrow i\text{th element} \tag{5}$$

is true for $i = j, j \leqslant m-1$. Then

$$M_{j-1} M_j \cdots M_m \mathbf{t}' = M_{j-1}(M_j \cdots M_m \mathbf{t}'),$$

so that, by explicitly multiplying the vector

$$M_j M_{j+1} \cdots M_m \mathbf{t}'$$

obtained from (5) by putting $i = j$, by M_{j-1}, it can be checked that we get (5) with $i = j-1$. Hence, as (4) shows that (5) is true for $i = m-1$, it follows by induction that it will also be true for $i = 2$, i.e., we have

$$\mathbf{t} = M_2 M_3 \cdots M_m \mathbf{t}'$$

$$= t \begin{bmatrix} \sin x_m \cos x_{m-1} \cdots \cos x_3 \cos x_2 \\ \sin x_m \cos x_{m-1} \cdots \cos x_3 \sin x_2 \\ \vdots \\ \sin x_m \cos x_{m-1} \sin x_{m-2} \\ \sin x_m \sin x_{m-1} \\ \cos x_m \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{m-2} \\ z_{m-1} \\ z_m \end{bmatrix}.$$

This proves that Y is the same point as P so that $P \in I'(m)$, as required.

Note that the above theorem essentially says that if a coordinate system in $\mathbb{C}^m$ is given a sequence of rotations through the polar angles of a given line (in proper order), then the new system obtained has the mth axis along the given line.

IV. EULER ANGLES OF $O^+(n, \mathbb{C})$

Let R be the complex rotation which transforms the coordinate system I of $\mathbb{C}^n$ into I'. We are not in a position to define the Euler angles of I' relative to I or, equivalently, those of R. We start by constructing a sequence of coordinate systems

$$I_n^i, \quad i = n, n-1, \ldots, 1,$$

satisfying

$$I_n^i(k) = I'(k), \quad k = i+1, i+2, \ldots, n, \quad i \leqslant n-1, \tag{6}$$

and a sequence of set of "angles"

$$\{x_{ii}, x_{i\,i-1}, \ldots, x_{i2}\}, \quad i = n, n-1, \ldots, 2,$$

according to the following algorithm:

Algorithm: Let us suppose that the systems

$$I_n^i, \quad i \geqslant j,$$

satisfying (6) and the set of sequence of angles

$$\{x_{ii}, x_{i\,i-1}, \ldots, x_{i2}\}, \quad i > j,$$

have already been constructed for some $j > 2$. As mentioned earlier, the coordinates of $Z(j)$ in I_j^j are

$$(z^j(j,1), z^j(j,2), \ldots, z^j(j,j)).$$

Let the spherical polar coordinates relative to I_j^j of $Z(j)$ be

$$(1, x_{jj}, x_{j\,j-1}, \ldots, x_{j2}).$$

Apply to I_n^j, the ordered sequence of rotations

$$T_{j2}, T_{j3}, \ldots, T_{jj},$$

where

T_{jk} = a rotation by an angle x_{jk} in the $(1, k)$ plane, $k = 2, 3, \ldots, j-1$,

T_{jj} = a rotation by an angle $-x_{jj}$ in the $(1, j)$ plane, and

let I_n^{j-1} be the coordinate system so obtained.

As the applied rotations do not affect the k th axis, $j+1 \leqslant k \leqslant n$, we shall have

$$I_n^{j-1}(k) = I_n^j(k) = I'(k), \quad k = j+1, j+2, \ldots, n,$$

by (6) with $i = j$. Further, the theorem proved in the last section implies that

$$Z(j) \in I_j^{j-1} \equiv I_n^{j-1} \Rightarrow I_n^{j-1}(j) = I'(j).$$

Thus we have obtained the system I_n^{j-1} satisfying (6) with $i = j-1$ and the sequence of angles

$$\{x_{jj}, x_{j\,j-1}, \ldots, x_{j2}\}.$$

If we put $I_n^n = I_n \equiv I$, we can construct by the above procedure using the method of induction, all the coordinate systems

$$I_n^i, \quad i = n-1, n-2, \ldots, 1,$$

and all the $\frac{1}{2}n(n-1)$ (complex) angles

$$x_{ij}, \quad 2 \leqslant j \leqslant i \leqslant n.$$

We call these the Euler angles of I' relative to I. Note that (6) with $i = 1$ implies that

$$I_n^1(2) = I'(2), I_n^1(3) = I'(3), \ldots, I_n^1(n) = I'(n), \Rightarrow I_n^1 = I'.$$

Thus starting from I and giving it a sequence of successive rotations by the above Euler angles in a particular order, we reach the system I'. This verifies the fact that these angles completely and uniquely determine the transformation from I to I', and hence also the corresponding rotation R. Clearly

R = matrix of the resultant of the above succession of rotations,

so that it can be calculated in terms of the angles in a straightforward manner. The converse of this, i.e., the problem of finding the Euler angles of a given matrix $R \in O^+(n, \mathbb{C})$, has already been solved by the author[14] for real rotations. As the spherical polar coordinates in $\mathbb{C}^n$ used here are formally identical with those used in Ref. 14 for $\mathbb{R}^n$, it is easy to check that the arguments and procedure of 14 remain valid for complex rotations as well; we therefore have the following rule for finding the Euler angles of a given $R \in O^+(n, \mathbb{C})$:

Suppose first that

$$\{x_{jk}; k = j, j-1, \ldots, 2, \ j \geqslant i+1\}$$

have been obtained so that the matrices

$$M_{jk}, \quad k = j, j-1, \ldots, 2, \ j \geqslant i+1,$$

and hence also

$$N_j = M_{j2} M_{j3} \cdots M_{jj}, \quad j \geqslant i+1,$$

can be easily computed. Here $M_{jk} \equiv M_{jk}(x_{jk})$ is the matrix of rotation T_{jk} described in the Algorithm. It can be shown that

$$R^i = N_{i+1}^T N_{i+2}^T \cdots N_n^T R$$

has all elements in the last $(n-i)$ rows and columns equal to zero except the diagonal ones which are all 1. Thus if $\mathbf{t}^i$ is the ith column of R, we compute

$$N_{i+1}^T N_{i+2}^T \cdots N_n^T \mathbf{t}^i,$$

and find it to be of the form

$$[c_1, c_2, \ldots, c_i, \ 0, 0, \ldots, 0]^T,$$

where the c's are complex numbers.

$$\{x_{ii}, x_{i\,i-1}, \ldots, x_{i2}\}$$

are now obtained as the polar angles of

$$[c_1, c_2, \ldots, c_i]^T.$$

Let us now make the following remarks:

Remark 1: If $R^{(i)}$ is the matrix obtained from R^i by deleting its last $(n-i)$ rows and columns, then

$$x_{jk}, \quad 2 \leqslant k \leqslant j \leqslant i,$$

are the Euler angles of $R^{(i)}$.

Remark 2: We have

$$N_2^T N_3^T \cdots N_n^T R = \mathbf{1}, \quad \text{the } n \times n \text{ unit matrix},$$

so that

$$R = N_n N_{n-1} \cdots N_2. \tag{7}$$

V. EULER ANGLES OF SUBGROUPS OF $O^+(n, \mathbb{C})$

A. Subgroups of $O^+(n, \mathbb{C})$

We now obtain some subgroups of $O^+(n, \mathbb{C})$ by considering those of its elements which keep certain subspaces of $\mathbb{C}^n$ invariant. To illustrate the technique, we consider first the almost trivial subgroup $O^+(n, \mathbb{R})$. For this, we start with the subspace $\mathbb{R}^n$ of $\mathbb{C}^n$; clearly, it may be considered as the real vector space spanned by

$$\{e_1, e_2, \ldots, e_n\},$$

where e_i is the vector with all components zero except the ith, one which is 1. If

$$\alpha = \{\alpha_{ij}\}$$

is any complex $n \times n$ matrix, it will keep $\mathbb{R}^n$ ivariant if and only if it is real-valued, i.e., if and only if

$\alpha_{ij} \equiv (\alpha e_i)_j$ is real for all i, j.

Thus $\mathbf{R}^n$ is invariant under the set of all real-valued elements of $O^+(n, \mathbf{C})$, and this set is precisely the subgroup $O^+(n, \mathbf{R})$ of $O^+(n, \mathbf{C})$. Applications of the facts, previously obtained

(i) the polar angles of a real vector of $\mathbf{C}^n$, i.e., a vector of $\mathbf{R}^n$, are real,

(ii) a rotation by a real angle transforms a real vector into a real vector,

(iii) all the rotations carried out in the definition of Euler angles of a real rotation are by polar angles of real vectors, and hence by real angles, show that all the Euler angles of an element of $O^+(n, \mathbf{R})$ are real; in fact, they are identical with those given in Refs. 1–4.

Consider next the subspace $\hat{M}(n, r)$, $0 < r < n$, of $\mathbf{C}^n$, which is the real vector space spanned by

$$\{e_1, e_2, \ldots, e_r, ie_{r+1}, \ldots, ie_n\};$$

an element $\hat{x} \in \hat{M}(n, r)$ is clearly of the form

$$\hat{x} = [x_1, x_2, \ldots, x_r, ix_{r+1}, \ldots, ix_n]^T, \quad x_j \text{ real for all } j.$$

If $\hat{\alpha} \in O^+(n, \mathbf{C})$ keeps $\hat{M}(n, r)$ invariant, we must have

$$\hat{\alpha}\hat{x} \in \hat{M}(n, r)$$

for arbitrary $\hat{x} \in \hat{M}(n, r)$; it is easy to see that this will be so if and only if

$$\alpha_{jk} \text{ is real for } \begin{cases} \text{(i) } j \leqslant r \text{ and } k \leqslant r, \\ \text{(ii) } j \geqslant r+1 \text{ and } k \geqslant r+1 \end{cases}$$

and

$$\alpha_{jk} \text{ is pure imaginary for } \begin{cases} \text{(i) } j \leqslant r \text{ and } k \geqslant r+1, \\ \text{(ii) } j \geqslant r+1 \text{ and } k \leqslant r. \end{cases}$$

Thus the collection $\hat{L}(n, r)$ of all the elements of $O^+(n, \mathbf{C})$ which keep $\hat{M}(n, r)$ invariant consists of the set of all matrices $\hat{\alpha} \in O^+(n, \mathbf{C})$ given by

$$\begin{aligned} \hat{\alpha}_{jk} &= \alpha_{jk} && \text{for } \begin{matrix} \text{(i) } j \leqslant r, k \leqslant r, \\ \text{(ii) } k \geqslant r+1, k \geqslant r+1, \end{matrix} \\ \hat{\alpha}_{jk} &= i\alpha_{jkK} && \text{for } j \geqslant r+1, k \leqslant r, \\ \hat{\alpha}_{jk} &= -i\alpha_{jk} && \text{for } j \leqslant r, k \geqslant r+1, \\ & && \alpha_{jk} \text{ real for all } j, k. \end{aligned} \tag{8}$$

Using the fact that $\hat{\alpha}^{-1} = \hat{\alpha}^T$, it is easy to check that $\hat{L}(n, r)$ is a subgroup of $O^+(n, \mathbf{C})$. As in the case of $O^+(n, \mathbf{R})$, we should now obtain the reality conditions characterizing the Euler angles of $\hat{L}(n, r)$. However, these happen to be exceedingly involved and cumbersome, and so, to get some idea as to what might be happening in the general case, we consider here the two simplest special cases, $n = 1, 2$, which are found to possess all the essential features of the general case.

B. Euler angles of $\hat{L}(n, 1)$ and $\hat{L}(n, 2)$

First let $\hat{\alpha} \in L(n, 1)$ so that

$$x_{nn}, x_{nn-1}, \ldots, x_{n2},$$

are the polar angles of

$$[-i\alpha_{1n}, -i\alpha_{2n}, \ldots, -i\alpha_{n-1\,n}, \alpha_{nn}]^T.$$

(2) therefore shows that:

$$\cos x_{nn} = \alpha_{nn} > 1 \Rightarrow x_{nn} = i\theta_{nn}, \quad \theta_{nn} \text{ real};$$

$$\sin x_{n\,n-1} = \frac{\pm i\alpha_{n-1\,n}}{i(\alpha_{n-1\,n}^2 + \alpha_{n-2\,n}^2 + \cdots + \alpha_{1n}^2)^{1/2}}$$

$$= \text{real and} < 1 \text{ in magnitude}$$

$$\Rightarrow x_{n\,n-1} = \theta_{n\,n-1}, \text{ a real number.}$$

Similarly

$$x_{n\,n-2} = \theta_{n\,n-2}, \ldots, x_{n2} = \theta_{n2},$$

where all the θ's which appear here as well as in all that follows are assumed to be real.

Consider next the matrix

$$N_n^T = M_{nn}^T M_{n\,n-1}^T \cdots M_{n2}^T$$

for the above x_{ni}. It is easy to verify (by considering the action of one M_{ni}^T at a time) that when each one acts on a vector of the form

$$[\beta_1, \beta_2, \ldots, \beta_{n-1}, i\beta_n]^T, \quad \text{the } \beta\text{'s all real,}$$

it gives a vector of the same for. It follows that

$$\hat{\alpha}^{n-1} = N_n^T \hat{\alpha}$$

is a matrix of the same form as $\hat{\alpha}$ so that $\hat{\alpha}^{(n-1)}$, obtained from $\hat{\alpha}^{n-1}$ by deleting its last row and column, will be real and orthogonal and, hence, will have all of its Euler angles real. Remark 1 now tells us that the rest of Euler angles of $\hat{\alpha}$ are real. Thus for an element of $\hat{L}(n, 1)$, we have

$$\begin{aligned} x_{nn} &= i\theta_{nn}, \\ x_{ij} &= \theta_{ij}, \quad (i, j) \neq (n, n). \end{aligned} \tag{9}$$

Conversely, it is easy to verify that Euler angles satisfy conditions (9) determine an element of $\hat{L}(n, 1)$, i.e., Eqs. (9) actually characterize the subgroup $\hat{L}(n, 1)$ of $O^+(n, \mathbf{C})$.

Next, let $\hat{\alpha} \leftarrow \hat{L}(n, 2)$. As now

$$\alpha_{nn}^2 + \alpha_{n-1\,n}^2 - \alpha_{n-2\,n}^2 - \cdots - \alpha_{1n}^2 = 1,$$

we have two possibilities:

$$\alpha_{nn} > 1 \quad \text{or} \quad \alpha_{nn} < 1;$$

we consider them separately.

Case 1: $\alpha_{nn} > 1$: As

$$(x_{nn}, x_{n\,n-1}, \ldots, x_{n2})$$

are the polar angles of

$$[-i\alpha_{1n}, -i\alpha_{2n}, \ldots, -i\alpha_{n-2\,n}, \alpha_{n-1\,n}, \alpha_{nn}]^T,$$

(2) leads to

$$x_{nn} = i\theta_{nnK}, \quad x_{n\,n-1} = i\theta_{n\,n-1},$$
$$x_{nj} = \theta_{nj}, \quad 2 - n \geqslant j \geqslant 2.$$

Proceeding as in the case of $\hat{L}(n, 1)$, we can now show that

$$\hat{\alpha}^{n-1} = N_n^T \hat{\alpha} \equiv M_{nn}^T M_{n\,n-1}^T \cdots M_{n2}^T \hat{\alpha}$$

is of the same form as $\hat{\alpha}$, so that $\hat{\alpha}^{(n-1)} \in \hat{L}(n-1, 1)$. Using again Remark 1, we see that in this case

$$\left.\begin{aligned} &x_{nn} = i\theta_{nn}, \quad x_{n\,n-1} = i\,\theta_{n\,n-1}, \quad x_{n-1\,n-1} = i\,\theta_{n-1\,n-1}, \\ &x_{ij} = \theta_{ij} \quad \text{for all other permissible values of } (i, j) \end{aligned}\right\} \tag{10}$$

characterize the elements of $\hat{L}(n, 2)$.

Case 2: $\alpha_{nn} < 1$: Here, we first get, as in the above case,

$$x_{nn} = -\theta_{nn}, \quad x_{n\,n-1} = \pi/2 - i\theta_{n\,n-1},$$
$$x_{nj} = \theta_{nj}, \quad n-2 \geqslant j \geqslant 2.$$

It is now found that N_n^T transforms a vector of the form

$$[i\beta_1, i\beta_2, \ldots, i\beta_{n-2}, \beta_{n-1}, \beta_n]^T$$

to one of the form

$$[\beta_1, i\beta_2, \ldots, i\beta_{n-2}, i\beta_{n-1}, \beta_n]^T;$$

the last two columns of $\hat{\alpha}^{n-1} = N_n^T \hat{\alpha}$ will therefore be of the form, i.e., the last column of $\hat{\alpha}^{(n-1)}$ is of the form

$$[\beta_1, i\beta_2, \ldots, i\beta_{n-1}]^T.$$

This leads to

$$x_{n-1\,n-1} = \pi/2 - i\theta_{n-1\,n-1},$$
$$x_{n-1\,j} = i\theta_{n-1\,j}, \quad n-2 \geqslant j \geqslant 2,$$

from which it can be verified that

$$N_{n-1}^T = M_{n-1\,n-1}^T M_{n-1\,n-2}^T \cdots M_{n-1\,2}^T$$

transforms a vector of the form

$$[i\beta_1, \beta_2, \ldots, \beta_{n-2}, \beta_{n-1}, i\beta_n]^T$$

to one of the form

$$[\beta_1, \beta_2, \ldots, \beta_{n-1}, i\beta_{n-1}, i\beta_n]^T.$$

Thus the first $(n-2)$ columns of $\hat{\alpha}^{n-2}$ are of this form so that $\hat{\alpha}^{(n-2)}$ is real; Remark 1 therefore implies that

$$x_{ij} = \theta_{ij}, \quad 2 \leqslant j \leqslant i \leqslant n-2.$$

It follows that elements of $\hat{L}(n, 2)$ in this case are characterized by the reality conditions:

$$\left.\begin{aligned}
&x_{nn} = -\theta_{nn}, && x_{nn-1} = \pi/2 - i\theta_{nn-1},\\
&x_{nj} = \theta_{nj}, && 2 \leqslant j \leqslant n-2,\\
&x_{n-1\,n-1} = \pi/2 - i\theta_{n-1\,n-1},\\
&x_{n-1\,j} = i\theta_{n-1\,j}, && 2 \leqslant j \leqslant n-2,\\
&x_{ij} = \theta_{ij}, && 2 \leqslant j \leqslant i \leqslant n-2.
\end{aligned}\right\} \tag{11}$$

To sum up the above discussion, we see that the Euler angles of $O^+(n, \mathbb{R})$ are all real, those of $\hat{L}(n, 1)$ are either real or pure imaginary, and this is also true for one set of elements of $\hat{L}(n, 2)$. However, for the rest of elements of $\hat{L}(n, 2)$, these angles are either real or pure imaginary or Barred, meaning that they are of the type $\pi/2 - i\theta$. In Paper II of this series, we show that the Euler angles of $\hat{L}(n, r)$, r arbitrary, are also of the above three types only, and explicitly state (and prove) conditions determining the type of any particular one of them.

C. The generalized Lorentz group

Recall that an element $\hat{x} \in \hat{M}(n, r)$ is of the form

$$\hat{x} = [x_1, x_2, \ldots, x_r, ix_{r+1}, \ldots, ix_n]^T, \quad x_j \text{all real},$$

so that if

$$x = [x_1, x_2, \ldots, x_n]^T$$

then $x \in \mathbb{R}^n$ and

$$\hat{x} = fx,$$

where $f = \{f_{jk}\}$ is the $n \times n$ diagonal matrix

$$f_{jk} = \begin{cases} \delta_{jk}, & j \leqslant r, \\ i\delta_{jk}, & j \geqslant r+1. \end{cases}$$

Note that if $g = f^2$, then

(i) $f^{-1} = f^x$, $f^T = f$,

(ii) the mapping $f: \mathbb{R}^n \to \hat{M}(n, r)$ is an isomorphism of vector spaces,

(iii) $g_{jk} = \begin{cases} \delta_{jk}, & j \leqslant r, \\ -\delta_{jk}, & j \geqslant r+1, \end{cases}$

(iv) $f^4 = g^2 = \mathbf{1}$, the $n \times n$ unit matrix.

Next, let $\hat{\alpha} \in \hat{L}(n, r)$ and α be the real $n \times n$ matrix

$$\alpha = \{\alpha_{jk}\}.$$

Then $\hat{\alpha} = f\alpha f^{-1}$, and the orthogonality condition on $\hat{\alpha}$ gives

$$\hat{\alpha}^T \hat{\alpha} = \mathbf{1}$$
$$\Rightarrow f^{-1}\alpha^T f f \alpha f^{-1} = \mathbf{1}$$
$$\Rightarrow \alpha^T g \alpha = g \tag{12}$$
$$\Rightarrow (\alpha x)^T g (\alpha x) = x^T \alpha^T g \alpha x = x^T g x. \tag{13}$$

If we now define a "norm" $\mathcal{N}(x)$ on the vector space $\mathbb{R}^n$ by

$$\mathcal{N}(x) = x^T g x$$

and call the metric space $M(n, r)$ so obtained a generalized Minkowski space, then (13) says that $\hat{\alpha} \in \hat{L}(n, r)$ if and only if α is a linear transformation in $M(n, r)$ which keeps the lengths of the vectors invariant, i.e., it is an orthogonal transformation in $M(n, r)$. These orthogonal transformations are called generalized Lorentz transformations and their collection $L(n, r)$, which is easily checked to be a group, is called the generalized Lorentz group. Clearly, $L(n, r)$ and $\hat{L}(n, r)$ are simply a relabelling of one another as $\hat{\alpha} = f\alpha f^{-1}$ gives an isomorphism of $L(n, r)$ onto $\hat{L}(n, r)$.

In the end, we note that $M(4, 1)$ and $L(4, 1)$ are simply the ordinary (four-dimensional) Minkowski space and the ordinary Lorentz group, respectively, of special relativity.

D. Euler angles for $L(n, 1)$ and $L(n, 2)$

We now come to the problem of parametrizing the generalized Lorentz group, using, of course, its isomorphism with $\hat{L}(n, r)$. Thus

$$\alpha \in L(n, r) \Rightarrow \hat{\alpha} \in \hat{L}(n, r)$$
$$\Rightarrow \hat{\alpha} = M_{n2}(x_{n2}) M_{n3}(x_{n3}) \cdots M_{nn}(x_{nn})$$
$$\times M_{n-1\,2}(x_{n-1\,2}) \cdots M_{22}(x_{22})$$
$$\Rightarrow \alpha = f^{-1} M_{n2}(x_{n2}) M_{n3}(x_{n3}) \cdots M_{22}(x_{22}) f. \tag{14}$$

However, for a parametrization of any group of transformations to be really elegant and useful, it should express a general transformation as a product of simpler transformations of the same type. Hence, we try to convert the rhs of (14) into a product of (real) generalized Lorentz transformations. Using the reality nature of the Euler angles of $\hat{L}(n, 1)$ and $\hat{L}(n, 2)$ obtained earlier, we now show that this conversion is indeed possible for elements of $L(n, 1)$ and $L(n, 2)$.

First, Let $\alpha \in L(n, 1)$, so that (9) shows that (14) takes the form

$$\alpha = f^{-1}M_{n2}(\theta_{n2})\cdots M_{n\,n-1}(\theta_{n\,n-1})M_{nn}(i\theta_{nn}) \times M_{n1\,2}(\theta_{n-1\,2})\cdots M_{22}(\theta_{22})f$$

$$\Rightarrow \alpha = \bar{M}_{n2}(\theta_{n2})\cdots \bar{M}_{n\,n-1}(\theta_{n\,n-1})\bar{M}_{nn}(\theta_{nn}) \times \bar{M}_{n-1\,2}(\theta_{n-1\,2})\cdots \bar{M}_{22}(\theta_{22}), \tag{15}$$

where

$$\bar{M}_{nn}(\theta_{nn}) = f^{-1}M_{nn}(i\theta_{nn})f$$

$$\bar{M}_{ij}(\theta_{ij}) = f^{-1}M_{ij}(\theta_{ij})f, \quad (i,j) \neq (n,n). \tag{16}$$

It is easy to check that for $(i,j) \neq (n,n)$,

$\bar{M}_{ij}(\theta_{ij}) = M_{ij}(\theta_{ij})$ = a real rotation in $(1,j)$ plane, while [recall that $M_{ii}(x_{ii})$ is the matrix of rotation $-x_{ii}$ in the $(1,i)$ plane]

$$\bar{M}_{nn}(\theta_{nn}) = \begin{bmatrix} \cosh\theta_{nn} & \cdots & -\sinh\theta_{nn} \\ \vdots & & \vdots \\ -\sinh\theta_{nn} & \cdots & \cosh\theta_{nn} \end{bmatrix}$$

and therefore represents a pure Lorentz transformation along the 1-axis with velocity $\tanh\theta_{nn}$ and with x_n behaving as time; we call it a simple Lorentz transformation by an angle θ_{nn} in the $(1, n)$ plane. Thus (15) gives a genuine parametrization of $L(n, 1)$ by $\frac{1}{2}n(n-1)$ real "Euler angles"

$$\theta_{ij}, \quad 2 \leqslant j \leqslant i \leqslant n,$$

each having an associated matrix $\bar{M}_{ij}(\theta_{ij})$, which is either a real rotation, when the angle is said to be of first kind, or a simple Lorentz transformation, when it is said to be of second kind. Note that the definition of Euler angles in the context of the generalized Lorentz group is somewhat different from that in the context of the (real or complex) orthogonal group; this is because the associated matrices in the two contexts, $\bar{M}_{ij}$ and M_{ij}, are not always the same.

It may be mentioned here that the above parametrization of $L(n, 1)$ is the same as that given by Wolf and Anderson,[4] Wolf,[5] and Wong[10]; as in the case of these authors, we have $\frac{1}{2}n(n-1) - 1$ real rotations θ_{ij} $(i,j) \neq (n,n)$, and a single Lorentz transformation θ_{nn} which corresponds to their single "boost."

Next, let $\alpha \in L(n, 2)$. As $\hat{\alpha}$ now belongs to $\hat{L}(n, 2)$, we have two possibilities. If the Euler angles of $\hat{\alpha}$ are of the type (10), we get, as before,

$$\alpha = \bar{M}_{n2}(\theta_{n2})\cdots \bar{M}_{nn}(\theta_{nn})\bar{M}_{n-1\,2}(\theta_{n-1\,2})\cdots \bar{M}_{22}(\theta_{22}),$$

where

$$\bar{M}_{ij}(\theta_{ij}) = f^{-1}M_{ij}(i\theta_{ij})f = \text{a simple Lorentz transformation}$$

for $(i,j) = (n,n), (n,n-1), (n-1,n-1)$, and

$$\bar{M}_{ij}(\theta_{ij}) = f^{-1}M_{ij}(\theta_{ij})f = \text{a real rotation}$$

for all other (i,j).

Thus the Euler angles

$$\theta_{nn}, \quad \theta_{n\,n-1}, \quad \text{and} \quad \theta_{n-1\,n-1}$$

are of second kind while the rest of them are of first kind.

On the other hand, if the Euler angles of $\hat{\alpha}$ are of type (11),

$$\alpha = \bar{M}_{n2}(\theta_{n2})\cdots \bar{M}_{nn}(\theta_{nn})\bar{M}_{n-1\,2}(\theta_{n-1\,2})\cdots \bar{M}_{n-1\,n-2}(\theta_{n-1\,n-2}) \times \bar{M}_{n-1\,n-1}(\theta_{n-1\,n-1})\bar{M}_{n-2\,2}(\theta_{n-2\,2})\cdots \bar{M}_{22}(\theta_{22}),$$

where

$$\bar{M}_{nj}(\theta_{nj}) = f^{-1}M_{nj}(\theta_{nj})f, \quad j = 2, 3, \ldots, n-2,$$

$$\bar{M}_{n\,n-1}(\theta_{n\,n-1}) = f^{-1}M_{n\,n-1}(\pi/2 - i\theta_{n\,n-1})d_1,$$

$$\bar{M}_{nn}(\theta_{nn}) = d_1^{-1}M_{nn}(-\theta_{nn})d_1,$$

$$\bar{M}_{n-1\,j}(\theta_{n-1\,j}) = d_1^{-1}M_{n-1\,j}(i\theta_{n-1\,j})d_1, \quad j = 2, 3, \ldots, n-2,$$

$$\bar{M}_{n-1\,n-1}(\theta_{n-1\,n-1}) = \bar{M}_{n-1\,n-1}(\theta_{n-1\,n-1})A,$$

$$\bar{M}_{n-1\,n-1}(\theta_{n-1\,n-1}) = d_1^{-1}M_{n-1\,n-1}(\pi/2 - i\theta_{n-1\,n-1})d_2,$$

$$\bar{M}_{ij}(\theta_{ij}) = f^{-1}M_{ij}(\theta_{ij})f, \quad 2 \leqslant j \leqslant i \leqslant n-2,$$

with d_1, d_2, and A as the $n \times n$ diagonal matrices (the entries within the square brackets give the diagonal elements)

$$d_1 = \text{diag}[-i, 1, 1, \ldots, 1, -1, i],$$

$$d_2 = \text{diag}[-1, 1, 1, \ldots, 1, -i, i],$$

$$A = \text{diag}[-1, 1, 1, \ldots, 1, -1, 1].$$

It is easy to check from these that the Euler angles

$$\left.\begin{array}{ll} \theta_{nj}, & j = 2, 3, \ldots, n-2, n \\ \theta_{ij}, & 2 \leqslant j \leqslant i \leqslant n-2 \end{array}\right\} \text{ are of first kind,} \tag{17a}$$

while

$$\theta_{n-1\,j}, \quad j = 2, 3, \ldots, n-2, \quad \text{are of second kind.} \tag{17b}$$

In addition,

$$\bar{M}_{n\,n-1}(\theta_{n\,n-1}) = \begin{bmatrix} \sinh\theta_{n\,n-1} & \cdots & \cosh\theta_{n\,n-1} & 0 \\ \vdots & & \vdots & \vdots \\ \cosh\theta_{n\,n-1} & \cdots & -\sinh\theta_{n\,n-1} & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

and $\bar{M}_{n-1\,n-1}(\theta_{n-1\,n-1})$ is a similar matrix with $\theta_{n\,n-1}$ replaced by $-\theta_{n-1\,n-1}$. We call the transformation represented by the above matrix [which is *not* an element of $L(n, r)$] a Barred Lorentz transformation by an angle $\theta_{n\,n-1}$ in the $(1, n-1)$ plane and the corresponding angle, an Euler angle (in context of Lorentz group) of third kind. Thus we get

$$\theta_{n\,n-1}, \theta_{n-1\,n-1} \text{ are of third kind,} \tag{17c}$$

so that the nature of the $\frac{1}{2}n(n-1)$ real Euler angles in this case is given by (17a)–(17c).

It follows that an arbitrary element of $L(n, 1)$ or $L(n, 2)$ can be factorized as a product of $\frac{1}{2}n(n-1)$ real matrices, each associated with a real Euler angle which is either of first kind or of second kind or of third kind. In Paper II of this series, we prove that this statement remains valid for $L(n, r)$, r arbitrary. Note in the end that these angles of third kind are the new and interesting feature which appear only when $r > 1$ and which were mentioned in the Introduction.

APPENDIX

We prove here the result that for given complex numbers z, z_1 and z_2 such that

$$z_1^2 + z_2^2 = 1,$$

each of the following three systems of equations for x has a unique solution:

$$\sin x = z, \tag{A1a}$$

$$-\pi/2 \leqslant \mathrm{Re}\, x \leqslant \pi/2; \tag{A1b}$$

$$\cos x = z, \tag{A2a}$$

$$0 \leqslant \mathrm{Re}\, x \leqslant \pi; \tag{A2b}$$

$$\begin{cases} \sin x = z_1, & \text{(A3a)} \\ \cos x = z_2, & \text{(A3b)} \\ 0 \leqslant \mathrm{Re}\, x \leqslant 2\pi. & \text{(A3c)} \end{cases}$$

(1a) gives

$$e^{ix} - e^{-ix} = 2iz$$

$$\Rightarrow e^{2ix} - 2ize^{ix} - 1 = 0$$

$$\Rightarrow e^{ix} = iz \pm (1 - z^2)^{1/2}$$

$$\Rightarrow x = \begin{cases} -i \log|iz + (1 - z^2)^{1/2}| + \arg[iz + (1 - z^2)^{1/2}] \\ \text{or} \\ -i \log|iz - (1 - z^2)^{1/2}| + \arg[iz - (1 - z^2)^{1/2}]. \end{cases} \tag{A4}$$

Now the identity

$$[iz - (1 - z^2)^{1/2}][iz + (1 - z^2)^{1/2}] = -1$$

$$\Rightarrow \arg[iz - (1 - z^2)^{1/2}] + \arg[iz + (1 - z^2)^{1/2}] = \pi.$$

This together with the fact that both the arguments lie between $-\pi$ and π implies that precisely one of them lies between $-\pi/2$ and $\pi/2$, i.e., precisely one of the two solutions (A4) satisfies (A1b). Thus the system (A1a), (A1b) has a unique solution.

(A2a) similarly leads to

$$x = \begin{cases} -i \log|z + (z^2 - 1)^{1/2}| + \arg[z + (z^2 - 1)^{1/2}] \\ \text{or} \\ -i \log|z - (z^2 - 1)^{1/2}| + \arg[z - (z^2 - 1)^{1/2}], \end{cases} \tag{A5}$$

while the identity

$$[z - (z^2 - 1)^{1/2}][z + (z^2 - 1)^{1/2}] = 1$$

gives

$$\arg[z - (z^2 - 1)^{1/2}] + \arg[z + (z^2 - 1)^{1/2}] = 0,$$

so that just one of the two arguments lie between 0 and π. Thus just one of the two solutions (5) satisfies both (2a) and (2b), as required.

Now we have already seen that (3a) gives two possible values of x, say x_1 and x_2, both of which satisfy (A3c). Clearly, we have

$$e^{ix_1} = iz_1 + (1 - z_1^2)^{1/2}, \quad e^{ix_2} = iz_1 - (1 - z_1^2)^{1/2}$$

$$\Rightarrow \cos x_1 = (1 - z_1^2)^{1/2}, \quad \cos x_2 = -(1 - x_1^2)^{1/2}.$$

Now as z_2 is equal to one of the two numbers

$$(1 - z_1^2)^{1/2}, -(1 - z_1^2)^{1/2},$$

(A3b) will be satisfied by one and only one of the two solutions x_1 and x_2. Hence the system (A3a), (A3b), and (A3c) has a unique solution.

[1] A. Syed, J. Natl. Sci. Math. (Pakistan) **9**, 271 (1969).
[2] R. C. Raffenetti and K. Ruedenberg, Int. J. Quantum. Chem. **38**, 625 (1970).
[3] D. K. Hoffman, R. C. Raffenetti, and K. Ruedenberg, J. Math. Phys. **13**, 528 (1972).
[4] K. B. Wolf and R. L. Anderson, J. Math. Phys. **11**, 3177 (1970).
[5] K. B. Wolf, J. Math. Phys. **12**, 197 (1971).
[6] K. B. Wolf, J. Math. Phys. **13**, 1634 (1972).
[7] T. Mackawa, J. Math. Phys. **13**, 655 (1972).
[8] T. Mackawa, J. Math. Phys. **16**, 334 (1975).
[9] T. Mackawa, J. Math. Phys. **16**, 2372 (1975).
[10] M. K. F. Wong, J. Math. Phys. **15**, 25 (1974).
[11] W. Rindler, *Essentials of Relativity* (Springer, New York, 1977), p. 187.
[12] A. Qadir, Int. J. Theo. Phys. **14**, 1 (1976); F. S. Kotz, J. Math. Phys. **15**, 2242 (1974).
[13] F. D. Murnaghan, *The Unitary and Rotation Groups* (Spartan, Washington, D.C., 1962).
[14] A. Syed, J. Sci. (Karachi Univ.) **7**, 39 (1979).

Euler angle parametrization of the complex rotation group and its subgroups. II

Ansaruddin Syed
Department of Mathematics, University of Karachi, Karachi, Pakistan

(Received 22 October 1981; accepted for publication 15 January 1982)

The work of Paper I is extended here by obtaining the reality nature of Euler angles of the subgroups $\hat{L}(n,r)$ (defined there) of $O^+(n,C)$, for arbitrary r, and then showing that these again enable one to define "real" Euler angles of the generalized Lorentz groups $L(n,r)$, r arbitrary.

PACS numbers: 02.20.Qs

I. INTRODUCTION

In the first paper of this series,[1] to be referred to as Paper I, we were able to define Euler angles

$$\{x_{jk},\ 2\leqslant k\leqslant j\leqslant n\}$$

of an element $R\in O^+(n,C)$. This was done with the help of induction, as follows. Suppose that

$$\{x_{jk},\ 2\leqslant k\leqslant j,\ j\geqslant i+1\}$$

have already been defined and that

T_{jk} = rotation by an angle x_{jk} in the $(1,k)$ plane, $k=2,3,\ldots,j-1$,

T_{jj} = rotation by an angle $-x_{jj}$ in the $(1,j)$ plane.

Let $M_{jk}=M_{jk}(x_{jk})$ be the matrix of the rotation T_{jk} and

$$N_j=M_{j2}M_{j3}\cdots M_{jj},\quad j\geqslant i+1.$$

If $\mathbf{t}^i$ is the ith column of R it is found that

$$N_{i+1}^T N_{i+2}^T\cdots N_n^T\mathbf{t}^i$$

is of the form

$$[c_1,c_2,\ldots,c_i,0,0,\ldots,0]^T.$$

We define

$$\{x_{ij},\ 2\leqslant j\leqslant i\}$$

as the polar angles of

$$[c_1,c_2,\ldots,c_i]^T.$$

It was noted there that

$$R=N_nN_{n-1}\cdots N_2, \tag{1}$$

and that if $R^{(i)}$ is obtained from

$$R^i=N_{i+1}^T N_{i+2}^T\cdots N_n^T R \tag{2}$$

by deleting the last $(n-r)$ rows and columns, then

$$x_{jk},\ 2\leqslant k\leqslant j\leqslant i \text{ are the Euler angles of } R^{(i)}. \tag{3}$$

We next defined the subgroup $\hat{L}(n,r)$ of $O^+(n,C)$ as the collection of those of its elements which keep the real vector space $\hat{M}(n,r)$ generated by the elements

$$\{e_1,e_2,\ldots,e_r,ie_{r+1},\ldots,ie_n\}$$

of C^n, invariant, e_j being the unit vector with all elements zero except the jth one, which is 1. The reality conditions characterizing the elements of $\hat{\alpha}\in\hat{L}(n,r)$ are

$$\begin{aligned}
&\hat{\alpha}_{jk}=\alpha_{jk} \quad \text{for} \begin{cases}\text{(i) } j\leqslant r,\ k\leqslant r,\\ \text{(ii) } j\geqslant r+1,\ k\geqslant r+1,\end{cases}\\
&\hat{\alpha}_{jk}=i\alpha_{jk} \quad \text{for } j\geqslant r+1,\ k\leqslant r,\\
&\hat{\alpha}_{jk}=-i\alpha_{jk} \quad \text{for } j\leqslant r,\ k\geqslant r+1,\\
&\qquad\alpha_{jk} \text{ real for all } j,k.
\end{aligned} \tag{4}$$

The nature of Euler angles of $\hat{L}(n,r)$ was considered in the special cases $r=1,2$, where they were found to be either real or pure imaginary or barred, meaning that they are of the form $\pi/2-i\theta$, θ real. We now prove in Sec. II that this statement remains valid for arbitrary r.

Generalized Lorentz groups $L(n,r)$ were then defined as (real) orthogonal linear transformations in a real n-dimensional space with norm

$$x_1^2+x_2^2+\cdots+x_r^2-x_{r+1}^2-\cdots-x_n^2,$$

and were found to be isomorphic with $\hat{L}(n,r)$. In the end, we proved, when $r=1$ or 2, that:

(i) There exists a set of real Euler angles for $\alpha\in L(n,r)$, each having an associated real matrix, such that α is a product of these matrices.

(ii) The matrix associated with any particular Euler angle of $\alpha\in L(n,r)$ is one out of a set of three different given forms; this leads naturally to the classification of the angle as either of first or of second or of third kind.

In Sec. IV below, we extend these results to all permissible values of r.

II. REALITY NATURE OF EULER ANGLES OF $\hat{L}(n,r)$

This nature is obtained with the help of the following two fundamental results and their consequences.

Theorem 1: Let $\{x_{ij}\}$ be the Euler angles of $\hat{\alpha}\in\hat{L}(n,r)$, and σ_i, $i\geqslant r$, be the number of barred ones of these among

$$\{x_{pq},\ i+1\leqslant p\leqslant n,\ 2\leqslant q\leqslant p\}.$$

Also let $\sigma_i(j)$ be the number of barred angles among

$$\{x_{pj},\ p=i+1,i+2,\ldots,n\}$$

for $2\leqslant j\leqslant i+1$, while for $i+2\leqslant j\leqslant n$, it is given by

$$\sigma_i(j)=\sigma_{i+1}(j)=\cdots=\sigma_{j-1}(j).$$

Then each of the last $(n-r)$ columns of $\hat{\alpha}^i$ is of the form

$$[(i)^{-1+\sigma_i}\mathrm{Re},\ (i)^{-1+\sigma_i(2)}\mathrm{Re},\ldots,(i)^{-1+\sigma_i(r)}\mathrm{Re},\ (i)^{\sigma_i(r+1)}\mathrm{Re},\ldots,(i)^{\sigma_i(n)}\mathrm{Re}]^T, \tag{5}$$

 0022-2488/83/030471-07$02.50

where Re stands for some real number. (Note that Re at different places stands, in general, for different real numbers.)

Lemma 1: If the ith column of $\hat{\alpha}^i$, $i \geqslant r+1$, is of the form (5) so that the ith column of $\hat{\alpha}^{(i)}$ is given by

$$\beta^i = [(i)^{-1+\sigma_i}\beta^i_1,\ (i)^{-1+\sigma_i(2)}\beta^i_2,\ldots,(i)^{-1+\sigma_i(r)}\beta^i_r,$$
$$(i)^{\sigma_i(r+1)}\beta^i_{r+1},\ldots,(i)^{\sigma_i(i)}\beta^i_i]^T, \tag{6}$$
$$\beta^i_p,\quad p=1,2,\ldots,i \text{ all real},$$

and if

$$(-1)^{-1+\sigma_i}(\beta^i_1)^2+(-1)^{-1+\sigma_i(2)}(\beta^i_2)^2+\cdots$$
$$+(-1)^{-1+\sigma_i(p)}(\beta^i_p)^2 \tag{7}$$

changes sign at

$$p=\lambda(1),\ p=\lambda(2),\ldots,p=\lambda(k_i),$$

for $p \leqslant r$, and

$$(-1)^{-1+\sigma_i}(\beta^i_1)^2+(-1)^{-1+\sigma_i(2)}(\beta^i_2)^2+\cdots$$
$$+(-1)^{-1+\sigma_i(r)}(\beta^i_r)^2$$
$$+(-1)^{\sigma_i(r+1)}(\beta^i_{r+1})^2+\cdots+(-1)^{\sigma_i(p)}(\beta^i_p)^2 \tag{8}$$

at

$$p=\lambda(k_i+1),\ldots,p=\lambda(l_i)$$

for $r+1 \leqslant p \leqslant i-1$, then:

(i) $x_{i\lambda(q)}$ is barred for $q=1,2,\ldots,l_i$;

(ii) x_{ii} is barred if and only if $\sigma_i(i)$ is odd; we put $\bar{l}_i = l_i$ if x_{ii} is not barred and

$$\bar{l}_i = l_i+1,\quad \lambda(\bar{l}_i)=i,$$

if x_{ii} is barred;

(iii) for

$$j \neq \lambda(q),\quad q=1,2,\ldots,\bar{l}_i,$$

we have

$$x_{ij} = \begin{cases} (i)^{\rho_i(j)}\theta_{ij}, & j \leqslant r, \\ (i)^{\rho_i(j)+1}\theta_{ij}, & j \geqslant r+1; \end{cases}$$

here

$$\rho_i(j) = \sigma_i + \sigma_i(j) + \tau_i(j),$$

$\tau_i(j)$ is the number of barred angles [according to (i) above] among

$$\{x_{ie},\ 2 \leqslant e \leqslant j-1\}$$

so that it is also equal to the number of changes of sign of (7) for $j \leqslant r+1$ and of (8) for $r+2 \leqslant j \leqslant i$, as p goes from 1 to $(j-1)$, and the θ's, here as well as everywhere else in this paper, are real.

We prove first the lemma and then use it to prove the theorem.

Proof of Lemma 1: We consider the three parts (i), (ii), and (iii) separately.

(i) The fact that

$$x_{ij},\quad j=2,3,\ldots,i,$$

are the polar angles of (6) implies that [see Eq. (2) of Paper I] for $q \leqslant k_i$,

$$\sin x_{i\lambda(q)} = \frac{(i)^{\sigma_i(\lambda(q))}\beta^i_{\lambda(q)}}{[(-1)^{\sigma_i}(\beta^i_1)^2+(-1)^{\sigma_i(2)}(\beta^i_2)^2+\cdots+(-1)^{\sigma_i[\lambda(q)]}(\beta^i_{\lambda(q)})^2]^{1/2}}.$$

If $\sigma_i(\lambda(q))$ is even, we get

$$\sin x_{i\lambda(q)} = \frac{\pm\beta^i_{\lambda(q)}}{[(-1)^{\sigma_i}(\beta^i_1)^2+\cdots+(-1)^{\sigma_i(\lambda(q)-1)}(\beta^i_{\lambda(q)-1})^2+(\beta^i_{\lambda(q)})^2]^{1/2}},$$

with

$$(-1)^{\sigma_i}(\beta^i_1)^2+\cdots+(-1)^{\sigma_i(\lambda(q)-1)}(\beta^i_{\lambda(q)-1})^2<0, \tag{9}$$
$$(-1)^{\sigma_i}(\beta^i_1)^2+\cdots+(-1)^{\sigma_i(\lambda(q)-1)}(\beta^i_{\lambda(q)-1})^2+(\beta^i_{\lambda(q)})^2>0, \tag{10}$$

as (7) changes sign at $p=\lambda(q)$. Thus $\sin x_{i\lambda(q)}$ is real. Further, (9) and (10) also imply

$$|\beta^i_{\lambda(q)}| > [(-1)^{\sigma_i}(\beta^i_1)^2+\cdots+(-1)^{\sigma_i(\lambda(q)-1)}(\beta^i_{\lambda(q)-1})^2+(\beta^i_{\lambda(q)})^2]^{1/2}$$

so that

$$|\sin x_{i\lambda(q)}|>1 \Rightarrow x_{i\lambda(q)} = \pi/2 - i\theta_{i\lambda(q)},$$

i.e., $x_{i\lambda(q)}$ is barred.

On the other hand, if $\sigma_i(\lambda(q))$ is odd,

$$\sin x_{i\lambda(q)} = \frac{\pm\beta^i_{\lambda(q)}}{[-(-1)^{\sigma_i}(\beta^i_1)^2-\cdots-(-1)^{\sigma_i(\lambda(q)-1)}(\beta^i_{\lambda(q)-1})^2+(\beta^i_{\lambda(q)})^2]^{1/2}},$$

and the fact that (7) changes sign at $p=\lambda(q)$ implies that

$$-(-1)^{\sigma_i}(\beta^i_1)^2-\cdots-(-1)^{\sigma_i(\lambda(q)-1)}(\beta^i_{\lambda(q)-1})^2<0,$$
$$-(-1)^{\sigma_i}(\beta^i_1)^2-\cdots-(-1)^{\sigma_i(\lambda(q)-1)}(\beta^i_{\lambda(q)-1})^2+(\beta^i_{\lambda(q)})^2>0.$$

As before, this leads to

$$\sin x_{i\lambda(q)} \text{ real},\ |\sin x_{i\lambda(q)}|>1 \Rightarrow x_{i\lambda(q)} = \pi/2 - i\theta_{i\lambda(q)}.$$

Thus $x_{i\lambda(q)}$ is barred for all $q \leqslant k_i$; that it is barred for $k_i+1 \leqslant q \leqslant l_i$ can be proved in the same way.

(ii) We have

$$\cos x_{ii} = (i)^{\sigma_i(i)}\beta^i_i.$$

Hence

x_{ii} is barred $\Leftrightarrow \cos x_{ii}$ is pure imaginary $\Leftrightarrow \sigma_i(i)$ is odd.

(iii) First, let $j \leqslant r$, so that we have

$$\sin x_{ij} = \frac{(i)^{-1+\sigma_i(j)}\beta^i_j}{[(-1)^{-1+\sigma_i}(\beta^i_1)^2+(-1)^{-1+\sigma_i(2)}(\beta^i_2)^2+\cdots+(-1)^{-1+\sigma_i(j)}(\beta^i_j)^2]^{1/2}}.$$

As $\tau_i(j)$ changes of sign have occurred in (7) as p goes from 1 to j, we will have

$$(-1)^{-1+\sigma_i+\tau_i(j)} = (-1)^{-1+\sigma_i}\cdot(-1)^{\tau_i(j)}$$
$$= \operatorname{sgn}\left[(-1)^{-1+\sigma_i}(\beta_1^i)^2 + \cdots + (-1)^{-1+\sigma_i(j)}(\beta_j^i)^2\right]$$
$$\Rightarrow\left[(-1)^{-1+\sigma_i}(\beta_1^i)^2 + \cdots + (-1)^{-1+\sigma_i(j)}(\beta_j^i)^2\right]^{1/2}$$
$$= \pm(-i)^{-1+\sigma_i+\tau_i(j)}\left[(-1)^{\tau_i(j)}(\beta_1^i)^2 + (-1)^{\sigma_i+\sigma_i(2)+\tau_i(j)}\right.$$
$$\left.\times(\beta_2^i)^2 + \cdots + (-1)^{\sigma_i+\sigma_i(j)+\tau_i(j)}(\beta_j^i)^2\right]^{1/2},$$

where the second factor on the rhs is real,

$$\Rightarrow \sin x_{ij} = \pm(-i)^{\rho_i(j)}\beta_j^i/\left[(-1)^{\tau_i(j)}(\beta_1^i)^2\right.$$
$$\left.+(-1)^{\sigma_i+\sigma_i(2)+\tau_i(j)}(\beta_2^i)^2 + \cdots + (-1)^{\sigma_i+\sigma_i(j)+\tau_i(j)}(\beta_j^i)^2\right]^{1/2}$$

with the denominator real. Thus, if $\rho_i(j)$ is odd, $\sin x_{ij}$ is pure imaginary so that

$$x_{ij} \text{ is pure imaginary} \Rightarrow x_{ij} = (i)^{\rho_i(j)}\theta_{ij}.$$

On the other hand, if $\rho_i(j)$ is even,

$$\sin x_{ij} = \pm\beta_j^i/\left[(-1)^{\tau_i(j)}(\beta_1^i)^2 + (-1)^{\sigma_i+\sigma_i(2)+\tau_i(j)}(\beta_2^i)^2\right.$$
$$\left.+\cdots+(-1)^{\sigma_i+\sigma_i(j)+\tau_i(j)}(\beta_j^i)^2\right]^{1/2}$$

is real, and as both

$$(-1)^{\tau_i(j)}(\beta_1^i)^2 + (-1)^{\sigma_i+\sigma_i(2)+\tau_i(j)}(\beta_2^i)^2 + \cdots$$
$$+(-1)^{\sigma_i+\sigma_i(j-1)+\tau_i(j)}(\beta_{j-1}^i)^2 + (\beta_j^i)^2$$

and

$$(-1)^{\tau_i(j)}(\beta_1^i)^2 + (-1)^{\sigma_i+\sigma_i(2)+\tau_i(j)}(\beta_2^i)^2 + \cdots$$
$$+(-1)^{\sigma_i+\sigma_i(j-1)+\tau_i(j)}(\beta_{j-1}^i)^2$$

are positive,

$$|\sin x_{ij}| < 1 \Rightarrow x_{ij} \text{ is real} \Rightarrow x_{ij} = (i)^{\rho_i(j)}\theta_{ij}.$$

Proceeding in exactly the same way, we can show that

$$x_{ij} = (i)^{1+\rho_i(j)}\theta_{ij}$$

for $r+1 \leqslant j \leqslant i-1$. For $j = i$, we have

$$\cos x_{ii} = (i)^{\sigma_i(i)}\beta_i^i.$$

Because of part (ii) above, we need to consider only the case

$$x_{ii} \text{ is not barred} \Rightarrow \sigma_i(i) \text{ is even}$$

so that

$$\cos x_{ii} = \pm\beta_i^i.$$

Thus, if $|\beta_i^i| < 1$, x_{ii} is real. However, in this case, (8) will be positive for $p = (i-1)$ (it is 1 for $p = i$), so that as the total number of changes of sign of (8) as p goes from 1 to $(i-1)$ is $l_i = \tau_i(i)$, the first sign is

$$(-1)^{-1+\sigma_i}$$

and the last sign is positive; we will have

$$(-1)^{-1+\sigma_i}\cdot(-1)^{\tau_i(i)} = 1 \Rightarrow \sigma_i + \tau_i(i) = \text{odd}$$
$$\Rightarrow 1 + \sigma_i + \sigma_i(i) + \tau_i(i) \equiv 1 + \rho_i(i) = \text{even} \Rightarrow x_{ii} = (i)^{1+\rho_i(i)}\theta_{ii}.$$

On the other hand, if $|\beta_i^i| > 1$, x_{ii} is pure imaginary; also (8) is negative for $p = (i-1)$, so that, as above, we get

$$(-1)^{-1+\sigma_i}\cdot(-1)^{\tau_i(i)} = -1 \Rightarrow \sigma_i + \tau_i(i) \text{ is even}$$
$$\Rightarrow 1 + \rho_i(i) = \text{odd} \Rightarrow x_{ii} = (i)^{1+\rho_i(i)}\theta_{ii}.$$

This completes the proof of the lemma.

Note here the following relations, to be used later, which are essentially the consequences of the definition of $\lambda(q)$:

$$\sigma_i(\lambda(q)) \equiv (\sigma_i + q) \bmod 2, \quad q \leqslant k_i, \tag{11}$$
$$\sigma_i(\lambda(q)) \equiv (1 + \sigma_i + q) \bmod 2, \quad k_i + 1 \leqslant q \leqslant \bar{l}_i. \tag{12}$$

For, if $q \leqslant k_i$, the fact that (7) changes sign at $p = \lambda(q)$ implies that its sign at this value of p will be that of the last term added, i.e., it will be

$$(-1)^{-1+\sigma_i(\lambda(q))}.$$

On the other hand, the facts that q changes of sign have ocurred in (7) as p goes from 1 to $\lambda(q)$ and that the sign of the first term is

$$(-1)^{-1+\sigma_i}$$

implies that its sign at $p = \lambda(q)$ will be

$$(-1)^{-1+\sigma_i+q}.$$

Hence

$$\sigma_i(\lambda(q)) \equiv (\sigma_i + q) \bmod 2.$$

(12) can be proved in the same way for $k_i \leqslant q \leqslant l_i$. For $q = \bar{l}_i = l_i + 1$, x_{ii} is barred and $i = \lambda(\bar{l}_i)$

$$\Rightarrow \sigma_i(i) \text{ is odd}$$
$$\Rightarrow (-1)^{\sigma_i(i)}(\beta_i^i)^2 \text{ is negative}$$
$$\Rightarrow (8) \text{ is positive (in fact} > 1) \text{ for } p = (i-1)$$
$$\Rightarrow (8) \text{ is positive for } p = \lambda(l_i)$$
$$\Rightarrow \sigma_i(\lambda(l_i)) \text{ is even}$$
$$\Rightarrow 1 + \sigma_i + l_i \text{ is even}$$
$$\Rightarrow \sigma_i + \bar{l}_i \text{ is even}$$
$$\Rightarrow \sigma_i(\lambda(\bar{l}_i)) \equiv (1 + \sigma_i + \bar{l}_i) \bmod 2.$$

Proof of Theorem 1: We use here the method of induction. As the theorem is obviously true for $i = n$ [σ_n as well as $\sigma_n(j)$, $j = 2,3,\ldots,n$, are all zero], we assume it for $i = \mu$ with

$$r+1 \leqslant \mu \leqslant n,$$

and show that it is then true for $i = (\mu - 1)$ also.

Let $\gamma^{\mu-1}$ be any of the last $(n-r)$ columns of $\hat{\alpha}^{\mu-1}$ and γ^{μ} be the corresponding column of $\hat{\alpha}^{\mu}$ so that

$$\gamma^{\mu-1} = N_{\mu}^{T}\gamma^{\mu} \equiv M_{\mu\mu}^{T}M_{\mu\mu-1}^{T}\cdots M_{\mu 2}^{T}\gamma^{\mu},$$

where the M's are as in the Introduction. As the theorem is true for $i = \mu$, the μth column of $\hat{\alpha}^{\mu}$ is of the form (5) with $i = \mu$, and so the the angles

$$x_{\mu j}, \quad 2 \leqslant j \leqslant \mu$$

are as in Lemma 1. Thus

$$x_{\mu 2} = (i)^{\sigma_\mu + \sigma_\mu(2)}\theta_{\mu 2},$$

so that $M_{\mu 2}$ is the matrix

$$\begin{bmatrix} \text{Re} & (i)^{\sigma_\mu+\sigma_\mu(2)}\text{Re} & \cdot & \cdot \\ (i)^{\sigma_\mu+\sigma_\mu(2)}\text{Re} & \text{Re} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix},$$

where, here as well as in all that follows, all the elements of the matrix which have not been explicitly written are assumed to be those of the unit $n\times n$ matrix. As γ^{μ} is of the form (5) with $i=\mu$, it is easy to verify, by actual multiplication, that

$$M^{T}_{\mu 2}\gamma^{\mu}$$

is also of the same form. It can be similarly shown that

$$M^{T}_{\mu 3}M^{T}_{\mu 2}\gamma^{\mu},\ldots,M^{T}_{\mu\lambda(1)-1}\cdots M^{T}_{\mu 2}\gamma^{\mu}$$

are all of the form (5) with $i=\mu$. Next, as

$$x_{\mu\lambda(1)}=\pi/2-i\theta_{\mu\lambda(1)},$$

$M_{\mu\lambda(1)}$ is the matrix

$$\begin{array}{c}\lambda\,(1)\text{th column}\\ \uparrow\\ \begin{bmatrix} i\cdot\mathrm{Re} & \cdot & \cdot & \mathrm{Re} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathrm{Re} & \cdot & \cdot & i\cdot\mathrm{Re} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}\leftarrow\lambda\,(1)\text{th row}\end{array}$$

As $\sigma_{\mu}(\lambda\,(1))=\sigma_{\mu}+1$ by (11), actual multiplication by a column of form (5) with $i=\mu$ shows that

$$M^{T}_{\mu\lambda(1)}M^{T}_{\mu\lambda(1)-1}\cdots M^{T}_{\mu 2}\gamma^{\mu}$$

is of the form

$$\begin{aligned}&[(i)^{\sigma_{\mu}}\mathrm{Re},(i)^{-1+\sigma_{\mu}(2)}\mathrm{Re},\ldots,(i)^{-1+\sigma_{\mu}(\lambda(1)-1)}\mathrm{Re},\\ &\quad(i)^{\sigma_{\mu}(\lambda(1))}\mathrm{Re},(i)^{-1+\sigma_{\mu}(\lambda(1)+1)}\mathrm{Re},\ldots,\\ &\quad(i)^{-1+\sigma_{\mu}(r)}\mathrm{Re},(i)^{\sigma_{\mu}(r+1)}\mathrm{Re},\ldots,(i)^{\sigma_{\mu}(n)}\mathrm{Re}]^{T}.\end{aligned}\tag{13}$$

Using now

$$x_{\mu\lambda(1)+1}=(i)^{1+\sigma_{\mu}+\sigma_{\mu}(\lambda(1)+1)}\theta_{\mu\lambda(1)+1},$$

explicit multiplication shows that $M^{T}_{\mu\lambda(1)+1}$ transforms a column of form (13) to one of the same form. Thus

$$M^{T}_{\mu\lambda(1)+1}M^{T}_{\mu\lambda(1)}\cdots M^{T}_{\mu 2}\gamma^{\mu},$$

and similarly

$$M^{T}_{\mu\lambda(2)-1}\cdots M^{T}_{\mu 2}\gamma^{\mu},$$

also, will be of form (13). Again as

$$x_{\mu\lambda(2)}=\pi/2-i\theta_{\mu\lambda(2)}$$

and $\sigma_{\mu}(\lambda\,(2))=2+\sigma_{\mu}$ by (11), it can be verified that

$$M^{T}_{\mu\lambda(2)}M^{T}_{\mu\lambda(2)-1}\cdots M^{T}_{\mu 2}\gamma^{\mu}$$

is of the form

$$\begin{aligned}&[(i)^{i+\sigma_{\mu}}\mathrm{Re},(i)^{i+\sigma_{\mu}(2)}\mathrm{Re},\ldots,(i)^{-1+\sigma_{\mu}(\lambda(1)-1)}\mathrm{Re},\\ &\quad(i)^{\sigma_{\mu}(\lambda(1))}\mathrm{Re},(i)^{-1+\sigma_{\mu}(\lambda(1)+1)}\mathrm{Re},\ldots,\\ &\quad(i)^{-1+\sigma_{\mu}(\lambda(2)-1)}\mathrm{Re},(i)^{\sigma_{\mu}(\lambda(2))}\mathrm{Re},(i)^{-1+\sigma_{\mu}(\lambda(2)+1)}\mathrm{Re},\\ &\quad\ldots,(i)^{-1+\sigma_{\mu}(r)}\mathrm{Re},(i)^{\sigma_{\mu}(r+1)}\mathrm{Re},\ldots,(i)^{\sigma_{\mu}(n)}\mathrm{Re}]^{T}.\end{aligned}$$

Continuing this process with the help of (11) and (12), we will ultimately see that

$$\gamma^{\mu-1}\equiv M^{T}_{\mu\mu}M^{T}_{\mu\mu-1}\cdots M^{T}_{\mu 2}\gamma^{\mu}$$

is of the form

$$\begin{aligned}&[(i)^{-1+\bar{l}_{\mu}+\sigma_{\mu}}\mathrm{Re},(i)^{-1+\sigma_{\mu}(2)}\mathrm{Re},\ldots,(i)^{-1+\sigma_{\mu}(\lambda(p)-1)}\mathrm{Re},\\ &\quad(i)^{\sigma_{\mu}(\lambda(p))}\mathrm{Re},(i)^{-1+\sigma_{\mu}(\lambda(p)+1)}\mathrm{Re},\ldots,(i)^{-1+\sigma_{\mu}(r)}\mathrm{Re},\\ &\quad(i)^{\sigma_{\mu}(r+1)}\mathrm{Re},\ldots,(i)^{\sigma_{\mu}(\lambda(q)-1)}\mathrm{Re},\ldots,\\ &\quad(i)^{1+\sigma_{\mu}(\lambda(q))}\mathrm{Re},(i)^{\sigma_{\mu}(\lambda(q)+1)}\mathrm{Re},\ldots,(i)^{\sigma_{\mu}(\mu)}\mathrm{Re},\\ &\quad\ldots,(i)^{\sigma_{\mu}(n)}\mathrm{Re}]^{T},\end{aligned}\tag{14}$$

where p and q take on all the values

$$1,2,\ldots,k_{\mu},$$

and

$$k_{\mu}+1,k_{\mu}+2,\ldots,\bar{l}_{\mu},$$

respectively.

Looking now at the definitions of σ_i and $\sigma_i(j)$ in the statement of the theorem and at the nature of the angles

$$x_{\mu j},\quad 2<j\leqslant\mu,$$

as given by Lemma 1, we see that

$$\begin{aligned}&\sigma_{\mu-1}=\sigma_{\mu}+\bar{l}_{\mu},\\ &\sigma_{\mu-1}(\lambda\,(q))=1+\sigma_{\mu}(\lambda\,(q)),\quad q=1,2,\ldots,\bar{l}_{\mu},\\ &\sigma_{\mu-1}(j)=\sigma_{\mu}(j),\quad j\neq\lambda\,(q),\quad q=1,2,\ldots,\bar{l}_{\mu},\end{aligned}$$

so that $\gamma^{\mu-1}$ is really of the form

$$[(i)^{-1+\sigma_{\mu-1}}\mathrm{Re},(i)^{-1+\sigma_{\mu-1}(2)}\mathrm{Re},\ldots,(i)^{\sigma_{\mu-1}(n)}]^{T},$$

i.e., of the form (5) with $i=\mu-1$. This proves that the theorem is true for $i=\mu-1$ as required.

Corollary 1: Each of the first r columns of $\hat{\alpha}^{i}$ is of the form

$$\begin{aligned}&[(i)^{\sigma_i}\mathrm{Re},(i)^{\sigma_i(2)}\mathrm{Re},\ldots,(i)^{\sigma_i(r)}\mathrm{Re},\\ &\quad(i)^{1+\sigma_i(r+1)}\mathrm{Re},\ldots,(i)^{-1+\sigma_i(n)}\mathrm{Re}]^{T},\end{aligned}$$

for each of the first r columns of $\hat{\alpha}$ is of the form

$$i\times\text{one of the last }(n-r)\text{ columns of }\hat{\alpha}.$$

Corollary 2: The reality nature of the Euler angles of an element $\hat{\alpha}\in\hat{L}\,(n,r)$ are such that:

(i) For $i<r$,

$\sigma_i(j)\equiv[\sigma_i+\tau_i(j)]\ \mathrm{mod}\ 2\Rightarrow x_{ij}$ is real,

$\sigma_i(j)\equiv[\sigma_i+\tau_i(j)+1]\ \mathrm{mod}\ 2\Rightarrow x_{ij}$ is pure imaginary or barred.

(ii) For $r+1\leqslant j\leqslant i-1$,

$\sigma_i(j)\equiv[\sigma_i+\tau_i(j)]\ \mathrm{mod}\ 2\Rightarrow x_{ij}$ is pure imaginary or barred,

$\sigma_i(j)\equiv[\sigma_i+\tau_i(j)+1]\ \mathrm{mod}\ 2\Rightarrow x_{ij}$ is real.

(iii) $\sigma_i(i)=\mathrm{odd}\Rightarrow x_{ii}$ is barred,

$\sigma_i(i)=\mathrm{even},\quad \sigma_i+l_i=\mathrm{even}\Rightarrow x_{ii}$ is pure imaginary,

$\sigma_i(i)=\mathrm{even},\quad \sigma_i+l_i=\mathrm{odd}\Rightarrow x_{ii}$ is real.

These follow immediately from Lemma 1 and Eqs. (11) and (12).

Corollary 3: $\sigma_i(i+1)=\mathrm{even}$ for every $i\geqslant r$.

For, consider $\sigma_{i+1}(i+1)$. If it is odd, $x_{i+1\,i+1}$ is barred

so that $\sigma_i(i+1)$ is even, while if it is even, x_{ii} is not barred so that

$$\sigma_i(i+1) = \sigma_{i+1}(i+1) = \text{even again.}$$

Theorem 2: Let

$$\bar{\sigma}_i \equiv \sigma_i \mod 2,$$

$$\bar{\sigma}_i(j) \equiv [\sigma_i(j)] \mod 2, \quad 2 \leqslant j \leqslant r,$$

$$\bar{\sigma}_i(j) \equiv [\sigma_i(j)+1] \mod 2, \quad r+1 \leqslant j \leqslant n,$$

with $\bar{\sigma}_i$ and each of

$$\bar{\sigma}_i(j), \quad j = 2,3,\ldots,n,$$

either 0 or 1. Then the number of zeros and ones in the set

$$\{\bar{\sigma}_i, \bar{\sigma}_i(2),\ldots,\bar{\sigma}_i(n)\} \tag{15}$$

is independent of i and equals r and $(n-r)$, respectively.

Note first of all that the required numbers are indeed r and $(n-r)$ in (15) with $i = n$. Hence, it is sufficient to prove that they do not change as we go from (15) with $i = \mu$ to that with $i = \mu - 1$.

Now among the angles

$$x_{\mu j}, \quad j = 2,3,\ldots,\mu,$$

$\bar{l}_\mu$ are barred and [by (11) and (12)] $\bar{\sigma}_\mu(j)$ of any two consecutive barred $x_{\mu j}$ are different. Hence, if $\bar{l}_\mu$ is even, $\bar{\sigma}_\mu = \bar{\sigma}_{\mu-1}$ while equal number of zeros and ones change over to ones and zeros, respectively, in going from the set

$$\{\bar{\sigma}_i(2), \bar{\sigma}_i(3),\ldots,\bar{\sigma}_i(n)\} \tag{16}$$

with $i = \mu$ to that with $i = \mu - 1$; the number of zeros and ones in (15) therefore remains the same as we go from $i = \mu$ to $i = \mu - 1$. If, however, $\bar{l}_\mu$ is odd and $\bar{\sigma}_\mu = 0$, then $\bar{\sigma}_{\mu-1} = 1$ and in going from $i = \mu$ to $i = \mu - 1$ in (16), $(\bar{l}_\mu + 1)/2$ ones change to zeros while only $(\bar{l} - 1)/2$ zeros change into ones [the first barred angle $x_{\mu j}$ corresponds to $\bar{\sigma}_\mu(j) = 1$] so that the numbers of zeros and ones in (15) again do not change as we go from $i = \mu$ to $i = \mu - 1$. Similar arguments can now be applied in the case $\bar{l}_\mu$ odd and $\bar{\sigma}_\mu = 1$, to complete the proof of the required result.

Corollary 4: The number of real and pure imaginary elements in the first r columns of $\hat{\alpha}^i$ are r and $(n-r)$, respectively, for all i, while those in the last $(n-r)$ columns are $(n-r)$ and r respectively.

For, according to Corollary 1 and Theorem 1, these numbers are essentially the number of zeros and ones in (15).

Corollary 5: $\bar{\sigma}_r, \bar{\sigma}_r(2),\ldots,\bar{\sigma}_r(r)$ are all zero.

For, by Corollary 3,

$$\bar{\sigma}_r(r+1), \bar{\sigma}_r(r+2) = \bar{\sigma}_{r+1}(r+2),\ldots,\bar{\sigma}_r(n) = \bar{\sigma}_{n-1}(n)$$

are all unity so that the result follows by Theorem 2. Note that an important consequence of this is that $\hat{\alpha}^{(r)}$ is real so that so also will be

$$x_{ij}, \quad 2 \leqslant j \leqslant i \leqslant r,$$

as these are the Euler angles of $\hat{\alpha}^{(r)}$.

We are now in a position to give complete description of the reality nature of the Euler angles of an element $\hat{\alpha} \in \hat{L}(n,r)$. For, by Theorem 1, the last column of $\hat{\alpha}^{(i)}$ is given by (6) so that

$$x_{ij}, \quad 2 < j \leqslant i,$$

will be as in Lemma 1, and hence also as in Corollary 2. Putting now

$$i = n, n-1,\ldots,r+1$$

in succession, we get the reality nature of

$$x_{ij}, \quad 2 \leqslant j \leqslant i, \quad r+1 \leqslant i \leqslant n.$$

The rest of the angles

$$x_{ij}, \quad 2 \leqslant j \leqslant i \leqslant r,$$

are all real.

III. EULER ANGLES GIVING A GENERAL ELEMENT OF $\hat{L}(n,r)$

We now come to the closely related problem of writing down a set of angles

$$x_{ij}, \quad 2 \leqslant j \leqslant i \leqslant n,$$

which represent, i.e., are the Euler angles of a general element of $\hat{L}(n,r)$. We do this according to the following rule

Rule: We write down the angles in the order

$$x_{n2}x_{n3},\ldots,x_{nn},x_{n-1\,2},\ldots,x_{n-1\,n-1},$$
$$x_{n-2\,2},\ldots,x_{32},x_{33},x_{22} \tag{17}$$

and take any given x_{ij} to be such that its reality nature satisfies the conditions of Corollary 2, but is otherwise arbitrary; this is obviously possible as the said conditions involve only the reality nature of those x_{pq} which come before x_{ij} in the order (17).

Now suppose that in the choice of

$$x_{ij}, \quad j = 2,3,\ldots,i,$$

made above, $x_{i\lambda(q)}$ with

$$q = 1,2,\ldots,l_i,$$

and

$$2 \leqslant \lambda(1) < \lambda(2) \cdots < \lambda(k_i) \leqslant r < \lambda(k_i+1) < \cdots < \lambda(l_i) \leqslant i-1,$$

are barred, and also that if x_{ii} is barred, we take $\bar{l}_i = l_i + 1$ and $\lambda(\bar{l}_i) = i$ while if it is not barred, we take $\bar{l}_i = l_i$. Then it is a simple matter to check that the conditions (i), (ii), and (iii) on the reality nature of x_{ij} given in the statement of Lemma 1, as well as Eqs. (11) and (12) are satisfied. It follows that Theorem 2 as well as Corollaries 3 and 5 will also hold.

Consider now the matrix

$$\delta^i = N_i N_{i-1} \cdots N_2, \quad i \geqslant r.$$

We show that each of its last $(n-r)$ columns is of the form (5). Note first of all that this is certainly so for $i = r$ as δ^r is real (Corollary 5) with its last $(n-r)$ columns identical with the corresponding columns of the $n \times n$ unit matrix, and

$$\sigma_r(r+1), \sigma_r(r+2),\ldots,\sigma_r(n)$$

are all even (Corollary 3). Next, as in the proof of Theorem 1, we have shown that N_μ^T operating on a column of form (5) with $i = \mu$ gives a column of the same form with $i = \mu - 1$; we also have the converse result that N_μ operating on a column of form (5) with $i = \mu - 1$ gives a column of the same form with $i = \mu$. Hence, keeping in view the fact that x_{pq} appearing here and in Theorem 1 satisfy the same reality

conditions, viz., the conditions of Corollary 2, it follows that if any of the last $(n-r)$ columns of $\delta^{\mu-1}$ is of the form (5) with $i=\mu-1$, the corresponding column of δ^{μ} will be of the same form with $i=\mu$. We therefore get the required result by induction.

Similar arguments can be used to show that each of the first r columns of δ^{i} is of the form

$$i\times\text{a column of form (5).}$$

As

$$\sigma_n(j)=0 \quad \text{for all } j\in\{2,3,\ldots,n\},$$

this implies that

$$\delta^{n}=N_nN_{n-1}\cdots N_2$$

will be an element of $\hat{L}(n,r)$. We thus have complete specification of Euler angles of $\hat{L}(n,r)$; they satisfy the conditions of Corollary 2, while any set of angles satisfying these conditions and being otherwise arbitrary, are the Euler angles of an element of $\hat{L}(n,r)$.

IV. EULER ANGLES OF GENERALIZED LORENTZ GROUP $L(n,r)$

We now define a set of $\frac{1}{2}n(n-1)$ real Euler angles of an element α of the generalized Lorentz group $L(n,r)$ and discuss their nature for general r between 1 and $(n-1)$. Recall from Paper I that

$$\alpha\in L(n,r)\Rightarrow\hat{\alpha}=f\alpha f^{-1}\in\hat{L}(n,r).$$

Let

$$x_{ij},\quad 2\leqslant j\leqslant i\leqslant n,$$

be the (complex) Euler angles of $\hat{\alpha}$, so that

$$\begin{aligned}\hat{\alpha}&=N_nN_{n-1}\cdots N_2\\&=M_{n2}(x_{n2})M_{n3}(x_{n3})\cdots M_{nn}(x_{nn})M_{n-12}(x_{n-12})\\&\quad\times\cdots M_{22}(x_{22})\\&\Rightarrow\alpha=f^{-1}M_{n2}(x_{n2})M_{n3}(x_{n3})\cdots M_{22}(x_{22})f.\end{aligned}$$

As

$$x_{ij}=\theta_{ij}\quad\text{for } 2\leqslant j\leqslant i\leqslant r,$$

we get

$$\bar{M}_{ij}(\theta_{ij})=f^{-1}M_{ij}(x_{ij})f=M_{ij}(\theta_{ij}),\quad 2\leqslant j\leqslant i\leqslant r,\tag{18}$$

and so

$$\begin{aligned}\alpha=&f^{-1}M_{n2}(x_{n2})M_{n3}(x_{n3})\cdots M_{r+1\,r+1}(x_{r+1\,r+1})f\\&\times\bar{M}_{r2}(\theta_{r2})\cdots\bar{M}_{22}(\theta_{22}).\end{aligned}\tag{19}$$

As we are now going to use a large number of diagonal matrices, we recall the usual convenient notation

$$\text{diag}[D_1,D_2,\ldots,D_n]$$

for the $n\times n$ diagonal matrix

$$D=\{D_i\delta_{ij}\},$$

and note that if

$$D=\text{diag}[D_1,D_2,\ldots,D_n],$$

$$D'=\text{diag}[D'_1,D'_2,\ldots,D'_n],$$

and

$$A=\{A_{ij}\}$$

is any $n\times n$ matrix, then

$$(DAD')_{ij}=D_iA_{ij}D'_j.\tag{20}$$

We now define a set of diagonal matrices

$$d_{ij},\quad 2\leqslant j\leqslant i+1,\ r\leqslant i\leqslant n,$$

as follows:

As we know how to calculate x_{ij}, their

$$\sigma_i,\quad \sigma_i(j),\quad\text{and}\quad\tau_i(j)$$

can all be considered known. We take, for $2\leqslant j\leqslant(r+1)$,

$$\begin{aligned}d_{ij}=\text{diag}[&(-i)^{\sigma_i+\tau_i(j)},(i)^{\sigma_i(2)},(i)^{\sigma_i(3)},\ldots,\\&(i)^{\sigma_i(\lambda(p)-1)},(i)^{1+\sigma_i(\lambda(p))},(i)^{\sigma_i(\lambda(p)+1)},\ldots,\\&(i)^{\sigma_i(\lambda(q)-1)},(i)^{\sigma_i(\lambda(q))},(i)^{\sigma_i(\lambda(q)+1)},\ldots,\\&(i)^{\sigma_i(r)},(i)^{1+\sigma_i(r+1)},\ldots,(i)^{1+\sigma_i(n)}],\end{aligned}\tag{21}$$

where p and q take on the values

$$p=1,2,\ldots,\tau_i(j),$$

$$q=\tau_i(j)+1,\ldots,k_i,$$

while, for $(r+1)\leqslant j\leqslant(i+1)$,

$$\begin{aligned}d_{ij}=\text{diag}[&(-i)^{\sigma_i+\tau_i(j)},(i)^{\sigma_i(2)},\ldots,(i)^{\sigma_i(\lambda(p)-1)},\\&(i)^{+1\sigma_i(\lambda(p))},(i)^{\sigma_i(\lambda(p)+1)},\ldots,(i)^{\sigma_i(r)},\\&(i)^{1+\sigma_i(r+1)},\ldots,(i)^{1+\sigma_i(\lambda(q)-1)},(i)^{2+\sigma_i(\lambda(q))},\\&(i)^{1+\sigma_i(\lambda(q)+1)},\ldots,(i)^{1+\sigma_i(\lambda(s)-1)},(i)^{1+\sigma_i\lambda(s))},\\&(i)^{1+\sigma_i(\lambda(s)+1)},\ldots,(i)^{1+\sigma_i(i)},(i)^{1+\sigma_i(i+1)},\\&\ldots,(i)^{1+\sigma_i(n)}]\end{aligned}\tag{22}$$

and p, q, s now take on the values

$$p=1,2,\ldots,k_i,$$

$$q=k_i+1,\ldots,\tau_i(j),$$

$$s=\tau_i(j)+1,\ldots,\bar{l}_i.$$

Note that these definitions imply that

$$d_{i\,i+1}=d_{i-12},\tag{23}$$

$$d_{n2}=f,$$

$$d_{i\,r+1}\text{ given by (21)}=d_{i\,r+1}\text{ given by (22)}.$$

(19) can now be written as

$$\begin{aligned}\alpha=&\bar{M}_{n2}(\theta_{n2})\bar{M}_{n3}(\theta_{n3})\cdots\bar{M}_{r+1\,r+1}(\theta_{r+1\,r+1})A\\&\times\bar{M}_{r2}(\theta_{r2})\cdots\bar{M}_{22}(\theta_{22}),\end{aligned}\tag{24}$$

where

$$\bar{M}_{ij}(\theta_{ij})=d_{ij}^{-1}M_{ij}(x_{ij})d_{ij+1},$$

$$r+1\leqslant i\leqslant n,\quad 2\leqslant j\leqslant i,\tag{25}$$

and

$$\begin{aligned}A&=d_{r+1\,r+2}^{-1}f=d_{r2}^{-1}f\\&=\text{diag}[(-1)^{(1/2)\sigma_r},(-1)^{(1/2)\sigma_r(2)},\ldots,(-1)^{(1/2)\sigma_r(n)}].\end{aligned}\tag{26}$$

We now show that (24) is the required factorization of $\alpha\in L(n,r)$, of the type mentioned in the Introduction, by proving the following theorem.

Theorem 3: (i) The $\frac{1}{2}n(n-1)$ angles

$$\theta_{ij},\quad 2\leqslant j\leqslant i\leqslant n,$$

are all real; we take them as the Euler angles of $\alpha\in L(n,r)$.

(ii) The corresponding matrices $\bar{M}_{ij}(\theta_{ij})$ are also real.

(iii) The form of $\bar{M}_{ij}(\theta_{ij})$ is such that each of the angles θ_{ij} is either of first or of second or of third kind.

(i)is clearly obvious (θ_{ij} are, by definition, all real) while (18) shows that

$$\bar{M}_{ij}(\theta_{ij}), \quad 2<j<i<r,$$

are all real and the corresponding angles θ_{ij} are all of first kind. We now fully establish (ii) and (iii) by considering the cases

$$i>r+1, \quad 2<j<r,$$

and

$$i>r+1, \quad r+1<j<i,$$

separately.

Case I: $i>(r+1)$, $2<j<r$: Here d_{ij} is given by (21). Suppose first that x_{ij} is not barred. Then

$$\tau_i(j)=\tau_i(j+1)\Rightarrow d_{ij}=d_{ij+1},$$

and also

$$(d_{ij})_j=(d_{ij+1})_j=(i)^{\sigma_i(j)},$$

where $(D)_j$ denotes the jth diagonal element of a diagonal matrix D. Hence, using (20) and (25), we get

$$\begin{aligned}(\bar{M}_{ij}(\theta_{ij}))_{11}&=(d_{ij}^{-1})_1(M_{ij}(x_{ij}))_{11}(d_{ij+1})_1\\&=(i)^{\sigma_i+\tau_i(j)}\cos((i)^{\rho_i(j)}\theta_{ij})\cdot(-i)^{\sigma_i+\tau_i(j)}\\&=\begin{cases}\cos\theta_{ij}, & \rho_i(j)\ \text{even},\\ \cosh\theta_{ij}, & \rho_i(j)\ \text{odd}.\end{cases}\end{aligned}$$

Similarly,

$$(\bar{M}_{ij}(\theta_{ij}))_{jj}=\begin{cases}\cos\theta_{ij}, & \rho_i(j)\ \text{even},\\ \cosh\theta_{ij}, & \rho_i(j)\ \text{odd}.\end{cases}$$

Also,

$$\begin{aligned}(\bar{M}_{ij}(\theta_{ij}))_{1j}&=(d_{ij}^{-1})_1(M_{ij}(x_{ij}))_{1j}(d_{ij+1})_j\\&=(i)^{\rho_i(j)}\cdot-\sin\left[(i)^{\rho_i(j)}\theta_{ij}\right]\\&=\begin{cases}-\sin\theta_{ij}, & \rho_i(j)\ \text{even},\\ \sinh\theta_{ij}, & \rho_i(j)\ \text{odd},\end{cases}\\(\bar{M}_{ij}(\theta_{ij}))_{j1}&=(-i)^{\rho_i(j)}\cdot\sin\left[(i)^{\rho_i(j)}\theta_{ij}\right]\\&=\begin{cases}\sin\theta_{ij}, & \rho_i(j)\ \text{even},\\ \sinh\theta_{ij}, & \rho_i(j)\ \text{odd}.\end{cases}\end{aligned}$$

In addition, the rest of elements of $\bar{M}_{ij}(\theta_{ij})$ are those of the unit $n\times n$ matrix. It follows that the matrix $\bar{M}_{ij}(\theta_{ij})$ is real and its form is such that

θ_{ij} is of first kind if $\rho_i(j)$ is even,

θ_{ij} is of second kind if $\rho_i(j)$ is odd.

Next, let x_{ij} be barred. Then

$$\begin{aligned}&j=\lambda(\tau_i(j)+1)\\&\Rightarrow\sigma_i(j)=\sigma_i(\lambda(\tau_i(j)+1))\equiv(\sigma_i+\tau_i(j)+1)\ \text{mod}\ 2\\&\Rightarrow\sigma_i(j)+1=\sigma_i+\sigma_i(j)+\tau_i(j)+1=\text{even}.\end{aligned}\tag{27}$$

Also here

$$\tau_i(j+1)=\tau_i(j)+1,$$

so that we get

$$\begin{aligned}&(d_{ij})_1=(-i)^{\sigma_i+\tau_i(j)}, \quad (d_{ij})_j=(i)^{\sigma_i(j)},\\&(d_{ij+1})_1=(-i)^{\sigma_i+\tau_i(j)+1}, \quad (d_{ij+1})_j=(i)^{1+\sigma_i(j)},\\&(d_{ij+1})_e=(d_{ij})_e, \quad e\neq 1,j.\end{aligned}$$

These give

$$\begin{aligned}&(\bar{M}_{ij}(\theta_{ij}))_{11}=(-i)\cdot i\cdot\sinh\theta_{ij}=\sinh\theta_{ij},\\&(\bar{M}_{ij}(\theta_{ij}))_{jj}=i\cdot i\cdot\sinh\theta_{ij}=-\sinh\theta_{ij},\\&(\bar{M}_{ij}(\theta_{ij}))_{1j}=(i)^{\rho_i(j)+1}\cdot-\cosh\theta_{ij}\\&\qquad=\begin{cases}-\cosh\theta_{ij}, & \rho_i(j)+1\ \text{divisible by 4},\\ \cosh\theta_{ij}, & \rho_i(j)+1\ \text{not divisible by 4},\end{cases}\\&(\bar{M}_{ij}(\theta_{ij}))_{j1}\\&\quad=(-i)^{\rho_i(j)+1}\cdot\cosh\theta_{ij}\\&\quad=\begin{cases}\cosh\theta_{ij}, & \rho_i(j)+1\ \text{divisible by 4},\\ -\cosh\theta_{ij}, & \rho_i(j)+1\ \text{not divisible by 4}.\end{cases}\end{aligned}$$

The rest of the elements of $\bar{M}_{ij}(\theta_{ij})$ are again those of the unit $n\times n$ matrix. Hence, $\bar{M}_{ij}(\theta_{ij})$ is again real and its form is such that θ_{ij} is of third kind. This proves (ii) and (iii) for the case under consideration.

Case II: $i>(r+1)$, $r<j<i$: Proceeding in exactly the same way as above, we find, as required, that $\bar{M}_{ij}(\theta_{ij})$ is again real, and its form is such that if x_{ij} is not barred,

θ_{ij} is of first kind for $\rho_i(j)$ odd,

θ_{ij} is of second kind for $\rho_i(j)$ even,

while if it is barred, θ_{ij} is of third kind. This completes the proof of the theorem.

Looking at the statements giving the kind of θ_{ij} in the above proof and keeping in mind part (iii) of Lemma 1, we see that θ_{ij} is of first or of second or of third kind if and only if the corresponding x_{ij} is real or pure imaginary or barred, respectively. It follows that the σ's may be related to θ_{ij} in exactly the same way as to x_{ij} if we replace "barred" by "of third kind"; as a result, the rule of Sec. III can be immediately transformed into one for writing down a set of real θ_{ij} which are Euler angles of a general element of $L(n,r)$.

In the end, we make the remark that our discussion of Euler angles of $L(n,r)$ is somewhat indirect in the sense that to get them for $\alpha\in L(n,r)$, we have first to form the matrix $\hat{\alpha}\in L(n,r)$, calculate its (complex) Euler angles x_{ij} and then from these obtain the θ_{ij}. Similarly, the kind of any given θ_{ij} [which determines the form of the corresponding matrix $\bar{M}_{ij}(\theta_{ij})$] is also obtained indirectly, from the reality nature of x_{ij}. However, in the absence of any alternative direct approach, we will have to be content with this indirect one.

[1]A. Syed, J. Math. Phys. **23**, 463 (1982), preceding paper.

The Clebsch–Gordan coefficients of the three-dimensional Lorentz algebra in the parabolic basis[a)]

Debabrata Basu[b)] and Kurt Bernardo Wolf
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS), Universidad Nacional Autónoma de México, México 20 D. F., Mexico

(Received 28 September 1981; accepted for publication 23 December 1981)

Starting from the oscillator representation of the three-dimensional Lorentz algebra so(2,1), we build a Lie algebra of second-order differential operators which realizes all series of self-adjoint irreducible representations. The choice of the common self-adjoint extension over a two-chart function space determines whether they lead to single- or multivalued representations over the corresponding Lie group. The diagonal operator defining the basis is the parabolic subgroup generator. The direct product of two such algebras allows for the calculation of all Clebsch–Gordan coefficients explicitly, as solutions of Schrödinger equations for Pöschl–Teller potentials over one ($\mathscr{D}\times\mathscr{D}$), two ($\mathscr{D}\times\mathscr{C}$), or three ($\mathscr{C}\times\mathscr{C}$) charts. All coefficients are given in terms of up to two ${}_2F_1$ hypergeometric functions.

PACS numbers: 02.20.Sv

I. INTRODUCTION

The reduction of the direct product of two self-adjoint irreducible representations (SAIR's) of a Lie algebra into a direct sum of SAIR's of the same algebra is a problem of fundamental importance in applications of both compact and noncompact algebras and groups in physics. In view of the Wigner–Eckart theorem, the keystone of many practical calculations, it can be considered among the central problems, on par with the classification and explicit construction of the SAIR matrices or integral kernels.

After the classification and construction of the unitary irreducible representations (UIR's) of the three-dimensional Lorentz group SO(2,1) was solved by Bargmann,[1] many authors considered the Clebsch–Gordan (CG) problem for the corresponding algebra so(2,1). This problem consists of two parts. The first is the determination of the CG series, and the second is the explicit evaluation of the Clebsch–Gordan coefficients (CGC's) which effect the reduction. The SAIR's of so(2,1) can be classified—following Bargmann's nomenclature—into discrete ($\mathscr{D}$) and continuous ($\mathscr{C}$) series, which in turn divide into various types: (positive and negative) discrete $D^{\pm}_{k}$, nonexceptional and exceptional continuous C^{ϵ}_{q} whose precise definition is given in the Appendix. The essentially distinct direct products to be considered are $D^{+}\times D^{+}$, $D^{+}\times D^{-}$, $\mathscr{D}\times\mathscr{C}$, and $\mathscr{C}\times\mathscr{C}$. The CG series for these four couplings changes nontrivially as one goes from case to case (they are listed in the Appendix). This structure, nevertheless, has an intrinsic meaning in that it does not depend on the particular choice of basis in the Hilbert space of representations. The CGC's, on the other hand, are manifestly basis-dependent.

The Clebsch–Gordan problem for so(2,1) was investigated by Pukánszky,[2] Holman and Biedenharn,[3,4] Ferretti and Verde,[5] Wang,[6] and by Mukunda and Radhakrishnan[7] among others.[8] Pukánszky confined his attention essentially to the structure of the CG series for the $\mathscr{C}\times\mathscr{C}$ couplings. He started from the realization of the UIR's of SO(2,1) in the Hilbert space of functions on the unit circle, decomposing their tensor product and restricting attention to the subspace for which the total magnetic quantum number is zero. He did not attempt the problem of explicit evaluation of the CGC's and of their orthonormalization. This aspect of the problem was considered by Holman and Biedenharn (HB) and by Wang. HB based their investigation on the fundamental recurrence relation satisfied by the CGC's in the compact subgroup basis. Their first paper[3] was mainly concerned with the coupling of two discrete-series representations and analytic continuation properties with the rotation group CGC's, while all coupling cases were considered in their second paper[4] in the same basis. The results are given in terms of ${}_3F_2$ generalized hypergeometric functions of unit argument. The problem is mathematically much simpler when the SAIR's coupled belong to one of the discrete series; the CGC's [in the elliptic so(2) basis] for this special coupling problem are well known, and their symmetry properties have been thoroughly investigated. The complexity of the problem progressively increases as one goes from the discrete to the continuous nonexceptional and to the exceptional types of SAIR's. Nonnormalized CGC's for the coupling of the continuous nonexceptional UIR's were derived by Ferretti and Verde.[5] Their method was based on the formula

$$d^{k_1}_{m_1'm_1}(z)d^{k_2}_{m_2'm_2}(z) = \mathop{\mathrm{S}}_{k} C\begin{pmatrix}k_1, & k_2; & k\\ m_1, & m_2; & m\end{pmatrix} C\begin{pmatrix}k_1, & k_2; & k\\ m_1', & m_2'; & m'\end{pmatrix}^{*} d^{k}_{m_1'+m_2',m_1+m_2}(z), \tag{1.1}$$

where the d's are Bargmann's SO(2,1) UIR functions,[1] the C's are the CGC's, and S denotes the summation over discrete and integration over the continuous series. The expansion is obtained by the use of the Burchnall–Chaundy[9] formula, followed by a Sommerfeld–Watson transformation.

[a)] Work performed under financial assistance from CONACYT, Project ICCBIND 790370.

[b)] On leave from Dept. of Physics, Indian Institute of Technology, Kharagpur, India.

 0022-2488/83/030478-23$02.50

The results of Ferretti and Verde formed the starting point of the investigations by Wang,[6] who attempted to orthonormalize the CGC's by adopting a summation prescription originally due to HB. [It should be noted that this apparently leads to a divergent expression in one case: Eq. (2.46) of Ref. 4 for $\mathscr{C}\times\mathscr{C}$; the convergence criterion for ${}_3F_2(1)$ is not fulfilled.]

All these authors used the maximal compact subalgebra basis for the evaluation of the CGC's. More recently, Mukunda and Radhakrishnan[7] have made a departure from this in evaluating the CG coefficients in the noncompact hyperbolic so(1,1) basis. A very attractive feature of their treatment is that they relate the CGC's with the UIR's of SO(4), SO(3,1), and SO(2,2) in various compact and noncompact bases, some of which require evaluation, and followed by Mellin transformation and composition. [Equations (5.4a), (5.4b), and (5.8c) for $\mathscr{D}\times\mathscr{C}$ in Ref. 7c and (5.8b) for $\mathscr{C}\times\mathscr{C}$ in Ref. 7d also exhibit divergent ${}_3F_2(1)$'s. These results cannot therefore be regarded as final.]

The purpose of the present paper is to give a comprehensive evaluation of the Clebsch–Gordan coefficients in the parabolic (or horocyclic) iso(1) basis, for all couplings. With the choice of this new basis the CGC's have especially simple forms: They are in general expressible in terms of single, real Gauss ${}_2F_1$ hypergeometric functions, or at most a sum of two of them

The plan of this article is as follows. We start in Sec. II reviewing the oscillator realization of the so(2,1) algebra[10] [Eqs. (2.1)]. This realization is unique in that it consists of second-order differential operators, rather than first-order ones, as Lie algebras of transformation groups do; as a result, the operator domain problem is not as trivial as it may appear at first sight. On $\mathscr{L}^2(R)$, the SAIR of this algebra exponentiates to an integral-transform UIR of the metaplectic group[11] Mp (2,R) [the fourfold covering of SO(2,1) or twofold covering of $SU(1,1)\approx SL(2,R)\approx Sp(2,R)$]. These are the linear canonical transforms.[12–14] In this paper we remain within the Lie algebra, however, so Secs. III and IV are devoted to building—through coupling—the SAIR's of the discrete and continuous nonexceptional series. This construction is important, even though it yields only the SAIR's of so(2,1) which exponentiate to single-valued UIR's of Sp(2,R) and does not include the exceptional continuous series, since it leads us to the consideration of $\mathscr{L}^2$-Hilbert spaces of functions on a space $\mathscr{S}=\{-1,+1\}\times R^+$, i.e., containing two R^+ charts (which can also be viewed as two-component vectors[14,15]). There, the so(2,1) generators have the form (4.14). In the $\mathscr{C}$-series they are Schrödinger Hamiltonians corresponding to harmonic or repulsive oscillators with a strong centripedal singularity at the origin. Singular potentials on a two-chart space is a feature not too familiar for physicists. Section V explores all common self-adjoint extensions of this algebra of operators in $\mathscr{L}^2(\mathscr{S})$, leading to all SAIR series of so(2,1), $\mathscr{D}$ as well as $\mathscr{C}$.

The distinct usefulness of the parabolic basis is that the defining subalgebra generator $J_-=J_0-J_1$ on $\mathscr{S}$ is $\sigma r^2/2$, $\sigma=\pm 1$ [Eq. (4.14e)], and its eigenfunctions are thus Dirac δ's on $\mathscr{S}$. The general CG problem of coupling two SAIR's is thus set up in Sec. VI as an eigenbasis problem in $\mathscr{L}^2(\mathscr{S}^2)$ where the coupled-basis Casimir operator is diagonal. The space $\mathscr{S}^2$ is parametrized into six charts, up to three of which are needed for any one coupling. In every chart, the Casimir operator eigenvalue equation takes the form of the Schrödinger equation for a Pöschl–Teller potential[16,17] of the first or of the second kind, with strong or weak barriers or wells at the end points. The solutions of these equations are given in terms of Gauss ${}_2F_1$ hypergeometric functions. When a coupling requires more than one chart, a cancellation of the boundary Wronskians must take place. The particulars of the computation of the CGC's for $D^+\times D^+$, $D^+\times D^-$, $\mathscr{D}\times\mathscr{C}$, and $\mathscr{C}\times\mathscr{C}$ couplings are given in Secs. VII, VIII, IX, and X, respectively. We omit inessential calculational labor, but give some details on the construction of multichart Kronecker- and Dirac-orthonormal solutions to the Casimir operator eigenfunction problem, which in Secs. IX and X has both point and continuous spectrum. Section XI, finally, presents some concluding remarks and offers some directions for further work.

Some convenient notational minutae: We denote by R the real line, by R^+ the interval $[0,\infty)$, by Z the integers, and by Z^+ the subset $\{0,1,2,\cdots\}$. In order to write compactly that some SAIR index k belongs to the continuous nonexceptional representation [i.e., $k=(1+i\kappa)/2$, $\kappa\in R^+$], we write $k\in C$, and to the continuous exceptional type [i.e., $k=(1+\kappa)/2$, $\kappa\in(0,1)$] we write $k\in E$. It is thereby understood that the multivaluation index ϵ (see the Appendix) ranges in the appropriate intervals, and $k\in\mathscr{C}$ is meant to stand for (k,ϵ). Similarly, $k\in D^\pm$ means that the representation k belongs to the upper- or lower-bound discrete series. If the latter distinction in unimportant, we write $k\in\mathscr{D}$ and imply $(k,\pm)$ when writing k. Effecting the product of a D^+ and C representations, and extracting from the direct sum a coupled D^+ representation is indicated as $D^+\times C\to D^+$. D^+_ρ stands for the lower-bound oscillator representation $D^+_{1/4}+D^+_{3/4}$. The Clebsch–Gordan coefficients in the first two sections are denoted as in (3.15) or (4.10); later, as the need for full generality arises, they are denoted as in (6.19).

II. THE OSCILLATOR REPRESENTATION

2.1: Consider the three formal differential operators in the real variable $x\in R$,

$$J_1^\rho=\frac{1}{4}\left(-\frac{d^2}{dx^2}-x^2\right), \tag{2.1a}$$

$$J_2^\rho=-\frac{i}{4}\left(x\frac{d}{dx}+\frac{1}{2}\right), \tag{2.1b}$$

$$J_0^\rho=\frac{1}{4}\left(-\frac{d^2}{dx^2}+x^2\right), \tag{2.1c}$$

and their linear combinations

$$J_+^\rho=J_0^\rho+J_1^\rho=-\frac{1}{2}\frac{d^2}{dx^2}, \tag{2.1d}$$

$$J_-^\rho=J_0^\rho-J_1^\rho=\tfrac{1}{2}x^2, \tag{2.1e}$$

among which we find the Schrödinger Hamiltonians for the harmonic (2.1c) and repulsive (2.1a) oscillators and the free particle (2.1d). The set of operators (2.1) exhibits the commutation relations of the generators of so(2,1):

$$[J_1, J_2] = -iJ_0, \quad [J_0, J_1] = iJ_2, \quad [J_2, J_0] = iJ_1. \tag{2.1f}$$

The Casimir operator is a multiple of the identity:

$$Q^{\circ} = (J_1^{\circ})^2 + (J_2^{\circ})^2 - (J_0^{\circ})^2 = \tfrac{3}{16}\mathbf{1}. \tag{2.1g}$$

The generators (2.1) are also known to have a unique common self-adjoint extension[18] in the Hilbert space $\mathscr{L}^2(R)$, whose defining inner product is

$$(f,g)_R = \int_{-\infty}^{\infty} dx\, f(x)^* g(x). \tag{2.2}$$

2.2: The realization (2.1) of so(2,1) in $\mathscr{L}^2(R)$ is called the positive oscillator[10] representation $D^+_{\circ}$. It is not an irreducible representation, but in fact consists of the direct sum of two SAIR's, $D^+_{1/4}$ and $D^+_{3/4}$. The Bargmann eigenvalue of (2.2), $q = k(1-k)$, has the value $\frac{3}{16}$ for both. In order to distinguish between the two direct summands, we may use the eigenvalues ι of the inversion operator $I: f(x) = f(-x) = \iota f(x)$. The even (odd) harmonic oscillator wavefunctions belong to the eigenvalues $\mu = \epsilon + n$, $n \in Z^+$, $\epsilon = (2-\iota)/4$, under J_0°, for $k = \frac{1}{4}$ ($k = \frac{3}{4}$). These functions form a complete orthonormal basis for $\mathscr{L}^2_{\iota}(R)$, the space of $\mathscr{L}^2(R)$ functions with that parity. They also define common self-adjoint extensions of the algebra operators (2.1) in $\mathscr{L}^2(R^+)$ with boundary conditions given by the vanishing of the function or of its derivative, respectively, at the origin.

2.3: The operator whose eigenbasis we want to exploit in this paper is J_-, given by (2.1e). Such a Dirac-orthonormal generalized eigenbasis, complete for $\mathscr{L}^2(R)$, is $\{\delta(x-\rho), \rho \in R\}$, with eigenvalues $\rho^2/2 \in R^+$. The spectrum of J°_- thus covers R^+ twice. The decomposition according to the irreducible components of $D^+_{\circ}$ is easily accomplished through demanding definite parity. Thus

$$\psi^{1/4}_{\rho}(x) = (2\rho)^{-1/2}[\delta(x-\rho) + \delta(x+\rho)], \tag{2.3a}$$

$$\psi^{3/4}_{\rho}(x) = (2\rho)^{-1/2}[\delta(x-\rho) - \delta(x+\rho)], \quad x \in R, \ \rho \in R^+, \tag{2.3b}$$

constitute Dirac bases for even ($\iota = +1$) and odd ($\iota = -1$) functions, respectively. They have been chosen orthonormal in the Dirac sense:

$$(\psi^k_{\rho_1}, \psi^k_{\rho_2})_R = \rho_1^{-1}\delta(\rho_1 - \rho_2) = \delta(\rho_1^2/2 - \rho_2^2/2), \tag{2.4a}$$

for k either $\frac{1}{4}$ or $\frac{3}{4}$. Orthogonality in the upper index also holds, but will not be used. They are complete in the sense

$$\int_0^{\infty} d(\rho^2/2)\psi^k_{\rho}(x)^* \psi^k_{\rho}(x') = \delta(x - x'), \tag{2.4b}$$

where the right-hand side is the reproducing kernel in the space of the corresponding parity ι. The operators (2.1) with these domains will be indicated, respectively, by $J^{1/4}_{\alpha}$ and $J^{3/4}_{\alpha}$ ($\alpha = 1,2,0,+,-$). In particular, the generalized spectrum of J^k_-, $k = \frac{1}{4}$ or $\frac{3}{4}$, is now simple: A single eigenfunction $\psi^k_{\rho}(x)$ corresponds to each eigenvalue $\rho^2/2 \in R^+$, $\rho \in R^+$.

2.4: The negative oscillator representation $D^-_{\circ}$ of so(2,1) is obtained from the $D^+_{\circ}$ representation seen above by means of the outer automorphism of the algebra $A: J^{\circ}_{\alpha} \mapsto J^{\circ}_{\alpha}{}^-$, where

$$J_1^{k-} = -J_1^k, \quad J_2^{k-} = J_2^k,$$
$$J_0^{k-} = -J_0^k, \quad J_{\pm}^{k-} = -J_{\pm}^k, \tag{2.5a}$$

for $k = \circ$, or $\frac{1}{4}, \frac{3}{4}$ after decomposition into its irreducible components. The Casimir operator does not distinguish between positive and negative representations:

$$Q^{k-} = Q^k = Q^{\circ} = \tfrac{3}{16}\mathbf{1} \quad (k = \tfrac{1}{4}, \tfrac{3}{4}). \tag{2.5b}$$

In this way we produce the $D^-_{1/4}$ and $D^-_{3/4}$ SAIR's out of the D^+_k ones. The spectrum of J^k_0 changes sign under the automorphism A, and so does the spectrum of J^k_- which is now $-\rho^2/2 \in R^-, \rho \in R^+$, still simple, and with generalized eigenfunctions formally identical to (2.3a) and (2.3b).

III. THE OSCILLATOR COUPLING TO THE DISCRETE SERIES

In this section we consider the direct product of two irreducible components of the oscillator representation, decomposed into a direct sum of SAIR's D^+_k belonging to the positive discrete series.

3.1: Consider the two sets of so(2,1) generators, mutually commuting, given as in (2.1), in terms of two independent variables x_j, $j = 1,2$, and denote them by $J^{k_j}_{(j)\alpha}$ ($j = 1,2$, respectively, $\alpha = 1,2,0,+,-$; $k_j = \frac{1}{4}$ in the space $\mathscr{L}^2_{+1}(R)$ of even functions in x_j, and $k_j = \frac{3}{4}$ in the space $\mathscr{L}^2_{-1}(R)$ of odd functions in x_j). Out of these we build the two-variable operators

$$J^{k_1,k_2}_{\alpha} = J^{k_1}_{(1)\alpha} + J^{k_2}_{(2)\alpha}. \tag{3.1}$$

These operators will have a correspondingly unique self-adjoint extension in the Hilbert space $\mathscr{L}^2_{\iota_1,\iota_2}(R^2) = \mathscr{L}^2_{\iota_1}(R) \times \mathscr{L}^2_{\iota_2}(R)$, with inner product

$$(f,g)_{R^2} = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 f(x_1,x_2)^* g(x_1,x_2) = \int_0^{\infty} r\,dr \int_{-\pi}^{\pi} d\theta\, f[r,\theta]^* g[r,\theta], \tag{3.2}$$

where we have introduced polar coordinates in the plane:

$$x_1 = r\cos\theta, \quad x_2 = r\sin\theta, \quad r \in R^+, \quad \theta \equiv \theta \pmod{2\pi}. \tag{3.3}$$

All functions $f(x_1,x_2)$ in the domain of J^{k_1,k_2}_{α} have parity ι_j under inversions $x_j \leftrightarrow -x_j$ when $k_j = \frac{1}{4}, \frac{3}{4}$ as above. The Casimir operator (2.2) corresponding to the factor operator set $J^{k_j}_{(j)\alpha}$ will be denoted by $Q^{k_j}_{(j)}$, and that of the coupled set J^{k_1,k_2}_{α} will be denoted by Q^{k_1,k_2}.

3.2: We shall define the product states as the generalized eigenfunctions of the four commuting operators

$$Q^{k_1}_{(1)}, \quad Q^{k_2}_{(2)}, \quad J^{k_1}_{(1)-}, \quad J^{k_2}_{(2)-}. \tag{3.4}$$

The first two operators are here identically $\frac{3}{16}\mathbf{1}$, while the second pair determine the product states to be

$$\psi^{k_1,k_2}_{\rho_1,\rho_2}(x_1,x_2) = \psi^{k_1}_{\rho_1}(x_1)\psi^{k_2}_{\rho_2}(x_2) \tag{3.5}$$

in terms of (2.3), with eigenvalues $\rho_1^2/2$ and $\rho_2^2/2$, respectively. From (2.4) they are Dirac-orthonormal and complete:

$$(\psi^{k_1,k_2}_{\rho_1,\rho_2}, \psi^{k_1,k_2}_{\rho_1',\rho_2'})_{R^2} = \delta(\rho_1^2/2 - \rho_1'^2/2)\delta(\rho_2^2/2 - \rho_2'^2/2). \tag{3.6}$$

Orthogonality for the indices k_1, k_2 also holds, but will not be

needed.

3.3: The coupled states are defined as the generalized eigenfunctions of the four commuting operators

$$Q_{(1)}^{k_1}, \quad Q_{(2)}^{k_2}, \quad Q^{k_1,k_2}, \quad J_{-}^{k_1,k_2}, \tag{3.7}$$

in the same space $\mathscr{L}^2_{\iota_1,\iota_2}(R^2)$. Again, the first two operators are $\frac{3}{16}\mathbf{1}$ while the last one may be written in terms of the polar coordinates (3.3) as

$$J_{-}^{k_1,k_2} = J_{(1)-}^{k_1} + J_{(2)-}^{k_2} = x_1^2/2 + x_2^2/2 = r^2/2. \tag{3.8}$$

The generalized eigenfunctions of (3.7) will therefore include a factor $\delta(r^2/2 - \rho^2/2)$ so that they be Dirac-orthogonal eigenfunctions of $J_{-}^{k_1,k_2}$ under (3.2), with eigenvalue $\rho^2/2$. This, in turn, will produce a selection rule $\rho_1^2/2 + \rho_2^2/2 = \rho^2/2$ in the CGC. Finally, the coupled Casimir operator Q^{k_1,k_2} may be written in terms of (2.1), (2.2), (3.1), and (3.3) as

$$\begin{aligned} Q^{k_1,k_2} &= Q_{(1)}^{k_1} + Q_{(2)}^{k_2} + 2(J_{(1)1}^{k_1} J_{(2)1}^{k_2} + J_{(1)2}^{k_1} J_{(2)2}^{k_2} - J_{(1)0}^{k_1} J_{(2)0}^{k_2}) \\ &= \frac{1}{4} + \frac{1}{4}\frac{\partial^2}{\partial\theta^2}. \end{aligned} \tag{3.9}$$

Since the spectrum of $\partial^2/\partial\theta^2$ on $\mathscr{L}^2(S_1)$ (the Hilbert space of square-integrable functions on the circle) is negative (and discrete), we are assured of having coupled representations belonging to the discrete series. Whereas the product states separate in Cartesian coordinates, the coupled states separate in polar coordinates, and consequently may be written as

$$\Psi_{k,\rho}^{k_1,k_2}(r,\theta) = \delta(r^2/2 - \rho^2/2) F_k^{k_1,k_2}(\theta), \tag{3.10}$$

with $F_k^{k_1,k_2}(\theta)$ a solution of

$$Q^{k_1,k_2} F_k^{k_1,k_2}(\theta) = k(1-k) F_k^{k_1,k_2}(\theta), \tag{3.11}$$

properly normalized in $\mathscr{L}^2(S_1)$ and such that it satisfies the parity properties of the space of functions.

The θ-dependent factors $F_k^{k_1,k_2}(\theta)$ of the coupled states must be of the form $\exp(im\theta)$ with m integer [so that it is periodic] and $k(1-k) = (1-m^2)/4$, i.e., $k = (1 + |m|)/2$. As functions of $\theta = \arctan(x_2/x_1)$ they must be even (odd) under $x_1 \leftrightarrow -x_1$, $\theta \leftrightarrow \pi - \theta$ when $k_1 = \frac{1}{4}$ ($k_1 = \frac{3}{4}$) and even (odd) under $x_2 \leftrightarrow -x_2$, $\theta \leftrightarrow -\theta$. These requirements are met, as a selection rule, when performing the inner product with the product states (3.5), but it is illuminating to impose them on the coupled state. What they imply is that: (i) The $(k_1,k_2) = (\frac{1}{4},\frac{1}{4})$ coupling may only contain $\cos m\theta$ with m even; (ii) the $(\frac{3}{4},\frac{3}{4})$ coupling, only $\sin m\theta$ with m even (excluding thus $m = 0$); (iii) the $(\frac{1}{4},\frac{3}{4})$ and $(\frac{3}{4},\frac{1}{4})$ couplings, only $\sin m\theta$ and $\cos m\theta$, respectively, with m odd. An elementary analysis involving the J_0 eigenvalues will confirm these selection rules. The choice of phases for the product and coupled states has been made out of simplicity.

3.4: The normalized coupled states are thus built out of (3.10) with

$$F_k^{1/4,1/4}(\theta) = \begin{cases} (2\pi)^{-1/2}, & k = \frac{1}{2}, \\ \pi^{-1/2}\cos(2k-1)\theta, & k = \frac{3}{2},\frac{5}{2},\cdots, \end{cases} \tag{3.12a}$$

$$F_k^{3/4,3/4}(\theta) = \pi^{-1/2}\sin(2k-1)\theta, \quad k = \tfrac{3}{2},\tfrac{5}{2},\cdots, \tag{3.12b}$$

$$F_k^{1/4,3/4}(\theta) = \pi^{-1/2}\sin(2k-1)\theta, \quad k = 1,2,\cdots, \tag{3.12c}$$

$$F_k^{3/4,1/4}(\theta) = \pi^{-1/2}\cos(2k-1)\theta, \quad k = 1,2,\cdots, \tag{3.12d}$$

and satisfy

$$(\Psi_{k,\rho}^{k_1,k_2}, \Psi_{k',\rho'}^{k_1,k_2})_{R^2} = \delta_{k,k'}\delta(\rho^2/2 - \rho'^2/2). \tag{3.13}$$

Again, orthogonality in k_1 and k_2 indices is present but will not be used.

3.5: The normalization coefficient $\pi^{-1/2}$ in (3.12) is trivial to obtain from integral tables or through a number of elementary analyses. We would like to present here, however, the method which will be used, especially in Secs. IX and X, for more complicated point-spectrum eigenfunctions. Write $F_k(\theta) = c_k f_k(\theta)$ and $F_l(\theta) = c_l f_l(\theta)$, eigenfunctions of Q in (3.9) with eigenvalues $k(1-k)$ and $l(1-l)$, respectively, $f_k(\theta) = \sin(2k-1)\theta$ or $\cos(2k-1)\theta$, and c_k to be determined. Then, the integral over any interval $(a,b) \subset R$ may be put, through multiplying the eigenvalue equation of f_k^* by f_l and vice versa, and taking the difference, in terms of the boundary Wronskians as

$$\begin{aligned} \int_a^b d\theta f_k^* f_l &= [4(l+k-1)(l-k)]^{-1} W(f_k^*, f_l)\,|_a^b \\ &= [4(l+k-1)(l-k)]^{-1}[f_k^{*\prime} f_l - f_k^* f_l']\,|_b^a, \end{aligned} \tag{3.14a}$$

where the prime indicates differentiation with respect to the argument θ. The boundary term on the right-hand side vanishes when $b - a = 2\pi$ and $k = l$ + integer, assuring the orthogonality of any eigenfunction pair. Now, Eq. (3.14a) is true even when f_l (and/or f_k) is not a periodic function in θ under the period $b - a$, i.e., when $2l - 1$ is not an integer. For the normalization problem at hand, we may choose for a a point where the boundary term is zero, such as $a = 0$ and for b some other point where the term is easily evaluated. We thus write

$$\begin{aligned} \int_0^{2\pi} d\theta\, |f_k(\theta)|^2 &= 2\lim_{l\to k}\int_0^{\pi} d\theta f_k^*(\theta) f_l(\theta) \\ &= [2(2k-1)]^{-1}\lim_{(l\to k)}(l-k)^{-1} W(f_k^*, f_l)|_{\theta=\pi} \\ &= [2(2k-1)]^{-1}\frac{\partial}{\partial l} W(f_k^*, f_l)\,|_{\theta=\pi, l=k}. \end{aligned} \tag{3.14b}$$

In the cases exemplified here, the limits and valuations may be manifestly exchanged, and the result is π; it follows that $c_k = \pi^{-1/2}$, as given in (3.12). This method will be used when f_k and f_l are given hypergeometric functions, whose normalization constants are not easily evaluated by other means.

3.6: The inner products between the product and coupled states constitute the standard definition of the Clebsch–Gordan coefficients:

$$\begin{aligned} C\begin{pmatrix} k_1,k_2;k \\ \rho_1,\rho_2;\rho \end{pmatrix} &= (\psi_{\rho_1,\rho_2}^{k_1,k_2}, \Psi_{k,\rho}^{k_1,k_2})_{R^2} \\ &= 2(\rho_1\rho_2)^{-1/2}\int_0^{\pi/2} d\theta\, \delta(\rho\cos\theta - \rho_1) \\ &\quad \times \delta(\rho\sin\theta - \rho_2) F_k^{k_1,k_2}(\theta) \\ &= 2(\rho_1\rho_2)^{-1/2}\delta(\rho_1^2/2 + \rho_2^2/2 - \rho^2/2) \\ &\quad \times F_k^{k_1,k_2}(\arctan(\rho_2/\rho_1)). \end{aligned} \tag{3.15}$$

We have used the matching symmetry properties in θ of the product and coupled states to reduce the θ-integration to the first quadrant and finally express the results in terms of the solutions (3.12) to (3.11).

Noting that

$$\arctan(\rho_2/\rho_1) = \arcsin(\rho_2/\rho) = \arccos(\rho_1/\rho), \quad (3.16)$$

we can write the expressions in terms of Chebyshev polynomials of the first and second kinds (Ref. 19, Eq. 22.3.15). The CGC for the $D_{\rho}^{+} \times D_{\rho}^{+} \to D_{k}^{+}$ couplings are thus

$$C\begin{pmatrix} \frac{1}{4}, & \frac{1}{4}; & \frac{1}{2} \\ \rho_1, & \rho_2; & \rho \end{pmatrix} = \delta(\rho_1^2/2 + \rho_2^2/2 - \rho^2/2)(\tfrac{1}{2}\pi\rho_1\rho_2)^{-1/2}, \quad (3.17a)$$

$$C\begin{pmatrix} \frac{1}{4}, & \frac{1}{4}; & k \\ \rho_1, & \rho_2; & \rho \end{pmatrix} = \delta(\rho_1^2/2 + \rho_2^2/2 - \rho^2/2)\pi^{-1/2}\rho_1^{-1/2}\rho_2^{-1/2} \times T_{2k-1}(\rho_1/\rho), \quad k = \tfrac{3}{2},\tfrac{5}{2},\cdots, \quad (3.17b)$$

$$C\begin{pmatrix} \frac{3}{4}, & \frac{3}{4}; & k \\ \rho_1, & \rho_2; & \rho \end{pmatrix} = \delta(\rho_1^2/2 + \rho_2^2/2 - \rho^2/2)2\pi^{-1/2}\rho_1^{-1/2}\rho_2^{1/2}\rho^{-1} \times U_{2k-2}(\rho_1/\rho), \quad k = \tfrac{3}{2},\tfrac{5}{2},\cdots, \quad (3.17c)$$

$$C\begin{pmatrix} \frac{1}{4}, & \frac{3}{4}; & k \\ \rho_1, & \rho_2; & \rho \end{pmatrix} = C\begin{pmatrix} \frac{3}{4}, & \frac{3}{4}; & k \\ \rho_1, & \rho_2; & \rho \end{pmatrix}, \quad k = 1,2,\cdots, \quad (3.17d)$$

$$C\begin{pmatrix} \frac{3}{4}, & \frac{1}{4}; & k \\ \rho_1, & \rho_2; & \rho \end{pmatrix} = C\begin{pmatrix} \frac{1}{4}, & \frac{1}{4}; & k \\ \rho_1, & \rho_2; & \rho \end{pmatrix}, \quad k = 1,2,\cdots. \quad (3.17e)$$

$$\cos(\nu \arcsin z) = F(-\nu/2,\nu/2;\tfrac{1}{2};z^2), \quad |\mathrm{Re}\, z| < 1, \quad (3.18a)$$

$$\sin(\nu \arcsin z) = \nu z(1-z^2)^{1/2}F(-\nu/2+1,\nu/2+1;\tfrac{3}{2};z^2), \quad (3.18b)$$

we can write the CGC's for $D_{\rho}^{+} \times D_{\rho}^{+} \to D_{k}^{+}$ in terms of hypergeometric functions ${}_2F_1$.

Through application of the outer automorphism A to both product and coupled states, we find the same values for the $D_{\rho}^{-} \times D_{\rho}^{-} \to D_{k}^{-}$ CGC's as those given in (3.17) with the simple changes $k \to k^{-}$. In the parabolic basis we do not have to observe the Bargmann phase convention [Ref. 1, Eqs. (6.23) and (7.10)] which introduces a factor of $(-1)^m$ to the so(2)-classified eigenbasis.

3.7: The description of the CG series and the evaluation of the coefficients, as presented in this section, could have been made also using the completeness relation of the Fourier series on $\mathscr{L}^2(S_1)$ with the proper parities, thereby expanding the product of Dirac δ's on the plane (3.5) in terms of a Dirac δ on the radius, times the limit of the Dirichlet kernel. This method has been implemented before for the integral transform group in projecting out the k-radial canonical transform[3b] belonging to the D_{k}^{+} UIR, out of the $\mathrm{Mp}(2,R) \times \mathrm{Mp}(2,R)$ oscillator integral transform UIR. Here, we have not left the algebra level. The completeness statements for the Pöschl–Teller Schrödinger equation solutions—to be seen in Secs. VI–X—are more difficult to make than those for Fourier series.

3.8: As a preparation for Secs. V–X, we now project the coupled operators (3.1) on a definite D_{k}^{+} SAIR space ($k = \frac{1}{2},1,\frac{3}{2},2,\cdots$) using the fact that in this space the coupled Casimir operator Q^{k_1,k_2} has a definite eigenvalue $k(1-k)$. We perform a similarity transformation by $r^{1/2}$ so that the resulting operator be manifestly symmetric in $\mathscr{L}^2(R^{+})$ with inner product

$$(f,g)_{R^{+}} = \int_0^\infty dr\, f(r)^* g(r). \quad (3.19)$$

Thus

$$r^{1/2}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}\right)r^{-1/2} = r^{1/2}\left(\frac{\partial^2}{\partial r^2} + r^{-1}\frac{\partial}{\partial r} + r^{-2}[4Q^{k_1,k_2} - 1]\right)r^{-1/2} = \frac{\partial^2}{\partial r^2} - \frac{\gamma}{r^2}, \quad \gamma = (2k-1)^2 - \tfrac{1}{4} \geqslant -\tfrac{1}{4}. \quad (3.20)$$

We obtain hence for the $D_{k}^{\pm}$ SAIR

$$J_1^{k\pm} = \pm\frac{1}{4}\left(-\frac{d^2}{dr^2} + \frac{\gamma}{r^2} - r^2\right), \quad (3.21a)$$

$$J_2^{k\pm} = -\frac{i}{2}\left(r\frac{d}{dr} + \frac{1}{2}\right), \quad (3.21b)$$

$$J_0^{k\pm} = \pm\frac{1}{4}\left(-\frac{d^2}{dr^2} + \frac{\gamma}{r^2} + r^2\right), \quad (3.21c)$$

$$J_{+}^{k\pm} = J_0^{k\pm} + J_1^{k\pm} = \pm\frac{1}{2}\left(-\frac{d^2}{dr^2} + \frac{\gamma}{r^2}\right), \quad (3.21d)$$

$$J_{-}^{k\pm} = J_0^{k\pm} - J_1^{k\pm} = \pm\tfrac{1}{2}r^2, \quad (3.21e)$$

$$k = (1 + [\gamma + \tfrac{1}{4}])/2 \in \{\tfrac{1}{2},1,\tfrac{3}{2},2,\cdots\}. \quad (3.21f)$$

Of course, for $\gamma = 0$ we are formally back at (2.1), although k is from here on taking a range of values which excludes this case.

The Casimir operator of the so(2,1) algebra (3.21), calculated as in (2.2), yields here

$$Q^k = q\mathbf{1}, \quad q = k(1-k) = -\gamma/4 + \tfrac{3}{16} \leqslant -\tfrac{1}{4}. \quad (3.22)$$

In the next section we shall describe the domains where the operators (3.21) have self-adjoint extensions. Here we would only like to point out that if I_1 and I_2 are the inversion operators along the two axes, the space $\mathscr{L}^2_{\iota_1}(R) \times \mathscr{L}^2_{\iota_2}(R)$ will have eigenvalue $\iota = \iota_1\iota_2$ under the product inversion $I = I_1I_2$. Under the projection to a definite $Q^{k_1k_2}$ eigenvalue in (3.20), this information is lost. We may keep, however, the J_0-spectrum starting point $\epsilon_j = (2-\iota_j)/4$ defined in subsection 2.2, in order to associate with the domain of (3.21) the index $\epsilon = \epsilon_1 + \epsilon_2 = 1 - (\iota_1 + \iota_2)/4$. The spectrum of $J_0^{k\pm}$ will be contained in $\{\epsilon + n, n \in Z\}$. It is immediate to see that $\epsilon = 0$ when k is integer, and $\epsilon = \frac{1}{2}$ when k is half-integer. In fact, thus $\epsilon \equiv k \pmod 1$.

IV. THE OSCILLATOR COUPLING TO THE CONTINUOUS SERIES

We consider now the coupling of a positive and a negative oscillator representation, and their decomposition into a direct integral of SAIR's belonging to the nonexceptional continuous series.

4.1: Consider two sets of generators, $J_{(1)\alpha}^{k_1+}$ and $J_{(2)\alpha}^{k_2-}$, the first belonging to the $D_{k_1}^{+}$ summand ($k_1 = \frac{1}{4}$ or $\frac{3}{4}$) of the positive oscillator representation and the second set, obtained through (2.5) from the first, belonging to the $D_{k_2}^{-}$ negative oscillator representation. Out of these we now build, *vis-à-vis* (3.1), the two-variable operators

$$J_{\alpha}^{k_1,k_2} = J_{(1)\alpha}^{k_1+} + J_{(2)\alpha}^{k_1-} = J_{(1)\alpha}^{k_1} + \tau_\alpha J_{(2)\alpha}^{k_2}, \quad (4.1a)$$

$$\tau_\alpha = -1 \quad \text{for } \alpha = 1,0,+,-, \quad \tau_\alpha = 1 \quad \text{for } \alpha = 2, \quad (4.1b)$$

also self-adjoint with respect to the inner product $(\cdot,\cdot)_{R^2}$ in

(3.2), in the space of functions $\mathscr{L}^2_{\iota_1,\iota_2}(R^2) = \mathscr{L}^2_{\iota_1}(R) \times \mathscr{L}^2_{\iota_2}(R)$.

In particular, the diagonal operator $J_-^{k_1,k_2}$ is the difference, rather than the sum of $J^{k_1}_{(1)-}$ and $J^{k_2}_{(2)-}$. It is thus convenient to parametrize the R^2 plane in hyperbolic coordinates (σ,r,θ), $\sigma\in\{-1,+1\}$, $r\in R$, $\theta\in R$ given by

$$\text{for } |x_1|>|x_2|: \quad \sigma=+1, \quad x_1 = r\cosh\theta, \quad x_2 = r\sinh\theta, \tag{4.2a}$$

$$\text{for } |x_1|<|x_2|: \quad \sigma=-1, \quad x_1 = r\sinh\theta, \quad x_2 = r\cosh\theta, \tag{4.2b}$$

disregarding the cone $|x_1| = |x_2|$. (The fact that we are using in this section the letters r and θ to denote the hyperbolic radius and angle should lead to no confusion with the former section, where they stood for the polar radius and angle. The ensuing uniformity of notation—with the ranges specified—will be seen to be quite economical.) Note that each value of σ in (4.2) defines one coordinate chart, and both $\sigma=+1$ and $\sigma=-1$ charts are needed to cover R^2.

We may express, correspondingly, the $(\cdot,\cdot)_{R^2}$ inner product (3.2) through

$$(f,g)_{R^2} = \sum_{\sigma=\pm1}\int_{-\infty}^{\infty}|r|dr\int_{-\infty}^{\infty}d\theta f\{\sigma,r,\theta\}^* g\{\sigma,r,\theta\}, \tag{4.3}$$

where $f\{\sigma,r,\theta\} = f(x_1,x_2)$ uses the hyperbolic parametrization. We may think of $f(r,\theta)$ as a two-component function with components labeled by $\sigma=\pm1$ or included them—as we do here—as a coordinate.

4.2: As before, we define the product states as the generalized eigenfunctions of $Q^{k_1}_{(1)}$, $Q^{k_2}_{(2)}$, $J^{k_1}_{(1)-}$ and $J^{k_2}_{(2)-}$ in $\mathscr{L}^2_{\iota_1,\iota_2}(R^2)$ as in (3.5)–(2.3) with the difference that their eigenvalue under the latter operator will be $-\rho_2^2/2$, $\rho_2\in R^+$, instead of $+\rho_2^2/2$. When $\rho_1>\rho_2$ the product state is nonzero only over the $\sigma=+1$ chart, while when $\rho_1<\rho_2$, it is nonzero over the $\sigma=-1$ chart. The orthogonality relation (3.6) continues to hold.

4.3: The coupled states, again, are defined as the Dirac-normalized generalized eigenfunctions of $Q^{k_1}_{(1)}$, $Q^{k_2}_{(2)}$, Q^{k_1,k_2} and $J_-^{k_1,k_2}$ in $\mathscr{L}^2_{\iota_1,\iota_2}(R^2)$, as in (3.7) *et seq.*, except that we find it best to express them in terms of hyperbolic coordinates (4.2). Corresponding to (3.8), we have the coupled generator

$$J_-^{k_1,k_2} = J^{k_1}_{(1)-} - J^{k_2}_{(2)-} = x_1^2/2 - x_2^2/2 = \sigma r^2/2, \tag{4.4}$$

the change of sign due to (2.5). The essential spectrum of (4.4) thus covers R. Corresponding to (3.9), we have

$$\begin{aligned} Q^{k_1,k_2} &= Q^{k_1}_{(1)} + Q^{k_2}_{(2)} \\ &\quad + 2(-J^{k_1}_{(1)1}J^{k_2}_{(2)1} + J^{k_1}_{(1)2}J^{k_2}_{(2)2} + J^{k_1}_{(1)0}J^{k_2}_{(2)0}) \\ &= \frac{1}{4} - \frac{1}{4}\frac{\partial^2}{\partial\theta^2}. \end{aligned} \tag{4.5}$$

Since the spectrum $\partial^2/\partial\theta^2$ on $\mathscr{L}^2(R)$ is negative, we are assured of having coupled representations in the nonexceptional continuous series (the point $k=\frac{1}{2}$ will be subject to further discussion). The coupled states separate in hyperbolic coordinates and may be written as

$$\begin{aligned} \Psi^{k_1+,k_2-}_{k,\tau,\rho}(\sigma,r,\theta) &= \delta(\sigma r^2/2 - \tau\rho^2/2)E^{k_1,k_2}(\operatorname{sgn} r)F^{k_1+,k_2-}_{k,\tau}(\theta) \\ &= \delta_{\sigma,\tau}\delta(r^2/2-\rho^2/2)E^{k_1,k_2}(\operatorname{sgn} r)F^{k_1+,k_2-}_{k,\tau}(\theta), \end{aligned} \tag{4.6}$$

which is generalized eigenfunction of $J_-^{k_1,k_2}$ with eigenvalue $\tau\rho^2/2\in R$ and of Q^{k_1,k_2} with eigenvalue $k(1-k)\geqslant\frac{1}{4}$ provided (3.11) holds (with $\theta\in R$ now). This last equation for $F^{k_1+,k_2-}_{k,\tau}(\theta)$ has solutions $\exp(i\kappa\theta)$ for $\kappa\in R$, corresponding to $k(1-k) = (1+\kappa^2)/4$, i.e., $k = (1+i\kappa)/2$. The representations of the continuous series with κ and $-\kappa$ (k and $1-k$) are equivalent[1] so we shall restrict ourselves to $\kappa\in R^+$. Note that $\Psi^{k_1+,k_2-}_{k,\tau,\rho}(\sigma,r,\theta)$ is zero for $\sigma=-1$ when its eigenvalue under $J_-^{k_1,k_2}$ is positive, and is zero for $\sigma=+1$ when its eigenvalue is negative.

We still have to see it, explicitly, that the parities of the product states are the same as those of the coupled states. (i) The $(k_1,k_2) = (\frac{1}{4},\frac{1}{4})$ product states are even under $x_1\leftrightarrow -x_1$ and under $x_2\leftrightarrow -x_2$, i.e., even under $r\leftrightarrow -r$ and even under $\theta\leftrightarrow -\theta$, implying $E^{1/4,1/4}(-1)$, and that only a term $\cos k\theta$ may be present in $F^{k_1+,k_1-}_{\tau,k}(\theta)$. (ii) the $(\frac{3}{4},\frac{3}{4})$ product states are odd in both x_1 and x_2, hence even in r and odd in θ, so $E^{3/4,3/4}(-1)$ and only a term $\sin\kappa\theta$ may be present in (4.6). (iii) $(\frac{1}{4},\frac{3}{4})$ product states are even in x_1 and odd in x_2; for $\sigma=-1$ they are odd in r and even in θ. (iv) $(\frac{3}{4},\frac{1}{4})$ states, odd in x_1 and even in x_2 are, for $\sigma=+1$, odd in r and even in θ, while for $\sigma=-1$ they are odd in both r and θ. In all of these cases, the spectrum of J_0 is contained in $\{\epsilon+n,\ n\in Z\}$, $\epsilon = \epsilon_1-\epsilon_2 = (\iota_2-\iota_1)/4$.

4.4: To sum up, cases (i) and (ii) (which from the J_0 eigenvalue content we know belong to the C^0_q representations) are even functions of r, and cases (iii) and (iv) (belonging to the $C^{1/2}_q$ representations), are odd in r. We may hence write $E^{k_1,k_2}(\operatorname{sgn} r) = (\operatorname{sgn} r)^{2\epsilon}$, $\epsilon = |k_1-k_2|$, cut the range of r to R^+ and henceforth attach the index ϵ in front of the representation index $k = (1+i\kappa)/2$ to distinguish between the C^0_q and $C^{1/2}_q$ SAIR's. The inversion operator $I: f(x_1,x_2) = f(-x_1,-x_2)$ has eigenvalue $\iota = \iota_1\iota_2 = 1-4\epsilon$ in each ϵ-labeled irreducible space, but is no longer related to k, as it was in (3.23).

The coupled eigenstates are thus, explicitly,

$$\Psi^{k_1+,k_2-}_{\epsilon,k,\tau,\rho}(\sigma,r,\theta) = \delta_{\sigma,\tau}\delta(r^2/2-\rho^2/2)F^{k_1+,k_2-}_{\epsilon,k,\tau}(\theta), \tag{4.7}$$

where

$$F^{1/4+,1/4-}_{0,k,\tau}(\theta) = \pi^{-1/2}\cos\kappa\theta, \quad \kappa\geqslant 0, \tag{4.8a}$$

$$F^{3/4+,3/4-}_{0,k,\tau}(\theta) = \pi^{-1/2}\sin\kappa\theta, \quad \kappa>0, \tag{4.8b}$$

$$F^{1/4+,3/4-}_{1/2,k,+1}(\theta) = F^{3/4+,1/4-}_{1/2,k,-1}(\theta) = \pi^{-1/2}\sin\kappa\theta, \quad \kappa>0, \tag{4.8c}$$

$$F^{3/4+,1/4-}_{1/2,k,+1}(\theta) = F^{1/4+,3/4-}_{1/2,k,-1}(\theta) = \pi^{-1/2}\cos\kappa\theta, \quad \kappa\geqslant 0, \tag{4.8d}$$

which are generalized eigenfunctions—we repeat for clarity—of the coupled Q^{k_1,k_2} with eigenvalues $q = k(1-k)\geqslant\frac{1}{4}$, $k = (1+i\kappa)/2$, $\kappa\in R^+$, and form a basis for the continuous nonexceptional C^ϵ_q representation series, $\epsilon = |k_1-k_2|$, and of the coupled $J_-^{k_1,k_2}$ with eigenvalues $\tau\rho^2/2\in R$. The Dirac orthonormality condition which is the analog of (3.13) holds, with $\delta_{k,k'}$ replaced by $\delta_{\tau,\tau'}\delta(\kappa-\kappa')$.

4.5: Again, as in subsection 3.5, we would like to present a method for finding the Dirac-normalization coefficient

$\pi^{-1/2}$ present in (4.8), with the purpose of following the same line of computation in Dirac-normalizing more difficult functions in later sections. Consider $F_k(\theta)=c_k f_k(\theta)$ and $F_l(\theta)=c_l f_l(\theta)$ with $f_k(\theta)=\sin\kappa\theta$ or $\cos\kappa\theta$, $2k-1=i\kappa$, $2l-1=i\lambda$ with support on a single σ chart. Equation (3.14a), stemming from the Casimir operator equation (4.5), holds, with a change of sign obtained through replacing $4(l+k-1)(l-k)\mapsto(\lambda+\kappa)(\lambda-\kappa)$, and since the functions involved have definite parity,

$$\begin{aligned}\lim_{L\to\infty}\int_{-L}^{L} d\theta f_k(\theta)^* f_l(\theta)\\ &=2\lim_{L\to\infty}[(\lambda+\kappa)(\lambda-\kappa)]^{-1}W(f_k^*,f_l)\,|_0^L\\ &=\lim_{L\to\infty}\left[\frac{\sin(\lambda-\kappa)\theta}{\lambda-\kappa}\mp\frac{\sin(\lambda+\kappa)\theta}{\lambda+\kappa}\right]\Bigg|_{\theta=L}\\ &=\pi[\delta(\lambda-\kappa)\mp\delta(\lambda+\kappa)]=\pi\delta(\lambda-\kappa),\end{aligned}\tag{4.9}$$

the upper or lower signs holding for the sine or cosine functions. The $\mp\delta(\lambda+\kappa)$ summand may be discarded on the grounds of the range of the representation indices. The behavior at infinity, being that of the Dirichlet kernel for Fourier transforms, determines the Dirac normalization constant of (4.8) to be $\pi^{-1/2}$. The existence of a limit in the mean will also serve to find the proper linear combination coefficient for the two independent solutions of the Casimir eigenvalue equation so that they form a Dirac-orthonormal set.

4.6: The CGC, finally, is obtained as the inner product between the product and coupled states, as in (3.14). The chart which supports the four δ's in the product state is given by $\sigma=\mathrm{sgn}(\rho_1^2-\rho_2^2)=\tau$, and thus in turn yields $\theta=\mathrm{arctanh}(\rho_2/\rho_1)=\mathrm{arcsinh}(\rho_2/\rho)$ for $\sigma=+1$ and $\theta=\mathrm{arccoth}(\rho_2/\rho_1)=\mathrm{arcsinh}(\rho_1/\rho)$ for $\sigma=-1$. We thus arrive at the general expression

$$\begin{aligned}C\begin{pmatrix}k_1+,k_2-;\epsilon k\\ \rho_1,\ \rho_2\ ;\tau,\rho\end{pmatrix}&=(\psi^{k_1+,k_2-}_{\rho_1,\rho_2},\Psi^{k_1+,k_2-}_{\epsilon,k,\tau,\rho})_{R^2}\\ &=2(\rho_1\rho_2)^{-1/2}\delta(\rho_1^2/2-\rho_2^2/2-\tau\rho^2/2)\\ &\quad\times F^{k_1+,k_2-}_{\epsilon,k,\tau}(\mathrm{arctanh}[(\rho_2/\rho_1)^\tau]).\end{aligned}\tag{4.10}$$

Comparing the above expression with its $D^+\times D^-$ counterpart (3.14), we see that the difference arises in the θ-dependent function [Eqs. (4.8) here vs (3.12) before], where now $2k-1=i\kappa$ and the angle θ is a hyperbolic one. Since $\mathrm{arsinh}\,x=-i\arcsin ix$, we may use (3.18) in order to express (4.10) as hypergeometric functions of ρ_2^2/ρ^2 or ρ_1^2/ρ^2.

For the $D^1\times D^-$ coupling we have thus

$$\begin{aligned}C\begin{pmatrix}\frac14+,\frac14-;0,k\\ \rho_1\ ,\rho_2\ ;\tau,\rho\end{pmatrix}&=\delta(\rho_1^2/2-\rho_2^2/2-\tau\rho^2/2)\\ &\quad\times 2\pi^{-1/2}\rho_1^{-1/2}\rho_2^{-1/2}F\left[\begin{matrix}k-\frac12,\frac12-k\\ \frac12\end{matrix};-\frac{\rho_m^2}{\rho^2}\right],\quad k\geqslant 0,\end{aligned}\tag{4.11a}$$

$$\begin{aligned}C\begin{pmatrix}\frac34+,\frac34-;0,k\\ \rho_1\ ,\rho_2\ ;\tau,\rho\end{pmatrix}&=\delta(\rho_1^2/2-\rho_2^2/2-\tau\rho^2/2)\\ &\quad\times 2\kappa\pi^{-1/2}\rho_1^{1/2}\rho_2^{1/2}\rho^{-2}F\left[\begin{matrix}k+\frac12,\frac32-k\\ \frac32\end{matrix};-\frac{\rho_m^2}{\rho^2}\right],\quad k>0,\end{aligned}\tag{4.11b}$$

$$C\begin{pmatrix}\frac14+,\frac34-;\frac12,k\\ \rho_1\ ,\rho_2\ ;+1,\rho\end{pmatrix}=C\begin{pmatrix}\frac34+,\frac14-;\frac12,k\\ \rho_1\ ,\rho_2\ ;-1,\rho\end{pmatrix}$$

$$=C\begin{pmatrix}\frac34+,\frac34-;0,k\\ \rho_1\ ,\rho_2\ ;+1,\rho\end{pmatrix},\tag{4.11c}$$

$$C\begin{pmatrix}\frac14+,\frac34-;\frac12,k\\ \rho_1\ ,\rho_2\ ;-1,\rho\end{pmatrix}=C\begin{pmatrix}\frac34+,\frac14-;\frac12,k\\ \rho_1\ ,\rho_2\ ;-1,\rho\end{pmatrix}$$

$$=C\begin{pmatrix}\frac14+,\frac14-;0,k\\ \rho_1\ ,\rho_2\ ;+1,\rho\end{pmatrix},\tag{4.11d}$$

where

$$\rho_m=\begin{Bmatrix}\rho_2:&\rho_1>\rho_2&\text{i.e.,}&\tau=+1\\ \rho_1:&\rho_1<\rho_2,&\text{i.e.,}&\tau=-1\end{Bmatrix}=\min(\rho_1,\rho_2).\tag{4.11e}$$

4.7: Note in particular that for $k=\frac12$ ($\kappa=0$), the representation which belongs to the nonexceptional continuous series $C^0_{1/4}$ is obtained from the couplings $D^+_{1/4}\times D^-_{1/4}$. It does not appear in $D^+_{3/4}\times D^-_{3/4}$ [Eqs. (4.11a) vs (4.11b)]. For $\epsilon=\frac12$, the point $k=\frac12$ ($\kappa=0$) belongs, instead, to the $D^\pm_{1/2}$ SAIR's. That these discrete representations occur among the $D^+_\rho\times D^-_\rho$ direct integral decomposition is evident from (4.11c) and (4.11d), and in fact can be seen from (4.8c) and (4.8d): the coupling $D^+_{3/4}\times D^-_{1/4}$ contains $D^+_{1/2}$, but not $D^-_{1/2}$, while the reverse is true for $D^+_{1/4}\times D^-_{3/4}$. The fact that a discrete series representation appears (and no spurious $\epsilon=\frac12$, $k=\frac12$ C-representation) is shown by the J_- eigenvalue of the coupled state being purely positive in the first coupling, and purely negative in the second. This is the signature of a discrete—as opposed to a continuous—SAIR.

4.8: The CGC for the $D^-_\rho\times D^+_\rho$ coupling may be obtained from the $D^+_\rho\times D^-_\rho$ ones displayed above, through the application of the algebra outer automorphism A described in (2.5). The effect of $A=A_1A_2$ is to invert the sign of the spectra of the factor and coupled states under the J_- operators, so that $\tau\leftrightarrow-\tau$ through (4.11). This operator leaves $\mathscr{L}^2_{\iota_1,\iota_2}(R^2)$ invariant and is unitary under the inner product (4.3). It follows that

$$C\begin{pmatrix}k_1-,k_2+;\epsilon,k\\ \rho_1,\ \rho_2\ ;\tau,\rho\end{pmatrix}=C\begin{pmatrix}k_1+,k_2-;\epsilon,k\\ \rho_1,\ \rho_2\ ;\tau,\rho\end{pmatrix},\tag{4.12}$$

for all CGC's.

4.9: As a preparation for Secs. V–X, we shall now project the coupled operator on a definite C^ϵ_q SAIR space, using again the fact that there the coupled Q^{k_1,k_2} is a multiple of the identity. Following (3.19) and (3.20), we use a similarity transformation:

$$\begin{aligned}r^{1/2}\left(\frac{\partial^2}{\partial x_1^2}-\frac{\partial^2}{\partial x_2^2}\right)r^{-1/2}&\\ &=\sigma r^{1/2}\left(\frac{\partial^2}{\partial r^2}+r^{-1}\frac{\partial}{\partial r}+r^{-2}[4Q-1]\right)r^{-1/2}\\ &=\sigma\left(\frac{\partial^2}{\partial r^2}-\frac{\gamma}{r^2}\right),\quad \gamma=(2k-1)^2-\tfrac14=-\kappa^2-\tfrac14\leqslant-\tfrac14.\end{aligned}\tag{4.13}$$

This allows us to write the so(2,1) generators as operators on the hyperbolic coordinates $\sigma=\pm1$, $r\in R^+$ (retaining the index ϵ related to the eigenvalue ι of the inversion operator).

The expressions are

$$J_1^{\epsilon,k}=\frac{\sigma}{4}\left(-\frac{d^2}{dr^2}+\frac{\gamma}{r^2}-r^2\right),\tag{4.14a}$$

$$J_2^{\epsilon,k}=-\frac{i}{2}\left(r\frac{d}{dr}+\frac12\right),\tag{4.14b}$$

$$J_0^{\epsilon,k} = \frac{\sigma}{4}\left(-\frac{d^2}{dr^2} + \frac{\gamma}{r^2} + r^2\right), \tag{4.14c}$$

$$J_+^{\epsilon,k} = \frac{\sigma}{2}\left(-\frac{d^2}{dr^2} + \frac{\gamma}{r^2}\right), \tag{4.14d}$$

$$J_-^{\epsilon,k} = \frac{\sigma}{2} r^2. \tag{4.14e}$$

These operators are self-adjoint on the space of two-chart functions[13d,14] in $\mathscr{L}^2_{\sigma=1}(R^+)\times\mathscr{L}^2_{\sigma=-1}(R^+)$ with inner product

$$(f,g)_{\mathscr{S}} = \sum_{\sigma=\pm 1}\int_0^\infty dr\, f(\sigma,r)^* g(\sigma,r),$$
$$\mathscr{S} = \{-1,+1\}\times R^+. \tag{4.15}$$

The automorphism A exchanges charts $\sigma\leftrightarrow -\sigma$ and is unitary under (4.15).

The index ϵ, as before in (3.21), seems to be absent from the right-hand sides of (4.14). As will be seen in full detail below, it continues to exist as a specification of the domain of the operators: It indicates that we are considering the common self-adjoint extension of (4.14) in which the spectrum of $J_0^{k_1,k_2}$ is contained in $\{\epsilon + n, n\in Z\}$. In the discrete series case, for $k\geqslant 1$ there is a unique self-adjoint extension, and the index ϵ could in principle be dropped. The above operators, on the other hand, for $\gamma < \frac{3}{4}$ (including $\gamma\leqslant -\frac{1}{4}$ here) have a one-parameter family of such self-adjoint extensions.

4.10: In each chart, the operators $J_1^{\epsilon,k}$, $J_0^{\epsilon,k}$, and $J_+^{\epsilon,k}$ represent Schrödinger Hamiltonians with a strongly singular behavior at zero,[20] due to the terms γ/r^2 for $\gamma\leqslant -\frac{1}{4}$. We would like to study the features of such singular operators in $\mathscr{L}^2(R^+)$, on a single σ-chart, in particular, give a short account of the one-parameter family of self-adjoint extensions of $J_0^{\epsilon,k}$, and their spectra, both for $k\in\mathscr{D}$ and $k\in\mathscr{C}$. Then we shall do the same for the two-chart inner product (4.15), and thus make precise the applicability of (4.14) for both D^+, C as well as for the exceptional continuous representation series $E(\frac{1}{2} < k < 1)$, which does not appear in the decomposition of the product of two oscillator representations. Operators with an unbounded spectrum are not commonly used in physics, and the sharper tools of analysis must be brought to bear on the subject. From our knowledge of the so(2,1) representation series, we do expect that J_0 in our case (4.14) and (4.15) will have a spectrum given by $\epsilon + n$, $n\in Z$.

V. THE GENERALIZED OSCILLATOR ALGEBRA

The purpose of this section is to study the generalized oscillator realization of the so(2,1) algebra given by the second-order differential operators (4.14) [including (3.21) as a particular case] and to show how, through an appropriate definition of the function domains on two charts, these provide for all self-adjoint representations of the algebra.

5.1: To this end, we consider first the eigenvalue equation

$$K^k\phi_\mu^k(r) = \mu\phi_\mu^k(r), \quad r\in R^+, \tag{5.1a}$$

where K^k is J_0^k as given by (4.14) on the single chart $\sigma = 1$:

$$K^k = \frac{1}{4}\left(-\frac{d^2}{dr^2} + \frac{\gamma}{r^2} + r^2\right), \quad \gamma = (2k-1)^2 - \tfrac{1}{4}\in R, \tag{5.1b}$$

representing the Schrödinger Hamiltonian of a harmonic oscillator with a centrifugal barrier ($\gamma > 0$) or centripedal well ($\gamma < 0$). The solutions to Eq. (5.1), which vanish at infinity, are given by the Whittaker functions (Ref. 19, Eq. 13.1.33):

$$\phi_\mu^k(r) = c_\mu^k[\Gamma(k-\mu)\Gamma(1-k-\mu)]^{1/2} r^{-1/2} W_{\mu,k-1/2}(r^2)$$
$$= \phi_\mu^{1-k}(r), \tag{5.2}$$

where the Γ-function factor has been chosed for convenience. Note that the functions $\phi_\mu^k(r)$, being invariant under $k\leftrightarrow 1-k$ are real for all $k\in\mathscr{D}$, $(k>0)$ or $k\in\mathscr{C}$ $[k = (1+i\kappa)/2, \kappa\in R^+]$.

5.2: The behavior of $\phi_\mu^k(r)$ near $r = 0$ is given by

$$\phi_\mu^k(r) \underset{r\to 0^+}{\sim} c_\mu^k\{\Gamma(1-2k)[\Gamma(k-\mu)/\Gamma(1-k-\mu)]^{1/2} r^{2k-1/2}$$
$$+ \Gamma(2k-1)[\Gamma(1-k-\mu)/\Gamma(k-\mu)]^{1/2} r^{-2k+3/2}\}. \tag{5.3}$$

(i) When $k\geqslant 1$ the first summand is locally square-integrable but the second is not, unless $\mu = k+n$, $n\in Z^+$, when the Γ function in the latter's denominator makes the summand vanish. The pole in the coefficient of the first summand is an artifice of the Γ factor, and is cancelled by an appropriate normalization coefficient c_μ^k. Hence $\phi_\mu^k \in\mathscr{L}^2(R^+)$. It is, in fact, a Laguerre polynomial in r^2 of degree $\mu - k$ and order $2k-1$, times $r^{2k-1/2}\exp(-r^2/2)$. The spectrum of K^k for $k\geqslant 1$ is thus $\mu = k+n$, $\mu\in Z^+$: equally spaced and lower-bound.

(ii) When $0 < k < 1$, i.e., k is in the exceptional interval, both summands in (5.3) are locally square-integrable, and no quantization condition on μ follows from $\mathscr{L}^2(R^+)$ alone. For $\frac{1}{2} < k < 1$ the first term has a locally square-integrable derivative (and the second term must be eliminated with $\mu = k+n$, $n\in Z^+$), while when $0 < k < \frac{1}{2}$, the same is true if we exchange the first and second terms, and keep $\mu = k+n$, $n\in Z^+$. These correspond to Laguerre polynomials as above, including the case $k = \frac{1}{2}$ when proper normalization is applied.

(iii) When $k\in C$ $[k = (1+i\kappa)/2, \kappa\in R^+]$, then for all $\mu\in R$ (5.3) is in $\mathscr{L}^2(R^+)$. Its derivative does not have this property.

5.3: Although K^k formally appears to be symmetric, we find that, under the usual $\mathscr{L}^2(R^+)$ inner product (3.19),

$$(K^k\phi_\mu^k,\phi_\nu^k)_{R^+} - (\phi_\mu^k,K^k\phi_\nu^k)_{R^+} = \tfrac{1}{2}W(\phi_\mu^{k*},\phi_\nu^k)\,|_0^\infty, \tag{5.4}$$

where $W(f,g)$ is as before the Wronskian of f and g. The boundary term vanishes at infinity, while at zero it can be evaluated from (5.3). The cases for $k\in C$ and k real differ slightly but lead to the same result. The $\mathscr{L}^2(R^+)$ inner product of ϕ_μ^k and ϕ_ν^k can be thus obtained as

$$(\phi_\mu^k,\phi_\nu^k)_{R^+} = [2(\mu-\nu)]^{-1} W(\phi_\mu^{k*},\phi_\nu^k)\,|_{r\to 0}$$
$$= c_\mu^{k*} c_\nu^k \pi\csc(2\pi k)\left\{\left[\frac{\Gamma(1-k-\mu)\Gamma(k-\nu)}{\Gamma(k-\mu)\Gamma(1-k-\nu)}\right]^{1/2}\right.$$
$$\left. - \left[\frac{\Gamma(k-\mu)\Gamma(1-k-\nu)}{\Gamma(1-k-\mu)\Gamma(k-\nu)}\right]^{1/2}\right\}. \tag{5.5}$$

For each fixed $k > 0$, (5.5) vanishes when $\mu = k+m$, $m\in Z^+$, and $\nu = k+n$, $n\in Z^+$, and K^k is self-adjoint then in the Hilbert space which has $\{\phi_\mu^k\}$ as its denumerable basis.

This Hilbert space is unique when $k \geqslant 1$, but for the exceptional interval $0 < k < 1$ it is merely one of a family of Hilbert spaces: that whose denumerable basis functions have square-integrable derivatives. [This also holds for $k = \frac{1}{2}$ in spite of the apparent singularity of (5.5).] The common feature of these domains is that the spectrum of K^k is lower-bound and equally spaced, signifying that we are in the discrete series D^+_k.

5.4: In order to give a common Hilbert space domain for all generators of so(2,1) [(4.14) for fixed k and $\sigma = +1$], it is necessary to know whether a proposed domain with basis $\{\phi^k_\mu\}$ is invariant under the action of a generator basis. We are using the spectrum of $K^k = J^{k+}_0$ to specify such domains. When we construct the raising and lowering operators $J_{\uparrow\downarrow} = J_1 \pm iJ_2$ and apply them to any ϕ^k_μ, μ in the spectrum of K^k, we may employ the usual algebraic argument to show that $J_{\uparrow\downarrow}\phi^k_\mu$ should be an eigenfunction of K^k with eigenvalue $\mu \pm 1$. Unless $\mu \pm 1$ is in the spectrum of K^k when μ is, the rest of the algebra generators will not leave the domain invariant.

Lower-boundedness is no problem since it is elementary to show that $J_\downarrow \phi^k_k = 0$, but a necessary condition to be able to specify a common self-adjoint extension of all operators in so(2,1) is that the domain be such that the spectrum of K^k be equally spaced.

5.5: The domain in which K^k, for k in the exceptional interval, has a lower-bound equally spaced spectrum, however, is only a particular self-adjoint extension of the operator.[18] As said before, this is specified through the vector space on which K^k is to act. This domain can here be in turn described as those $\mathscr{L}^2(R^+)$ functions f whose boundary Wronskian $W(f,\phi_{\mu_0})\,|_0$ is to vanish, thus setting the boundary term in (5.5) to zero. This is equivalent to imposing that μ_0 be a point in the spectrum of the operator.

It can be shown that, for k in $(\frac{1}{2},1)$, the spectrum of each self-adjoint extension of K^k has a lower bound, but this minimum eigenvalue has itself no lower bound, so that extensions can be found where this minimum is as negative large as we please. For $k \in C$ the spectrum has no lower bound. Two more general results[18] are that the union of all self-adjoint extensions provide spectra which cover the real line. No eigenvalue μ_0 can belong to the spectrum of more than one self-adjoint extension of K^k.

Given thus some μ_0, finding the rest of the spectrum is an exercise in solving for ν the transcendental equation

$$\Gamma(k-\nu)/\Gamma(1-k-\nu) = \Gamma(k-\mu_0)/\Gamma(1-k-\mu_0), \tag{5.6}$$

which puts to zero the last expression in (5.5). As a numerical example, if for $k = \frac{5}{6}$ we choose $\mu_0 = -1$, the spectrum of K^k will be (up to three decimal places) $\{-1.0, 1.046, 2.097, 3.123, 4.141, 5.153, 6.163, 7.170, 8.176, 8.179, 9.182, \cdots\}$, while if for $k = \frac{1}{2} + i$ we choose $\mu_0 = 0$, the spectrum will be $\{\cdots, -8.881, 0.0, 1.449, 2.619, 3.727, 4.806, 5.869, 6.921, 7.965, \cdots\}$. These eigenvalues tend asymptotically toward equal spacing; but the spectrum is not linear, and the domain specified by $\mu_0 \in R$, for $\mu_0 - k$ not a positive integer, is therefore not fit to serve as a common invariant domain for a set of operators which are to be elements of a Lie algebra.

5.6: We shall now see what the choice of self-adjoint extensions is for the two-chart operator $J^{\epsilon,k}_0$ in (4.14c) with inner product (4.15). The eigenvalue equation to be solved now has two forms, one in each chart $\sigma = \pm 1$. Since $J^{\epsilon,k}_0 = \sigma K^k$, the equations are

$$K^k\psi^k_\mu(\sigma = +1, r) = \mu\psi^k_\mu(\sigma = +1, r), \tag{5.7a}$$

$$-K^k\psi^k_\mu(\sigma = -1, r) = \mu\psi^k_\mu(\sigma = -1, r). \tag{5.7b}$$

The difference with (5.1) appears to be slight. We can set

$$\psi^k_\mu(\sigma = +1, r) = \phi^k_\mu(r), \quad \psi^k_\mu(\sigma = -1, r) = a^k_\mu \phi^k_{-\mu}(r), \tag{5.7c}$$

with a^k_μ a constant to be determined. We now repeat (5.4) under the inner product (4.15):

$$(J^k_0\psi^k_\mu, \psi^k_\nu)_{\mathscr{S}} - (\psi^k_\mu, J^k_0\psi^k_\nu)_{\mathscr{S}} = \tfrac{1}{2}W(\phi^{k*}_\mu, \phi^k_\nu)\,|_{r=0} - \tfrac{1}{2}a^{k*}_\mu a^k_\nu W(\phi^{k*}_{-\mu}, \phi^k_{-\mu})\,|_{r=0}. \tag{5.8}$$

Upon performing the necessary algebra with (5.5), we can see that whenever $\mu - \nu$ is an integer, a full cancellation of the boundary terms occurs for $a^k_\mu = 1 = a^k_\nu$. The normalization constant c^k_μ of the constituent functions (5.2) may be found along the lines of subsection 3.5, or directly through integral tables (Ref. 2, Eqs. 7.6114 and 8.365.9). If we denote $\mu = \epsilon + n$, $\epsilon \in (-\frac{1}{2}, \frac{1}{2}]$, $n \in Z$, c^k_μ depends only on k and ϵ [this feature determined the Γ-function choice in (5.2)]. It is

$$c^k_\mu = c^{\epsilon k} = \pi^{-1}[2\sin\pi(k-\epsilon)\sin\pi(k+\epsilon)]^{1/2} = c^{\epsilon,1-k} = c^{-\epsilon,k}. \tag{5.9}$$

If $k \in C$, the functions $\psi^{\epsilon k}_\mu(\sigma, r)$ are the eigenfunctions of the self-adjoint extension of the operator $J^{\epsilon k}_0$ with spectrum $\mu \in \{\epsilon + n, n \in Z\}$. This extension may be labeled through the index $\epsilon \in (-\frac{1}{2}, \frac{1}{2}]$, which generalizes the previously used homonym index $\epsilon \in (0, \frac{1}{2})$ to multivalued representations of the group.

When $\epsilon = \frac{1}{2}$, the limit $k = (1 + i\kappa)/2$, $\kappa \to 0$, results in $c^{\epsilon k}$ becoming zero. For $\psi^k_\mu(+1, r)$, $\mu = \frac{1}{2}, \frac{3}{2}, \cdots$, this zero is compensated, however, by a matching pole from the Γ functions in (5.2); similarly, only the $\sigma = -1$ component of $\psi^k_\mu(\sigma, r)$, $\mu = -\frac{1}{2}, -\frac{3}{2}, \cdots$, is nonzero. These functions therefore belong to the discrete series of SAIR's, having a single-chart support.

5.7: The algebra generators (4.14) belong to the exceptional continuous SAIR series E when $0 < k < 1$ ($-\frac{1}{4} \leqslant \gamma < \frac{3}{4}$ twice). The domain is that labeled by ϵ, with $|\epsilon| < \min(k, 1-k)$ so that the radicand in (5.9) is positive. Since reflection symmetry $k \leftrightarrow 1-k$ holds, we may reduce to $\frac{1}{2} < k < 1$ and $|\epsilon| < 1 - k$. The spectrum of $J^{\epsilon,k}_0$ is in this case again given by $\mu \in \{\epsilon + n, n \in Z\}$.

We shall now examine the boundary of the exceptional series region (Fig. 1), $|\epsilon| = \min(k, 1-k)$, which gives rise to the lower- and upper-bound discrete series $D^\pm_k$ in the interval $0 < k < 1$, including the two direct summands of the oscillator representation. There is an interplay between the zeros of $c^{\epsilon k}$ in (5.9) and the Φ-function poles in the factor of (5.2). Consider the $\sigma = +1$ component and see Fig. 1. When $0 < \epsilon = k \leqslant \frac{1}{2}$ or $0 > \epsilon = k - 1 \geqslant -\frac{1}{2}$, the only nonzero func-

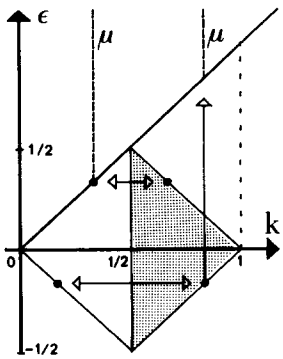


FIG. 1. The exceptional interval. The shaded region indicates the range of k and ϵ for the continuous exceptional series, $|\epsilon| < 1 - k$, where the $k \leftrightarrow 1 - k$ symmetry is used to reduce k to $(\frac{1}{2},1)$. The boundary corresponds to the discrete series. The points $0 < \epsilon = k$ (for $0 < k \leqslant \frac{1}{2}$) and $0 < \epsilon = 1 - k$ (for $\frac{1}{2} \leqslant k < 1$) lead to eigenfunctions ψ^k_μ with spectrum $\{\epsilon + n,\ n \in Z^+\}$, while $0 > \epsilon = k - 1$ (for $\frac{1}{2} > k > 1$) and $0 > \epsilon = -k$ (for $0 < k \leqslant \frac{1}{2}$) lead to $\{\epsilon + n + 1,\ n \in Z^+\}$.

tions with $\mu = \epsilon + n$, $n \in Z$, are those with $\mu = k, k+1, k+2, \cdots$ while when $0 > \epsilon = -k \geqslant -\frac{1}{2}$ or $0 < \epsilon = 1 - k \leqslant \frac{1}{2}$, the only nonzero functions with $\mu = \epsilon + n$, $n \in Z$, are those with $\mu = 1 - k, 2 - k, 3 - k, \cdots$. As an example, $\epsilon = \frac{1}{4}$, $k = \frac{1}{4}$, and $\epsilon = \frac{1}{4}$, $k = \frac{3}{4}$ characterize the same self-adjoint extension in the even positive harmonic oscillator state space, while $\epsilon = -\frac{1}{4}$, $k = \frac{1}{4}$ and $\epsilon = \frac{1}{4}$, $k = \frac{3}{4}$ characterize the odd oscillator state space. The operator K^k, is, of course, the same under the exchange $k \leftrightarrow 1 - k$. The above arguments, however, lead us to drop this symmetry and adopt $k \in (0,1)$ as the sole indicator of the self-adjoint extension of the algebra meant for the discrete series in the exceptional interval, as was done in point (ii) early in this section.

The same reasoning holds for the $\sigma = -1$ chart of the wavefunctions $\psi^k_\mu(\sigma,r)$, under the exchange $\mu \leftrightarrow -\mu$, leading to basis functions for the upper-bound negative discrete series D^-_k. For $k \geqslant 1$ we are left, thus, with a description of the discrete series in terms of a direct sum of two function spaces, where $\psi^k_\mu(\sigma,r)$ with $\mu = k, k+1, \cdots$ has only a $\sigma = +1$ chart support, and with $\mu = -k, -k-1, \cdots$ has only a $\sigma = -1$ chart support. These constitute single-chart SAIR bases for D^+_k and D^-_k, respectively.

The outer automorphism A effecting (2.5a) may be properly defined on functions of $\mathcal{S}$ through

$$Af(\sigma,r) = f(-\sigma,r). \tag{5.10}$$

It exchanges the lower- and upper-bound discrete SAIR spaces and leaves invariant the continuous series SAIR spaces.

5.8: As a conclusion to the last four sections, we should like to restate that the so(2,1) algebra realized through the set of formal second-order differential operators given in (4.14) for any real γ, on spaces of functions on $\mathcal{S} = \{-1, +1\} \times R^+$ with self-adjoint extensions in $\mathcal{L}^2(\mathcal{S})$ specified through the index ϵ in the manner brought out above, accommodates all the SAIR's of the algebra. We have devoted some space to show that this is true for the continuous $\mathcal{C}$ series, and justified that the more naive treatments found in the literature for the $\mathcal{D}$ case are also appropriate. In the next section we proceed to generalize the construction of the CGC's for general SAIR products.

VI. THE COUPLING OF TWO GENERAL REPRESENTATIONS

The tactics seen in Secs. III and IV provide us now with the strategy for coupling any pair of so(2,1) SAIR's as described in Sec. V.

6.1: We shall have two sets of operators (4.14) in two independent sets of variables $\sigma_1, r_1 \in \mathcal{S}$ and $\sigma_2, r_2 \in \mathcal{S}$ generating $\mathcal{D}$ and/or $\mathcal{C}$ representations. The factor operator sets $J^{k_j}_{(j)\alpha}, j = 1,2$, will refer to these variables, while the coupled operators will be, as before,

$$J_\alpha = J^{k_1}_{(1)\alpha} + J^{k_2}_{(2)\alpha}, \quad \alpha = 1,2,0,+,-. \tag{6.1}$$

The labels k_j are assumed to contain all the SAIR information, appearing as $k_j \pm$ for $\mathcal{D}$ and ϵ_j, k_j for $\mathcal{C}$. The domain of the operators will be the set of two-variable functions which belong, as functions of each, to the Hilbert spaces specified in the last section, where the spectrum of $J^{k_j}_{(j)0}$ is contained in $\{\epsilon_j + n,\ n \in Z\}$ understanding that for D^+, $\epsilon_j \equiv k_j$, modulo unity. The defining inner product may be written as

$$(\phi,\chi)_{\mathcal{S}^2} = \sum_{\sigma_1,\sigma_2 = \pm 1} \int_0^\infty dr_1 \int_0^\infty dr_2\, \phi(\sigma_1,r_1,\sigma_2,r_2)^* \times \chi(\sigma_1,r_1,\sigma_2,r_2). \tag{6.2}$$

We recall, finally that when one or both of the factor SAIR's belong to $D^\pm$, the support of the functions will be the $\sigma_j = \pm 1$ chart only.

6.2: The product states

$$\psi^{k_1,k_2}_{\tau_1,\rho_1,\tau_2,\rho_2}(\sigma_1,r_1,\sigma_2,r_2) = \delta_{\sigma_1,\tau_1}\delta_{\sigma_2,\tau_2}(\rho_1\rho_2)^{-1/2} \times \delta(r_1 - \rho_1)\delta(r_2 - \rho_2) \tag{6.3}$$

are the generalized eigenfunctions of the following set of commuting operators:

(i) $Q^{k_1}_{(1)}$ and $Q^{k_2}_{(2)}$ (realized as multiples of the identity) with eigenvalues $q_1 = k_1(1 - k_1)$, $q_2 = k_2(1 - k_2)$.

(ii) $J^{k_1}_{(1)-}$ and $J^{k_2}_{(2)-}$ with eigenvalues $\tau_1\rho_1^2/2$ and $\tau_2\rho_2^2/2$, respectively $(\tau_j \in \{-1,1\}, \rho_j \in R^+, j = 1,2)$. For $D^+_{k_j}$, τ_j is identically $+1$, and for $D^-_{k_j}$, $\tau_j = -1$.

(iii) The common domain of the above two-variable operators is taken to be the space described in the last section, specified by ϵ_1, ϵ_2. A denumerable basis for this space is $\psi^{k_1}_{\mu_1}(\sigma_1,r_1)\psi^{k_2}_{\mu_2}(\sigma_2,r_2)$ with $\pm\mu_j \in \{k_j + n,\ n \in Z^+\}$ for $D^\pm \ni k_j \equiv \epsilon_j \pmod 1$ and $\mu_j \in \{\epsilon_j + n,\ n \in Z\}$ for $k_j \in \mathcal{C}$. In the special case when ϵ_1, ϵ_2 are in $\{0, \frac{1}{2}\}$ the product states can be alternatively placed in the domain where the inversion operators $I_{(1)}$ and $I_{(2)}$ have eigenvalues ι_1 and ι_2 in $\{-1, +1\}$. The product states (6.3) are Dirac-orthonormal under (5.2):

$$(\psi^{k_1,k_2}_{\tau_1,\rho_1,\tau_2,\rho_2}, \psi^{k_1,k_2}_{\tau_1',\rho_1',\tau_2',\rho_2'})_{\mathcal{S}^2} = \delta(\tau_1\rho_1^2/2 - \tau_1'\rho_1'^2/2) \times \delta(\tau_2\rho_2^2/2 - \tau_2'\rho_2'^2/2). \tag{6.4}$$

6.3: The coupled states, denoted by $\Psi^{k_1,k_2}_{k,\tau,\rho}$, are the generalized eigenfunctions of the commuting operators:

(i') $Q^{k_1}_{(1)}$ and $Q^{k_2}_{(2)}$ with eigenvalues $q_1 = k_1(1 - k_1)$, $q_2 = k_2(1 - k_2)$ as before.

(ii') of the parabolic subalgebra generator

$$J_- = J^{k_1}_{(1)-} + J^{k_2}_{(2)-} = \sigma_1 r_1^2/2 + \sigma_2 r_2^2/2 = \sigma r^2/2 \quad (6.5)$$

[where in the last expression $\sigma \in \{-1,1\}$ and $r \in R^+$ will be properly defined below] with eigenvalue $\tau\rho^2/2$ ($\tau \in \{-1,1\}$ $\rho \in R^+$). Again, if the coupled SAIR belongs to $D^\pm$, the eigenvalue will be in $R^\pm$, so only $\tau = +1$ or $\tau = -1$ is allowed, while if it belongs to $\mathscr{C}$ it will be in R. Generalized eigenfunctions of (6.5) will thus have a factor $\delta(\sigma r^2/2 - \tau\rho^2/2) = \delta_{\sigma,\tau}\delta(r^2/2 - \rho^2/2)$ as in (4.6).

(iii′) The spectrum of the compact generator $J_0 = J^{k_1}_{(1)0} + J^{k_2}_{(2)0}$ as seen in Sec. V defines the self-adjoint extension domain. Here it is $\mu = \mu_1 + \mu_2$. Modulo integers, this implies that the coupled SAIR space is characterized by

$$\epsilon \equiv \epsilon_1 + \epsilon_2 \pmod 1. \quad (6.6)$$

Again, in the case when $\epsilon_1, \epsilon_2 \in \{0, \frac{1}{2}\}$ we may make use of the inversion operator $I = I_{(1)}I_{(2)}$ with eigenvalue $\iota = \iota_1\iota_2$, implying (6.6). This restricts the coupled SAIR's in a form analogous to the familiar rotation group coupling between vector and spinor UIR's.

(iv′) of the coupled Casimir operator Q, built as in (2.2) out of the coupled generators J_α, with eigenvalue $q = k(1-k)$. The explicit form of Q in terms of the $\mathscr{S}^2$ space variables $(\sigma_1,r_1,\sigma_2,r_2)$ is

$$\begin{aligned} Q &= (J_1)^2 + (J_2)^2 - (J_0)^2 = Q^{k_1}_{(1)} + Q^{k_2}_{(2)} \\ &\quad + 2\left[J^{k_1}_{(1)1}J^{k_2}_{(2)1} + J^{k_1}_{(1)2}J^{k_2}_{(2)2} - J^{k_1}_{(1)0}J^{k_2}_{(2)0}\right] \\ &= \frac{1}{4}\left[\sigma_1\sigma_2\left(r_1\frac{\partial}{\partial r_2} - \sigma_1\sigma_2 r_2\frac{\partial}{\partial r_1}\right)^2 \right. \\ &\quad \left. - \gamma_1\left(1 + \frac{\sigma_1\sigma_2 r_1^2}{r_1^2}\right) - \gamma_2\left(1 + \frac{\sigma_1\sigma_2 r_1^2}{r_2^2}\right) + 1\right]. \end{aligned} \quad (6.7)$$

6.4: We shall now parametrize the space $\mathscr{S}^2$ in a manner which encompasses the polar and hyperbolic coordinates used in Sec. III and IV. Only the case $\sigma_1 = \sigma_2$ appeared in Sec. III, and only $\sigma_1 = -\sigma_2$ in Sec. IV. On account of the two dichotomic indices σ_1,σ_2, we have four coordinate charts. In attention to the defintion in the last equality in (6.5), in accordance with $r_1 \gtrless r_2$, two of these charts will be reparametrized, so that we have a total of six charts as follows:

(i) $P^\pm$-charts:

$$\begin{aligned} &\text{for } \sigma_1 = \pm 1 = \sigma_2, \quad r_1, r_2 \in R^+; \quad \sigma = \pm 1, \text{ respectively,} \\ &r_1 = r\cos\theta, \quad r_2 = r\sin\theta, \qquad r \in R^+, \quad \theta \in [0, \pi/2]. \end{aligned} \quad (6.8a)$$

(ii) $H^\pm_>$-charts:

$$\begin{aligned} &\text{for } \sigma_1 = \pm 1 = -\sigma_2, \quad r_1 > r_2; \quad \sigma = \pm 1, \text{ respectively,} \\ &r_1 = r\cosh\theta, \quad r_2 = r\sinh\theta, \qquad r \in R^+, \quad \theta \in [0, \infty). \end{aligned} \quad (6.8b)$$

(iii) $H^\pm_<$-charts:

$$\begin{aligned} &\text{for } \sigma_1 = \mp 1 = -\sigma_2, \quad r_1 < r_2; \quad \sigma = \pm 1, \text{ respectively,} \\ &r_1 = r\sinh\theta, \quad r_2 = r\cosh\theta, \qquad r \in R^+, \quad \theta \in [0, \infty). \end{aligned} \quad (6.8c)$$

To this list we should add two more charts for the line manifolds $\sigma_1 = -\sigma_2, r_1 = r_2$ but, since these are of lower dimension we may disregard them in what follows.

A single function $\phi(\sigma_1,r_1,\sigma_2,r_2)$ over $\mathscr{S}^2$, when the coordinates of $\mathscr{S}^2$ are expressed in terms of (r,θ) in chart C, will be represented by a function $\phi^C(r,\theta)$. The set of six such functions $\{\phi^C, C = P^\sigma, H^\sigma_>, H^\sigma_<, \sigma = \pm 1\}$ constitutes the original ϕ in the new coordinate system (C,r,θ).

We have found it useful to depict the six charts as in Fig. 2, considering the product $\sigma_j r_j$ as if it were a coordinate for R. The $D^+ \times D^+$ coupling (Sec. VII) will require only one chart: the first quadrant P^+, since $\sigma_1 = +1 = \sigma_2$, in complete analogy with Sec. III, had we halved r to R^+ through parity. The $D^+ \times D^-$ coupling (Sec. VIII), having $\sigma_1 = +1 = -\sigma_2$, will require the two charts, $H^+_>$ and $H^-_<$, in the fourth quadrant. In Sec. IX we treat the $D^+ \times \mathscr{C}$ coupling, where $\sigma_1 = +1, \sigma_2 = \pm 1$, so that the right half-plane of Fig. 2 is needed and will be covered by the three-coordinate charts P^+, $H^+_>$, and $H^-_<$. Finally, the $\mathscr{C} \times \mathscr{C}$ coupling in Sec. X requires all six charts. The joint consideration of the required number of charts is important since the product and coupled operators will continue to be self-adjoint in the corresponding space (6.6) only if the formal replacement of variables $(\sigma_1,r_1,\sigma_2,r_2) \to (C,r,\theta)$ through (6.8) is made on the appropriate union of charts. The form of J_- in (6.5) has been tailored to that purpose.

6.5: The Casimir operator (6.7) has the three following forms in each of the charts (6.8):

(i) in $P^\pm$:

$$Q = \frac{1}{4}\left[\frac{d^2}{d\theta^2} - \gamma_1\sec^2\theta + \gamma_2\csc^2\theta + 1\right], \quad (6.9a)$$

(ii) in $H^\pm_>$:

$$Q = \frac{1}{4}\left[-\frac{d^2}{d\theta^2} - \gamma_1\operatorname{sech}^2\theta + \gamma_2\operatorname{csch}^2\theta + 1\right], \quad (6.9b)$$

(iii) in $H^\pm_<$:

$$Q = \frac{1}{4}\left[-\frac{d^2}{d\theta^2} + \gamma_1\operatorname{csch}^2\theta - \gamma_2\operatorname{sech}^2\theta + 1\right], \quad (6.9c)$$

generalizing thus (3.9) and(4.5).

6.6: The inner product (6.2) defining the Hilbert space will be expressible as

$$\begin{aligned} (\phi,\chi)_{\mathscr{S}^2} = \sum_{\sigma=\pm 1} &\left[(\phi^{P^\sigma},\chi^{P^\sigma})_P + (\phi^{H^\sigma_>},\chi^{H^\sigma_>})_H \right. \\ &\left. + (\phi^{H^\sigma_<},\chi^{H^\sigma_<})_H\right], \end{aligned} \quad (6.10a)$$

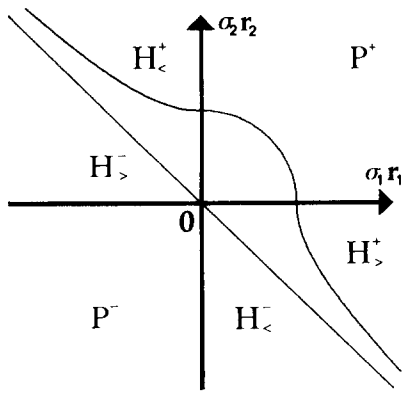


FIG. 2. The six coordinate charts in $\mathscr{S}^2$.

$$(\alpha,\beta)_P = \int_0^\infty r\,dr \int_0^{\pi/2} d\theta\, \alpha(r,\theta)^*\beta(r,\theta), \tag{6.10b}$$

$$(\alpha,\beta)_H = \int_0^\infty r\,dr \int_0^\infty d\theta\, \alpha(r,\theta)^*\beta(r,\theta). \tag{6.10c}$$

Although each of the formal operators (6.9) may be separately self-adjoint under (6.10b) or under one of (6.10c) in some appropriate one-chart Hilbert space, Q itself, as represented by all three forms (6.9), will have the proper spectrum only when placed in the full inner product (6.10), and only there will the orthogonality of the expected eigenfunctions hold consistent with (6.6). When a subspace of functions is identically zero in some charts (as in the $D^+ \times D^+$ case where the support lies entirely in the P^+ chart), the inner product may be reduced to a subset of the summands [to (6.10b) alone, with $\sigma = +1$ in the $D^+ \times D^+$ case].

6.7: Since the coupled states are eigenstates of J_- with eigenvalue $\tau\rho^2/2$, they will contain, as discussed above, a factor of $\delta_{\sigma\tau}\,\delta(r^2/2 - \rho^2/2)$. The Kronecker δ restricts the support of the function to the upper ($\sigma = \tau = +1$) or lower ($\sigma = \tau = -1$) diagonal half-planes in Fig. 2; the Dirac δ restricts the support further to a constant value of r^2; as shown in the figure, this is a quarter-circle in the P^σ-chart, and a quarter-hyperbola in the $H^\sigma_>$ and $H^\sigma_<$ charts.

As before [cf. Eqs. (3.10) and (4.6)], we write the product state as

$$\Psi^{k_1,k_2,C^\sigma}_{k,\tau,\rho}(r,\theta) = \delta_{\sigma,\tau}\delta(r^2/2 - \rho^2/2)F^C_{k,\tau}(\theta), \tag{6.11}$$

the index C^σ standing for a specification of the chart. The requirement that they be eigenfunctions of the Casimir operator (6.9) with eigenvalue $k(1-k)$ leads to the following differential equations for the θ-dependent factor:

(i) on the P-charts, $\theta\in[0,\pi/2]$:

$$\left[-\frac{d^2}{d\theta^2} + \gamma_1\sec^2\theta + \gamma_2\csc^2\theta\right]F^P_{k,\tau}(\theta) = (2k-1)^2F^P_{k,\tau}(\theta); \tag{6.12a}$$

(ii) on the $H_>$-charts, $\theta\in[0,\infty)$:

$$\left[-\frac{d^2}{d\theta^2} - \gamma_1\operatorname{sech}^2\theta + \gamma_2\operatorname{csch}^2\theta\right]F^{H_>}_{k,\tau}(\theta) = -(2k-1)^2F^{H_>}_{k,\tau}(\theta); \tag{6.12b}$$

(iii) on the $H_<$-charts, $\theta\in[0,\infty)$:

$$\left[-\frac{d^2}{d\theta^2} + \gamma_1\operatorname{csch}^2\theta - \gamma_2\operatorname{sech}^2\theta\right]F^{H_<}_{k,\tau}(\theta) = -(2k-1)^2F^{H_<}_{k,\tau}(\theta). \tag{6.12c}$$

We recall that $\gamma_j = (2k_j - 1)^2 - \frac{1}{4}$ for all cases, so that $\gamma_j > -\frac{1}{4}$ for $k_j\in\mathscr{D}$, and $\gamma_j < (1-2|\epsilon_j|)^2 - \frac{1}{4}$ for $k_j\in\mathscr{C}$, and the exceptional interval $0<k<1$ corresponds to $-\frac{1}{4}\leqslant\gamma_j<\frac{3}{4}$. Equation (6.12a) is the Schrödinger equation for a Pöschl–Teller potential[16] of the first kind, while Eqs. (6.12b) and (6.12c) (which only exchange $\gamma_1\leftrightarrow\gamma_2$) are Schrödinger equations for Pöschl–Teller potentials of the second kind. The singularities of the potentials [$\theta = 0$ and $\pi/2$ in (6.12a), and $\theta = 0$ in (6.12b) and (6.12c)] are of the inverse-quadratic type, and in Fig. 2 can be seen to lie on the coordinate axes: The value of k_1 provides the coefficient γ_1 of the singularity on the $r_1 = 0$ axis ($\theta = \pi/2$ in P, $\theta = 0$ in $H_<$), and k_2 the coefficient γ_2 of that on the $r_2 = 0$ axis ($\theta = 0$ in P, $\theta = 0$ in $H_>$).

In the two H cases, the potential tends exponentially to zero at $\theta\to\infty$. As we shall see in detail in the following four sections, when the support of the functions $F_{k,\tau}(\theta)$ is further restricted by a discrete-series factor SAIR $D^\pm$ through the values of σ_1, σ_2 to a quadrant or half-plane in Fig. 2, the region is delimited: A strong barrier[21] (i.e., an inverse-quadratic singularity with coefficient $\gamma_j\geqslant\frac{3}{4}$) or a weak barrier/well (of coefficient $-\frac{1}{4}\leqslant\gamma_j\leqslant\frac{3}{4}$) appears at the edge, according to whether $k\geqslant 1$ or $0<k<1$. For continuous series factor SAIR's, when two charts must be "joined" through the cancellation of the boundary Wronskians, the potential singularity which lies at the "common boundary" is a strong well[21] (i.e., of coefficient $\gamma_j\leqslant -\frac{1}{4}$) or a weak barrier/well according to whether $k_j\in C$ or $k_j\in E$.

A strong barrier at θ_0 imposes a choice of one of the two solutions of (6.12) through requiring that it be locally $\mathscr{L}^2$. A weak barrier/well singularity exhibits two locally $\mathscr{L}^2$-solutions, and allows for a choice: The solutions with locally $\mathscr{L}^2$-derivatives are to be taken for $\mathscr{D}$. A strong well at θ_0 also has two solutions, but no naturally distinguished ones. The proper cancellation of the boundary Wronskians will determine—up to constants in some cases—the total multichart solutions where the Casimir operator is to be self-adjoint and which must match (6.6). We shall defer further discussion on this aspect to Secs. IX and X, where this is done in detail for the cases at hand.

Lastly, the sign of the "energy" eigenvalue $(2k-1)^2$ should be noticed on the right-hand side of Eqs. (6.12). For continuous-series coupled SAIR's, $F^{H_>}$ and $F^{H_<}$ are "positive-energy" free states, while for $k\in\mathscr{D}$ they appear as "negative-energy" bound states in the well of the Pöschl–Teller potential. For the P^σ-chart, (6.12a), the roles are reversed and the discrete-series coupled SAIR's appear as positive-energy quantized states between potential barriers in the $D^+ \times D^+$ coupling. In the $D^+ \times \mathscr{C}$ and $\mathscr{C} \times \mathscr{C}$ couplings, the P^σ-chart must be matched with one or two H-charts, so it should not come as a surprise that in the P-chart the continuum C states appear under the D states. The $k = \frac{1}{2}$ level corresponds to zero "energy," and may in principle belong to $D^\pm_{1/2}$ or $C^\epsilon_{1/4}$ ($\epsilon\neq\frac{1}{2}$). Couplings to the exceptional continuous SAIR E can only appear as levels between values 0 and -1 of the H-chart "energy" $-(2k-1)^2$; they may be recognized through seeing that $|\epsilon| < \min(k, 1-k)$.

6.8: The general solution for the Pöschl–Teller Schrödinger equations (6.12) can be written as a linear combination of two of the fundamental solutions of the hypergeometric equation. In order to use uniformly the nomenclature of Ref. 19, p. 563, we may define, in each chart, the variable Θ as:

(i) P-charts, $\theta\in[0,\pi/2]$:

$$\Theta = \theta, \quad \sin\Theta = \sin\theta = r_2/r, \quad \cos\Theta = \cos\theta = r_1/r; \tag{6.13a}$$

(ii) $H_>$-charts, $\theta\in[0,\infty)$:

$$\Theta = i\theta, \quad \sin\Theta = i\sinh\theta = ir_2/r,$$

$$\cos\Theta = \cosh\theta = r_1/r; \tag{6.13b}$$

(iii) $H_<$-charts, $\theta\in[0,\infty)$:

$$\Theta = i\theta + \pi/2, \quad \sin\Theta = \cosh\theta = r_2/r,$$

$$\cos\Theta = -i\sinh\theta = -ir_1/r. \tag{6.13c}$$

The solutions of (6.12) can now be written in terms of the Gauss hypergeometric series expansions around $\Theta = 0,\pi/2,\infty$:

$$w^{k_1,k_2,k}_{1(0)}(\Theta) = w^{1-k_1,k_2,k}_{1(0)}(\Theta) = w^{k_1,k_2,1-k}_{1(0)}(\Theta)$$
$$= |\cos\Theta|^{2k_1-1/2}|\sin\Theta|^{2k_2-1/2}$$
$$\times F\left[\begin{matrix}k_1+k_2-k, k_1+k_2+k-1\\ 2k_2\end{matrix};\sin^2\Theta\right], \tag{6.14a}$$

$$w^{k_1,k_2,k}_{2(0)}(\Theta) = w^{k_1,1-k_2,k}_{1(0)}(\Theta), \tag{6.14b}$$

$$w^{k_1,k_2,k}_{1(1)}(\Theta) = w^{k_1,1-k_2,k}_{1(1)}(\Theta) = w^{k_1,k_2,1-k}_{1(1)}(\Theta)$$
$$= w^{k_2,k_1,k}_{1(0)}(\Theta \pm \pi/2)$$
$$= |\cos\Theta|^{2k_1-1/2}|\sin\Theta|^{2k_2-1/2}$$
$$\times F\left[\begin{matrix}k_1+k_2-k, k_1+k_2+k-1\\ 2k_1\end{matrix};\cos^2\Theta\right], \tag{6.15a}$$

$$w^{k_1,k_2,k}_{2(1)}(\Theta) = w^{1-k_1,k_2,k}_{1(1)}(\Theta), \tag{6.15b}$$

$$w^{k_1,k_2,k}_{1(\infty)}(\Theta) = w^{k_1,1-k_2,k}_{1(\infty)}(\Theta) = S_C^{2k_1-1}w^{1-k_1,k_2,k}_{1(\infty)}(\Theta)$$
$$= S_C^{-k_1-k_2+k}w^{k_2,k_1,k}_{1(\infty)}(\Theta \pm \pi/2)$$
$$= |\cos\Theta|^{2k_1-1/2}|\sin\Theta|^{-2k_1+2k-1/2}$$
$$\times F\left[\begin{matrix}k_1+k_2-k, k_1-k_2-k+1\\ 2-2k\end{matrix};\csc^2\Theta\right], \tag{6.16a}$$

$$w^{k_1,k_2,k}_{2(\infty)}(\Theta) = w^{1-k_1,1-k_2,1-k}_{1(\infty)}(\Theta) = S_C^{2k_1-1}w^{k_1,k_2,1-k}_{1(\infty)}(\Theta). \tag{6.16b}$$

In the last two equations, $S_C = +1$ for $C = H_>, H_<$, while $S_C = -1$ for $C = P$; the multivaluation is due to the Gauss function being evaluated on the branch cut (these cases will not be used, however). An identity due to a Kummer transformation (Ref. 19, Eq. 15.3.4) is

$$w^{k_1,k_2,k}_{n(\infty)}(\theta_{H_<}) = w^{k_2,k_1,k}_{n(\infty)}(\theta_{H_>}), \quad n = 1,2, \tag{6.16c}$$

between the hyperbolic cases, which exchanges $\cosh\theta \leftrightarrow \sinh\theta$ in the first two factors, and $1/\cosh^2\theta \leftrightarrow -1/\sinh^2\theta$ in the Gauss function argument. The absolute value of $\cos\Theta$ and $\sin\Theta$ has been placed on the first two factors of (6.14)–(6.16) in order to have real solutions for the $\mathscr{C}$-cases. If $2k_1$, $2k_2$, and/or $2k$ are integers, one of each pair of solutions is degenerate, and the logarithmic solution should replace it (these cases will also never be used).

In the P-charts ($\Theta\in[0,\pi/2]$) the argument of the hypergeometric functions in (5.16) ranges, for $w_{n(0)}$, $n = 1,2$, over $[0,1]$, and for $w_{n(1)}$, over $[1,0]$; for $w_{n(\infty)}$ however, the range is $(\infty,1]$, which lies on the branch cut of the Gauss function. The latter pair of solutions, therefore, will be avoided in describing solutions in the P^σ-chart. Similarly, and for the same reason, in the $H_>$-charts ($\Theta\in[0,i\infty)$) we shall avoid the $w_{n(1)}$ functions, and in the $H_<$-charts ($\Theta\in[\pi/2, \pi/2+i\infty)$) we avoid the $w_{n(1)}$ functions.

Elementary properties of these solutions which will be used time and again are their behavior near the expansion centers: $w^{k_1,k_2,k}_{1(0)}(\Theta) \sim \Theta^{2k_2-1/2}$ for $\Theta\to 0^+$, $w^{k_1,k_2,k}_{1(1)}(\Theta) \sim (\pi/2-\Theta)^{2k_1-1/2}$ for $\Theta\to\pi/2^-$, and $w^{k_1,k_2,k}_{1(\infty)}(\Theta) \sim (e^\Theta/2)^{2k-1/2}$ for $\Theta = i\theta$ or $i\theta+\pi/2$ when $\theta\to\infty$. The behavior of $w^{k_1,k_2,k}_{2(N)}(\Theta)$ can be obtained through exchanging $k_2\leftrightarrow 1-k_2$, $k_1\leftrightarrow 1-k_1$, and $k\leftrightarrow 1-k$, respectively.

6.9: In the following sections we shall use the above solutions to build the properly normalized product state through the condition that they be (Dirac-) orthonormal under the inner product (6.10a):

$$(\Psi^{k_1,k_2}_{k,\tau,\rho}, \Psi^{k_1,k_2}_{k',\tau',\rho'})_{\mathscr{S}^2} = \delta_{\tau,\tau'}\delta(\rho^2/2-\rho'^2/2)\delta[k,k'], \tag{6.17a}$$

$$\delta[k,k'] = \begin{cases}\delta_{k,k'}, & k,k'\in D,\\ \delta(\kappa-\kappa'), & k=(1+i\kappa)/2\in C \text{ or } k=(1+\kappa)/2\in E,\\ 0, & \text{otherwise.}\end{cases} \tag{6.17b}$$

The factorization (6.11) placed in (6.10) and the above requirement lead to the (Dirac-) orthonormality condition

$$(F_{k,\tau}, F_{k',\tau'})_{s^2} = \delta_{\tau,\tau'}\delta[k,k'], \tag{6.17c}$$

where we define the reduced inner product over θ to be

$$(f,g)_{s^2} = (f^P,g^P)_p + (f^{H_>},g^{H_>})_h + (f^{H_<},g^{H_<})_h, \tag{6.18a}$$

$$(a,b)_p = \int_0^{\pi/2} d\theta\, a(\theta)^* b(\theta), \tag{6.18b}$$

$$(a,b)_h = \int_0^{\infty} d\theta\, a(\theta)^* b(\theta). \tag{6.18c}$$

Here again, $f^C(\theta)$ ($C = P^\sigma, H^\sigma_>, H^\sigma_<$) gives the three forms of a single function f over the support of the coupled state (i.e., the line $\sigma_1 r_1^2 + \sigma_2 r_2^2 = \sigma r^2$ in Fig. 2). The inner product $(\cdot,\cdot)_{s^2}$ will thus in general involve integration over more than one θ range.

6.10: In essence, the solution of Eqs. (6.14)–(6.18) is the solution of the Clebsch–Gordan problem for so(2,1), for once the proper $F^C_{k,\tau}(\theta)$ are found, the CGC's are obtained as the inner product between the coupled states (6.11) and the product states (6.3). The latter have support on a single point in the plane $\mathscr{S}^2$ of Fig. 2, which due to (6.5) lies on the line support of the coupled state. The CGC's are thus

$$C\begin{pmatrix}k_1, & k_2 & ;k\\ \tau_1,\rho_1, & \tau_2,\rho_2 & ;\tau\rho\end{pmatrix} = (\psi^{k_1,k_2}_{\tau_1,\rho_1,\tau_2,\rho_2}, \Psi^{k_1,k_2}_{k,\tau,\rho})_{\mathscr{S}^2}$$
$$= \delta(\tau_1\rho_1^2/2 + \tau_2\rho_2^2/2 - \tau\rho^2/2)(\rho_1\rho_2)^{-1/2}F^C_{k,\tau}(T), \tag{6.19}$$

where C refers to the chart where $(\tau_1,\rho_1,\tau_2,\rho_2)$ lies, according to (6.8a), and T is then given by

$$P^\tau\text{-chart:} \quad T = \operatorname{arcsin}(\rho_2/\rho) = \operatorname{arccos}(\rho_1/\rho), \tag{6.20a}$$

$$H^\tau_>\text{-chart:} \quad T = \operatorname{arcsinh}(\rho_2/\rho) = \operatorname{arccosh}(\rho_1/\rho), \tag{6.20b}$$

$$H^\tau_<\text{-chart:} \quad T = \operatorname{arccosh}(\rho_2/\rho) = \operatorname{arcsinh}(\rho_1/\rho). \tag{6.20c}$$

6.11: One obvious symmetry relation between the CGC's comes from the automorphism $A = A_1A_2$, whose role is to invert all σ's. Because of the factors δ_{σ_j,τ_j} and $\delta_{\sigma,\tau}$ in (6.3) and (6.11) and the unitarity of A under (6.10), it follows that

$$C\begin{pmatrix}k_1, & k_2 & ;k\\ \tau_1,\rho_1, & \tau_2,\rho_2 & ;\tau,\rho\end{pmatrix} = C\begin{pmatrix}k_1, & k_2 & ;k\\ -\tau_1,\rho_1, & -\tau_2,\rho_2, & ;-\tau,\rho\end{pmatrix} \tag{6.21}$$

for all SAIR couplings. A special case of this symmetry was noted in (4.13). We shall henceforth reduce ourselves to the calculation of CGC's inequivalent under this relation.

VII. THE COUPLING $D^+ \times D^+$

The coupling of two lower-bound (or upper-bound) discrete-series SAIR's into a direct sum of representations of the same type, constitutes the simplest application of the general method outlined in the preceding section.

7.1: Given the range of the $\mathscr{S}^2$ variables ($\sigma_1 = 1, \sigma_2 = 1$ and hence $\sigma = 1$) the coupled function $F_{k,\tau}(\theta)$ in (6.11) will have support only in the P^+-chart (6.8a). For $k_j \geqslant 1$, in the D^+ SAIR's outside the exceptional interval the Pöschl–Teller potential Schrödinger equation (Fig. 3) exhibits strong barriers at the interval ends, and the square-integrable solutions to (6.12a) are $w^{k_1,k_2,k}_{1(0)}(\theta)$ [Eq. (6.14a)] and $w^{k_1,k_2,k}_{1(1)}(\theta)$ [Eq. (6.15a)] at each of the P-chart endpoints. They represent the same solution only when $-k_1 - k_2 + k$ is zero or a positive integer, and this determines through (6.6) that only discrete-series SAIR's appear in the reduction of $D^+_{k_1} \times D^+_{k_2}$. The hypergeometric function in the solution is a polynomial in $\sin^2\theta$, and is easily converted into a Jacobi polynomial (Ref. 19, Eq. 15.4.6) $P^{(2k_2-1,2k_1-1)}_{k-k_1-k_2}(\cos 2\theta)$, multiplied by powers of $\cos\theta$ and $\sin\theta$ as required to provide an orthogonal set of solutions.

When k_1 and/or k_2 lie on the exceptional interval (0,1), the strong barrier at $\theta = \pi/2$ and/or 0 becomes a weak barrier/well, and any linear combination of $w^{k_1,k_2,k}_{1(N)}(\theta)$ and $w^{k_1,k_2,k}_{2(N)}(\theta)$ [N indicating the end point at which such a weak singularity occurs] would provide square-integrable solutions, thereby destroying the unique quantization of k. Equation (6.6), however, fixes the self-adjoint extension of the operators and hence the spectrum of the Pöschl–Teller Schrödinger equation to $k = k_1 + k_2 + n$, $n \in Z^+$. For all $k > 0$, hence, the solution vanishes at the end points, and its derivative is square integrable. The proper normalization constants for the solutions are easily found (Ref. 19, Eq. 22.2.1), and the rest of the program proceeds through (6.19) and (6.20a).

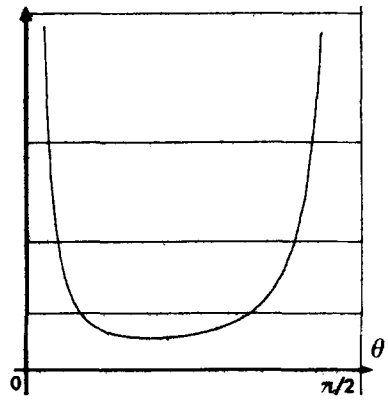


FIG. 3. Pöschl–Teller potential of the first kind for the coupling of $D^+ \times D^+$. The "energy levels" represent the direct summands in the product.

7.2: The CGC's for $D^+ \times D^+ \to D^+$ can be thus written as

$$C\begin{pmatrix} k_1, & k_2 & ; k \\ +1,\varphi_1, & +1,\varphi_2; & +1,\varphi \end{pmatrix} = \delta(\rho_1^2/2 + \rho_2^2/2 - \rho^2/2)\, p_k^{k_1,k_2} \rho_1^{2k_1-1} \rho_2^{2k_2-1} \rho^{1-2k_1-2k_2} P^{(2k_2-1,2k_1-1)}_{k-k_1-k_2}([\rho_1^2 - \rho_2^2]/\rho^2)$$

$$= \delta(\rho_1^2/2 + \rho_2^2/2 - \rho^2/2)\, c_k^{k_1,k_2} \rho_1^{2k_1-1} \rho_2^{2k_2-1} \rho^{1-2k_1-2k_2} F\left[\begin{matrix} k_1+k_2-k, k_1+k_2+k-1 \\ 2k_2 \end{matrix}; \rho_2^2/\rho^2\right], \tag{7.1a}$$

where the normalization coefficients are

$$p_k^{k_1,k_2} = \left[\frac{2(2k-1)\Gamma(k_1+k_2+k-1)\Gamma(-k_1-k_2+k+1)}{\Gamma(k_1-k_2+k)\Gamma(-k_1+k_2+k)}\right]^{1/2}, \tag{7.1b}$$

$$c_k^{k_1,k_2} = \frac{1}{\Gamma(2k_2)}\left[\frac{2(2k-1)\Gamma(-k_1+k_2+k)\Gamma(k_1+k_2+k-1)}{\Gamma(k_1-k_2+k)\Gamma(-k_1-k_2+k+1)}\right]^{1/2}, \tag{7.1c}$$

and where the range of the coupled SAIR's is

$$k = k_1 + k_2 + n, \quad n \in Z^+. \tag{7.1d}$$

Note that if we set k_1 and k_2 to be $\frac{1}{4}$ or $\frac{3}{4}$, we obtain the $D^+_{\circ} \times D^+_{\circ}$ coupling coefficients (2.18).

For the $D^- \times D^-$ coupling the developments are the same, except that now $\sigma_1 = -1 = \sigma_2$ and hence $\sigma = -1$, so we are in the P^--chart. Through (6.21) we obtain CGC's identical to the above ones.

VIII. THE COUPLING $D^+ \times D^-$

The coupling of a positive and a negative discrete SAIR follows the general pattern of Sec. III. Here, the decomposition yields a direct integral of continuous nonexceptional representations plus, depending on whether $k_1 > k_2$ or $k_1 < k_2$, a finite sum of positive or negative discrete-series SAIR's.

8.1: As now $\sigma_1 = +1$ and $\sigma_2 = -1$ we have functions with support on the lower-right quadrant of Fig. 2, i.e., on the union of the $H^+_>$ and $H^-_<$ coordinate charts. Furthermore, depending on whether $\rho_1 > \rho_2$ or $\rho_1 < \rho_2$, the support lies entirely within one or the other. We have thus to satisfy Eqs. (6.12b) or (6.12c) separately. These are two Schrödinger equations for Pöschl–Teller potentials of the second kind, shown in Fig. 4. The singularity coefficient at the origin of the equation belonging to the $\sigma = +1$ chart depends on k_2, and that of the $\sigma = -1$ chart, on k_1. When $k_j > 1$, as in the

figures, the potential has a strong barrier at the origin. When one or both k_j lie on the exceptional interval, one or both of the two equations will present a weak barrier/well at the origin. In every case, a continuum of "free" generalized eigenfunctions exists.

When $k_1 > k_2 > \frac{3}{4}$ (Fig. 4), there is a minimum in the $\sigma = +1$ chart potential. When this happens, the quantum system exhibits a finite number of bound states. When either $k_2 < \frac{3}{4}$ (for $\sigma = +1$) or $k_1 < \frac{3}{4}$ (for $\sigma = -1$), the minimum shifts into the weak well at the origin, which continues, nevertheless, to have a finite number of bound states. The "negative-energy" bound states correspond to discrete-series coupled functions since their eigenvalue under J_-, i.e., $\sigma\rho^2/2$, is purely positive ($k_1 > k_2$) or purely negative ($k_1 < k_2$), telling us they belong to the D_k^+ or D_k^- SAIR's respectively. When $k_1 = k_2$, the $H_>^+$ and $H_<^-$ chart equations are identical. No minima exist. In the continuum, the "energy" eigenvalue in (6.8b) and (6.8c) is positive, i.e., $-(2k-1)^2 \geqslant 0$ so that the coupled states built with these solutions belong to the continuous nonexceptional series $k = (1 + i\kappa)/2$, $\kappa > 0$. The zero eigenvalue will be subject to further scrutiny.

8.2: We begin with the coupling to the continuous series in the $H_>^+$ chart, $\rho_1 > \rho_2$. The solutions $w_{1(0)}^{k_1,k_2,k}(\theta)$ in (6.13b)–(6.14a) are the appropriate ones, for they vanish at the origin when $k_2 \geqslant 1$ and have locally $\mathscr{L}^2$ derivatives when k_2 lies in the exceptional interval, as required by the function domain of the operators. Since these solutions oscillate at infinity, the Dirac-normalization constant may be obtained through a process parallel to (4.9), so that (6.17) may be satisfied. The rest of the program follows through (6.19) and (6.20b).

The CGC's for the $D^+ \times D^- \to C$ coupling are thus

$$C_> \begin{pmatrix} k_1, & k_2 & ;\epsilon k \\ +1,\rho_1, & -1,\rho_2; & +1,\rho \end{pmatrix} = \delta(\rho_1^2/2 - \rho_2^2/2 - \rho^2/2) \times c_k^{k_1,k_2} \rho_1^{2k_1-1} \rho_2^{2k_2-2} \rho^{1-2k_1-2k_2} \times F\left[\begin{matrix} k_1+k_2-k, k_1+k_2+k-1 \\ 2k_2 \end{matrix} ; -\rho_2^2/\rho^2 \right], \quad (8.1a)$$

where we have indicated that $\rho_1 > \rho_2$ through the ">" subindex, which will be henceforth used for similar cases, and where

$$c_k^{k_1,k_2} = [\pi\Gamma(2k_2)]^{-1} \times [\kappa \sinh \pi\kappa \Gamma(k_1+k_2-k)\Gamma(-k_1+k_2+k) \times \Gamma(k_1+k_2+k-1)\Gamma(-k_1+k_2-k+1)]^{1/2}. \quad (8.1b)$$

As usual, we have used $k = (1 + i\kappa)/2$ and we imply that

$$\epsilon \equiv \epsilon_1 + \epsilon_2 \equiv k_1 - k_2 \pmod 1. \quad (8.2)$$

We remind the reader that, for the discrete series, as detailed in Sec. VI, $\epsilon_1 \equiv k_1 \pmod 1$ in $D_{k_1}^+$ and $\epsilon_2 \equiv -k_2 \pmod 1$ in $D_{k_2}^-$. The sign is irrelevant only when the ϵ_j's are 0 and/or $\frac{1}{2}$. Finally, if we set $k_1, k_2 = \frac{1}{4}$ or $\frac{3}{4}$ in (8.1), we reproduce exactly the CGC's for the coupling of two oscillator representations given in (4.11).

When $\rho_1 < \rho_2$ we are in the $H_<^-$-chart with $\sigma = -1$. This case leads to a similar differential equation—(6.12b) vs (6.12c)— with k_1 and k_2 exchanged, as well as ρ_1 and ρ_2—(6.13b) vs (6.13c)—in all the ensuing developments. We thus find immediately

$$C_< \begin{pmatrix} k_1, & k_2 & ;\epsilon k \\ +1,\rho_1, & -1,\rho_2; & -1,\rho \end{pmatrix} = C_> \begin{pmatrix} k_2, & k_1 & ;\epsilon k \\ +1,\rho_2, & -1,\rho_1; & +1,\rho \end{pmatrix}. \quad (8.3)$$

8.3: We turn finally to the discrete series present in the decomposition of $D^+ \times D^-$. Consider first $k_1 > k_2$, when the potential has a minimum in the $H_>^+$-chart, as in Fig. 4(a). There is a unique self-adjoint extension of the operator in (6.12b) determined by (8.2) and, consequently, unique quantization. Analysis of the asymptotic form of the solution in (6.13b)–(6.14a) shows that $w_{1(0)}^{k_1,k_2,k}(\theta)$ decays exponentially for $k = k_1 - k_2, k_1 - k_2 - 1, \cdots > \frac{1}{2}$ (i.e., down to 1 or $\frac{1}{2}$ when $k_1 - k_2$ is an integer or half-integer). When $k = \frac{1}{2}$ (i.e., for $k_1 - k_2$ half-integer) $w_{1(0)}^{k_1,k_2,k}(\theta)$ tends asymptotically to a constant and hence is not in $\mathscr{L}^2(R^+)$. We conclude that a finite number of discrete representations occur for this range of values of k. Since $\sigma = +1$ in the $H_>^+$-chart, and no corresponding bound states appear for $\sigma = -1$ in the $H_<^-$-chart, we confirm (8.2) in that the representations belong to the D^+-series. The bound and free states in the $H_>^+$-chart are orthogonal. A Kummer transformation on (6.13b)–(6.14a) for $k - k_1 + k_2 = -n$, $n \in Z^+$, informs us that the hypergeometric function is a Jacobi polynomial of degree n in $-\sinh^2\theta$, $\theta \in [0, \infty)$. We may use Bargmann's hint [Ref. 1, Eq. (10.26)] to evaluate the proper normalization coefficient and finally express the resulting series in terms of a Jacobi polynomial in $1 - 2\tanh\theta \in [-1, 1]$. We follow with (6.19) and (6.20b).

The CGC's for the $D^+ \times D^- \to D^+$ coupling are thus

$$C_v \begin{pmatrix} k_1, & k_2 & ;k \\ +1,\rho_1, & -1,\rho_2; & +1,\rho \end{pmatrix} = \delta(\rho_1^2/2 - \rho_2^2/2 - \rho^2/2) p_k^{k_1,k_2} \rho_1^{-k_1-k_2-k+1} \rho_2^{2k_2-1} \rho^{k_1-k_2+k-1} P_{k_1-k_2-k}^{2k_2-1,2k-1}([\rho^2 - \rho_2^2]/\rho_1^2)$$

$$= \delta(\rho_1^2/2 - \rho_2^2/2 - \rho^2/2) c_k^{k_1,k_2} \rho_1^{-2k_1+1} \rho_2^{2k_2-1} \rho^{2k_1-2k_2-1} F\left[\begin{matrix} -k_1+k_2+k, -k_1+k_2-k+1 \\ 2k_2 \end{matrix} ; -\rho_2^2/\rho^2 \right], \quad (8.4a)$$

$$p_k^{k_1,k_2} = \left[\frac{2(2k-1)\Gamma(k_1+k_2+k-1)\Gamma(k_1-k_2-k+1)}{\Gamma(k_1+k_2-k)\Gamma(k_1-k_2+k)} \right]^{1/2}, \quad (8.4b)$$

$$c_k^{k_1,k_2} = \frac{1}{\Gamma(2k_2)} \left[\frac{2(2k-1)\Gamma(k_1+k_2-k)\Gamma(k_1+k_2+k-1)}{\Gamma(k_1-k_2+k)\Gamma(k_1-k_2-k+1)} \right]^{1/2}, \quad (8.4c)$$

$$k = k_1 - k_2, k_1 - k_2 - 1, \cdots > \tfrac{1}{2}. \quad (8.4d)$$

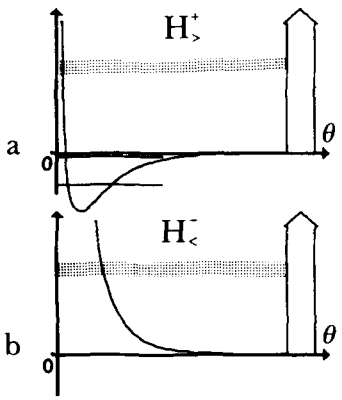


FIG. 4. Pöschl–Teller potentials of the second kind representing the coupling $D^+ \times D^-$ in (a) the $H^+_>$ chart and (b) the $H^-_<$ chart. The arrows indicate the coupling to the continuous series, while the discrete levels in (a) indicate the discrete-series summands in the coupling (a finite set).

The powers and hypergeometric function in (8.4a) and (8.1a) are the same, as can be seen through a Kummer transformation. The normalization coefficients are different, however.

When $k_1 < k_2$, the "bound states" appear in the $H^-_<$-chart, with $\sigma = -1$, and the same arguments which we used for (8.3) apply here. We thus conclude that the D^-_k SAIR's occur with CGC's related to the above through

$$C_< \begin{pmatrix} k_1, & k_2 & ;k \\ +1,\rho_1, & -1,\rho_2; & -1,\rho \end{pmatrix} = C_> \begin{pmatrix} k_2, & k_1 & ;k \\ +1,\rho_2, & -1,\rho_1; & +1,\rho \end{pmatrix}, \quad k_1 < k_2. \tag{8.5}$$

8.4: We would like to clarify now the appearance of the $k = \frac{1}{2}$ SAIR's in the $D^+ \times D^-$ coupling. They did not appear, we saw, among the "bound states" producing (8.4). The normalization coefficient (8.1b) appears to be zero when $k = (1 + i\kappa)/2$, $\kappa \to 0$; there is a cancelling pole, however, when either $k_1 + k_2 = \frac{1}{2}$ or when $k_1 - k_2 = \frac{1}{2}$ [the $k_1 - k_2 = -\frac{1}{2}$ case appears through (8.3)]. The former case applied to both σ-charts exhibiting $\epsilon \equiv \epsilon_1 + \epsilon_2 \equiv k_1 - k_2$ (mod 1), $\epsilon \in (-\frac{1}{2}, \frac{1}{2})$, while the latter cases exhibit $\epsilon = \frac{1}{2}$ for $\sigma = +1$ only and $\epsilon = -\frac{1}{2}$ for $\sigma = -1$ only. Hence $k_1 + k_2 = \frac{1}{2}$ represents a nonzero coupling to the continuous series $C^{\epsilon}_{1/4}$ for $\epsilon \in (-\frac{1}{2}, \frac{1}{2})$ (belonging to the continuum of Fig. 4), while $k_1 - k_2 = \frac{1}{2}$, a coupling to $D^+_{1/2}$ belonging to the discrete series. The latter appeared, we recall, in $D^+_o \times D^-_o$ of Sec. IV.

IX. THE COUPLING $\mathscr{D} \times \mathscr{C}$

The analysis of the coupling of one discrete- and one continuous-series representation requires a more careful analysis: it involves three of the charts in Fig. 2, two of them joined as described briefly in Sec. VI.

9.1: We analyze first the $D^+ \times \mathscr{C}$ case, where $\sigma_1 = +1$ and $\sigma_2 = \pm 1$, so the relevant portion of Fig. 2 is the right half-plane constituted by the P^+-, $H^+_>$-, and $H^-_<$-charts. The $r_1 = 0$ axis carries a strong potential barrier (for $k_1 > 1$) or a weak barrier/well (for $0 < k_1 < 1$) in the associated Pöschl–Teller potentials; the $r_2 = 0$ axis carries a similar singularity depending on k_2: a strong well for $k_2 \in C$ and a weak barrier/well for $k_2 \in E$.

As described in Sec. VI solutions in the P^+- and $H^+_>$-charts must join through the cancellation of the boundary Wronskians, while the solutions in the $H^-_<$-chart remain disconnected, as shown in Fig. 5.

9.2: We treat first the latter case [Fig. 5(a), where the $H^-_<$ coordinates are (6.8c), the inner product is (6.10c), and the Pöschl–Teller potential is (6.12c), exhibiting only "free-state" solutions $w^{k_1,k_2,k}_{1(0)}(\theta)$ given by positive $-(2k-2)^2 > 0$, which represent couplings to the C-series and vanish, or are locally $\mathscr{L}^2$ at the origin]. These elements and boundary conditions were present in the derivation of the $D^+ \times D^-$ coupling in the same $H^-_<$-chart, in the last section. The result, Eq. (8.3), yields the $D^+ \times \mathscr{C} \to C$ CGC's, which we write explicitly as

$$C_< \begin{pmatrix} k_1, & \epsilon_2, k_2 & ;\epsilon, k \\ +1,\rho_1, & -1,\rho_2; & -1,\rho \end{pmatrix} = \delta(\rho_1^2/2 - \rho_2^2/2 + \rho^2/2) \times c_k^{k_1,k_2} \rho_1^{2k_1 - 1} \rho_2^{2k_2 - 1} \rho^{-2k_1 - 2k_2 + 1} \times F\left[\begin{matrix} k_1 + k_2 - k, k_1 + k_2 + k - 1 \\ 2k_1 \end{matrix} ; -\rho_1^2/\rho^2 \right], \tag{9.1a}$$

$$c_k^{k_1,k_2} = [\pi \Gamma(2k_1)]^{-1} [\kappa \sinh \pi\kappa \times \Gamma(k_1 + k_2 - k)\Gamma(k_1 - k_2 + k)\Gamma(k_1 + k_2 + k - 1) \times \Gamma(k_1 - k_2 - k + 1)]^{1/2}, \tag{9.1b}$$

$$\epsilon \equiv \epsilon_1 + \epsilon_2 \equiv k_1 + \epsilon_2 \pmod 1, \tag{9.1c}$$

where again $k = (1 + i\kappa)/2$, but with the addendum that $k_2 = (1 + i\kappa_2)/2$ for $k_2 \in C$ or $k_2 = (1 + \kappa_2)/2, 0 < \kappa_2 < 1$ for $k_2 \in E$ [thus ϵ_2 in (9.1c) replaces k_2 in (8.2c).] The normalization coefficient (9.1b) is, of course, real.

9.3: The new feature which the $\mathscr{D} \times \mathscr{C}$ coupling introduces over previous cases is the form of the solution in the $\sigma = +1$ region, constituted by the union of the P^+- and $H^+_>$-charts. The inner product will contain the first two summands in (6.10a), where the coupled states will have one form in each chart, eigenfunctions of the Casimir operator, itself having forms (6.9a) and (6.9b) in each of the two charts,

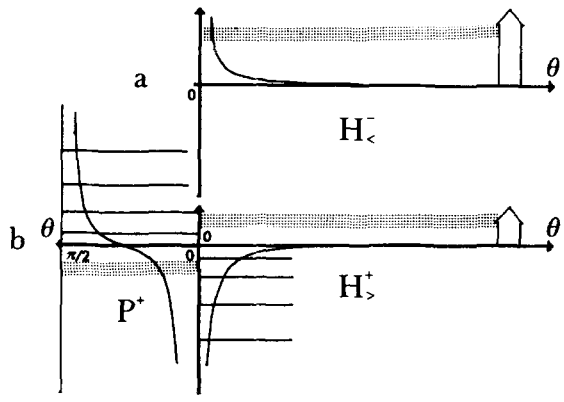


FIG. 5. The $\mathscr{D} \times \mathscr{C}$ coupling in (a) the $H^-_<$ chart and (b) in the region composed of the P^+ and $H^+_>$ charts. The former contains only coupling to the continuous series, while the latter includes both continuous and discrete series.

with a single eigenvalue $k(1-k)$.

The two Pöschl–Teller potentials are shown in Fig. 5(b). They do not join, but the two parts of the figure have only been placed so that the boundary Wronskians at adjacent points are indicated to cancel [see Eq. (9.4) ahead]. We shall consider first the $D^+\times\mathscr{C}\to D^+$ coupling to the discrete series, which corresponds to bound ("negative-energy") states in the $H_>^+$-chart potential, joined with a "positive-energy" state in the P^+-chart potential, due to the difference in the sign of the right-hand side of (6.12). The barrier due to the $D_{k_1}^+$ factor closes the system from the left, while on the right the potential rises exponentially to zero. A strong well or weak barrier/well due to the $C_{q_2}^{\epsilon}$ factor lies at the join of the two regions. The cases when k_1 and/or k_2 lie in the exceptional interval are important, but constitute minor modifications of the basic construction. In the $H_>^+$-chart, the rising potential imposes one constraint and reduces the solution there to be $f_k^{H_>}(\theta)=C_k^{H_>}w_{2(\infty)}^{k_1,k_2,k}(\theta)$, with an as yet undetermined constant $C_k^{H_>}$. A second constraint in the P^+-chart comes from square-integrability at $\theta_P=0$, and similarly for the derivative, reducing the solution there to $f_k^P(\theta)=C_k^P w_{1(1)}^{k_1,k_2,k}(\theta)$, with another constant C_k^P.

The inner product (6.18) of such two-chart functions, corresponding to eigenvalues $k(1-k)$ and $l(1-l)$ can be evaluated, as was done in (5.8), in terms of the boundary Wronskians:

$$\begin{aligned}(f_k,f_l)_{s^2}&=(f_k^P,f_l^P)_P+(f_k^{H_>},f_l^{H_>})_h\\&=[4(k+l-1)(l-k)]^{-1}\left[C_k^{P*}C_l^P W(w_{1(1)}^{k*},w_{1(1)}^{l})\,|_0^{\pi/2}\right.\\&\left.-C_k^{H_>*}C_l^{H_>}W(w_{2(\infty)}^{k*},w_{2(\infty)}^{l})\,|_0^{\infty}\right].\end{aligned}\tag{9.2}$$

This expression should equal $\delta_{k,l}$, and thus it must be zero for $k\neq l$; the two nonzero boundary values at $\theta_P=0$ and $\theta_{H_>}=0$ must therefore be equal in order to cancel. The relative minus sign between the Wronskians is due to the sign difference in the eigenvalue in (6.12). Since (Ref. 19, Eq. 15.3.6)

$$w_{1(1)}^{k_1,k_2,k}(\theta)=A_k w_{1(0)}^{k_1,k_2,k}(\theta)+\tilde{A}_k w_{2(0)}^{k_1,k_2,k}(\theta),\tag{9.3a}$$

$$\begin{aligned}A_k&=\Gamma(2k_1)\Gamma(-2k_2+1)/\Gamma(k_1-k_2+k)\\&\quad\times\Gamma(k_1-k_2-k+1),\\\tilde{A}_k&=A_k(k_2\leftrightarrow 1-k_2),\end{aligned}\tag{9.3b}$$

and (Ref. 19, Eq. 15.3.7)

$$w_{2(\infty)}^{k_1,k_2,k}(\theta)=B_k w_{1(0)}^{k_1,k_2,k}(\theta)+\tilde{B}_k w_{2(0)}^{k_1,k_2,k}(\theta),\tag{9.4a}$$

$$\begin{aligned}B_k&=\Gamma(2k)\Gamma(-2k_2+1)/\Gamma(k_1-k_2+k)\\&\quad\times\Gamma(-k_1-k_2+k+1),\\\tilde{B}_k&=B_k(k_2\leftrightarrow 1-k_2),\end{aligned}\tag{9.4b}$$

we may evaluate the boundary Wronskians explicitly. We note that for $k_2\in C$ the two summands in (8.3a) are complex conjugates, and similarly in (8.4a), while for $k\in E$ they are all real. One obtains thus

$$W(w_{1(1)}^{k*},w_{1(1)}^{l})|_{\theta_P\to 0^+}=2(2k_2-1)(\tilde{A}_k A_l-A_k\tilde{A}_l),\tag{9.5a}$$

$$W(w_{2(\infty)}^{k*},w_{2(\infty)}^{l})\,|_{\theta_{H_>}\to 0^+}=2(2k_2-1)(\tilde{B}_k B_l-B_k\tilde{B}_l).\tag{9.5b}$$

When the ratio of the latter two is asked to factor into $(C_k^P/C_k^{H_>})^*(C_l^P/C_l^{H_>})$, one finds that this does not generally happen. The requirement may be imposed only when $k-l$ is an integer. The domain restriction (9.1c) implies then that $k=\epsilon,\epsilon+1,\epsilon+2,\cdots$ and similarly for l. Hence, one finds

$$\begin{aligned}C_k^P/C_k^{H_>}&=\pi^{-1}[\sin\pi(k_2+\epsilon_2)\sin\pi(k_2-\epsilon_2)]^{1/2}\\&\quad\times\Gamma(2k_1)^{-1}\Gamma(2k)\Gamma(k_1+k_2-k)\\&\quad\times\Gamma(k_1-k_2-k+1).\end{aligned}\tag{9.6}$$

From (9.1c), this ratio is real. We also note the close resemblance between the radicands in (9.6) and in (5.9). Indeed, it tells us that when $k_2\in E$ is such that $\pm\epsilon_2\to k_2$, the ratio (9.3) develops a zero or a pole, so that the P- and $H_>$-function pieces become independent and reduce each to the $D^+\times D^+$ or $D^+\times D^-$ couplings seen in Secs. VII and VIII.

The last step is to normalize the function f_k so that $(f_k,f_k)_{s^2}=1$. This is done on (9.2) in the manner of (3.14), which essentially reduces to obtaining the derivative of the two boundary terms with respect to l and valuate at $l=k$. (The boundary terms cancel for $k-l$ integer, but lead separate lives for all other $l-k$; hence the derivative of the boundary term difference is not zero.) The resulting ψ functions combine to form trigonometric ones, and the result simplifies drastically. Use of (6.20a) or (6.20b) for the corresponding chart leads to the CGC's as in (6.19).

The CGC's for the $D^+\times\mathscr{C}\to D^+$ coupling are thus

$$C\begin{pmatrix}k_1, & \epsilon_2,k_2 & ;k\\ +1,\rho_1, & +1,\rho_2; & +1,\rho\end{pmatrix}=\delta(\rho_1^2/2+\rho_2^2/2-\rho^2/2)\times c_{k,P}^{k_1,k_2}\rho_1^{2k_1-1}\rho_2^{2k_2-1}\rho^{-2k_1-2k_2+1}F\left[\begin{matrix}k_1+k_2-k,k_1+k_2+k-1\\ 2k_1\end{matrix};\rho_1^2/\rho^2\right],\tag{9.7a}$$

$$c_{k,P}^{k_1,k_2}=\frac{1}{\Gamma(2k_1)}\left[\frac{2(2k-1)\Gamma(k_1-k_2+k)\Gamma(k_1+k_2+k-1)}{\Gamma(-k_1+k_2+k)\Gamma(-k_1-k_2+k+1)}\right]^{+1/2},\tag{9.7b}$$

and

$$C_>\begin{pmatrix}k_1, & \epsilon_2,k_2 & ;k\\ +1,\rho_1, & -1,\rho_2; & +1,\rho\end{pmatrix}=\delta(\rho_1^2/2-\rho_2^2/2-\rho^2/2)\times c_{k,H}^{k_1,k_2}\rho_1^{-2k_1+1}\rho_2^{2k_1-2k-1}\rho^{2k-1}F\left[\begin{matrix}-k_1+k_2+k,-k_1-k_2+k+1\\ 2k\end{matrix};-\rho_2^2/\rho^2\right],\tag{9.8a}$$

$$c_{k,H}^{k_1,k_2}=\frac{1}{\Gamma(2k)}\left[\frac{2(2k-1)\Gamma(k_1-k_2+k)\Gamma(k_1+k_2+k-1)}{\Gamma(k_1+k_2-k)\Gamma(k_1-k_2-k+1)}\right]^{1/2},\tag{9.8b}$$

where it is implied that (8.1c) holds, and the subscript " $>$ " of the CGC menas that $\rho_1 > \rho_2$.

9.4: It should be noted that for $\frac{3}{4} > k_2 > 1$ the Casimir operator Q has a negative spectrum $k(1-k)$ and nonvanishing "wavefunctions," even though the Pöschl–Teller potential in the $H_>^+$-chart [Fig. 5(b)] is purely positive. Of course, the joined P^+-chart must also be considered, but this remark only underlines the fact that a physicist's intuition on Schrödinger equation solutions may go astray for two- and three-chart such operators.

In contradistinction to the matching of solutions in Schrödinger equations with finite discontinuous potentials, where the two sides of a single wavefunction and its derivative across the discontinuity are made equal, here we must only cancel the boundary Wronskians of a pair of functions on two charts. The functions themselves at $\theta \to 0^+$ behave as $A_k \theta^{1/2+i\kappa_2} + \tilde{A}_k \theta^{1/2-i\kappa_2}$, $i\kappa_2 = 2k_2 - 1$. They are real, tend to zero, but their derivative is not defined at the boundary. The boundary is thus not simply a point where the potential has a discontinuity/singularity as Figs. 2 and 5(b) might suggest at first sight.

9.5: We now examine the next case: the $D^+ \times \mathscr{C} \to C$ coupling to a continuous-series representation. The proper function on the P^+-chart continues to be $w^{k_1,k_2,k}_{1(1)}(\theta)$, with boundary Wronskian (9.5a). In the $H_>^+$-chart, however, we may have any linear combination of $w^{k_1,k_2,k}_{1(\infty)}(\theta)$ and $w^{k_1,k_2,k}_{2(\infty)}(\theta)$; the former behaves asymptotically as $\exp(i\kappa\theta)$ and the latter as $\exp(-i\kappa\theta)$. It can be verified that when the solution family is taken in the linear combination

$$f_k^{H_>}(\theta) = C_k w^{k_1,k_2,k}_{1(\infty)}(\theta) + C_k^* w^{k_1,k_2,k}_{2(\infty)}(\theta), \quad C_k^* = C_{1-k} \tag{9.9a}$$

[the second summand being the complex conjugate of the first: see Eq. (6.16b)], the Wronskian of any pair of functions $f_k^{H_>}$, $f_l^{H_>}$ in the family behaves asymptotically as the Fourier transform Dirichlet kernel, representing $2\pi|C_k|^2 \times [\delta(\lambda-\kappa) + \delta(\lambda+k)]$. This is analogous to (4.9) and provides the Dirac normalization.

The cancellation of the pair's Wronskian at $\theta_{H_>} \to 0$ with the companion P-chart Wronskian (9.5a) over a continuous range of k and l requires that $f_k^{H_>}(\theta)$ have, in terms of solutions around zero, the real form

$$f_k^{H_>}(\theta) = \alpha A_k w^{k_1,k_2,k}_{1(0)}(\theta) + \alpha^{-1} \tilde{A}_k w^{k_1,k_2,k}_{2(0)}(\theta), \tag{9.9b}$$

with α an as-yet free parameter, independent of k, subject only to the restriction $|\alpha| = 1$ for $k_2 \in C$ and $\alpha = \alpha^*$ for $k_2 \in E$. This freedom is allowed since the boundary Wronskian [Eq. (9.5b) with $B_k \to \alpha A_k$ and $\tilde{B}_k \to \alpha^{-1}\tilde{A}_k$] is quadratic in the coefficients of (9.9b) and independent of α. We may write the coefficients C_k in (9.9a) in terms of those in (9.9b) through (Ref. 19, Eq. 15.3.7)

$$C_k = C_{1-k}^* = \frac{\Gamma(2k-1)\Gamma(2k_1)}{\Gamma(k_1-k_2+k)\Gamma(k_1+k_2+k-1)}\left[\frac{\sin\pi(k-k_1)}{\sin\pi k_2}\,\frac{\alpha+\alpha^{-1}}{2} + \frac{\cos\pi(k-k_1)}{\cos\pi k_2}\,\frac{\alpha-\alpha^{-1}}{2}\right]. \tag{9.10a}$$

The condition to fix α turns out to be the necessary orthogonality between the continuous-k "free" generalized eigenstates and the discrete-k "bound" eigenstates seen before. Indeed, the Wronskians of two functions, each in one family, at $\theta_P = 0$ and $\theta_{H_>} = 0$, cancel only when

$$\alpha = [\sin\pi(k_2-\epsilon_2)/\sin\pi(k_2+\epsilon_2)]^{1/2}. \tag{9.10b}$$

Replacing this into (9.10a) and taking (9.1c) into account, we find the proper normalization constants for the two-chart functions. Finally, we use (6.19) with (6.20a) and (6.20b) for the rest of the process.

The CGC's for the $D^+ \times \mathscr{C} \to C$ coupling are thus

$$C\begin{pmatrix} k_1, & \epsilon_2, k_2 & ; \epsilon, k \\ +1, \rho_1, & +1, \rho_2 & ; +1, \rho \end{pmatrix} = \delta(\rho_1^2/2 + \rho_2^2/2 - \rho^2/2) c_{k,P}^{k_1,k_2} \rho_1^{2k_1-1} \rho_2^{2k_2-1} \rho^{-2k_1-2k_2+1} \times F\left[\begin{matrix} k_1+k_2-k, k_1+k_2+k-1 \\ 2k_1 \end{matrix}; \rho_1^2/\rho^2\right], \tag{9.11a}$$

$$c_{k,P}^{k_1,k_2} = [\pi\Gamma(2k_1)]^{-1}[\tfrac{1}{2}\kappa \sinh\pi\kappa\,\Gamma(k_1+k_2-k)\Gamma(k_1-k_2+k)\Gamma(k_1+k_2+k-1)\Gamma(k_1-k_2-k+1)] \times \sin\pi(k_2+\epsilon_2)\sin\pi(k_2-\epsilon_2)/\sin\pi(k+\epsilon)\sin\pi(k-\epsilon)]^{1/2}, \tag{9.11b}$$

and

$$C_>\begin{pmatrix} k_1, & \epsilon_2, k_2 & ; \epsilon, k \\ +1, \rho_1, & -1, \rho_2 & ; +1, \rho \end{pmatrix} = \delta(\rho_1^2/2 - \rho_2^2/2 - \rho^2/2) \times \left\{ c_{k,H}^{k_1,k_2} \rho_1^{2k_1-1} \rho_2^{2k_2-1} \rho^{-2k_1-2k_2+1} F\left[\begin{matrix} k_1+k_2-k, k_1+k_2+k-1 \\ 2k_2 \end{matrix}; -\rho_2^2/\rho^2\right] + (k_2 \leftrightarrow 1-k_2)\right\}, \tag{9.12a}$$

$$c_{k,H}^{k_1,k_2} = (-1)^{k_1-\epsilon_1}\frac{\Gamma(2k-1)}{\pi}\left[\tfrac{1}{2}\kappa \sinh\pi\kappa\,\frac{\Gamma(k_1+k_2-k)\Gamma(k_1-k_2-k+1)}{\Gamma(k_1-k_2+k)\Gamma(k_1+k_2+k-1)}\,\frac{\sin\pi(k-\epsilon)}{\sin\pi(k+\epsilon)}\right]^{1/2}, \tag{9.12b}$$

where, again, (9.1c) is implied.

The expressions for the $\tau = 1$ CGC's follow from (6.21).

9.6: We have given in some detail the considerations which allowed the introduction of the parameter α in (9.9b). Some further remarks may place its origin in a different context: The CGC's are being found, basically, as the solutions

to a single differential equation [Eqs. (6.12)] written in terms of a single complex variable Θ [Eqs. (6.13)] along three line segments in the complex plane. The two-chart $D^+ \times \mathscr{C}$ coupling seen in this section requires the two segments $[0,\pi/2]$ and $[0,i\infty)$ which meet at $\Theta = 0$, at the common boundary suggested by Figs. 2 and 5(b). The potential barrier at $\Theta = \theta_P = \pi/2$ fixes the solution to be $w^{k_1,k_2,k}_{1(1)}(\Theta)$. If we take seriously the idea of analytic continuation of the solutions from one chart to another, through the L-shaped contour in the complex Θ-plane (thus automatically cancelling boundary Wronskians), the same $w^{k_1,k_2,k}_{1(1)}(\Theta)$ solution ought to continue into the $H_>$ region up along the contour. As was pointed out after Eqs. (6.14)–(6.16), however, $w^{k_1,k_2,k}_{1(1)}(\Theta)$, $\Theta \in iR^+$ lies along the branch cut of the hypergeometric function, namely $[1,\infty)$. Moreover, as a closer examination will reveal, although our solution (9.9) on the $H_>$-chart has the appropriate coefficients to reconstitute precisely $w^{k_1,k_2,k}_{1(1)}(i\theta)$, the relative constant phase factor is undetermined insofar as the Riemann sheet of the function is not specified. Finally, as one can again verify, if we choose any one Riemann sheet, the obtained $w^{k_1,k_2,k}_{1(1)}(i\theta)$ will not exhibit the appropriate asymptotic behavior (9.9a) to allow for Dirac orthonormalization of the function set. Through linear combinations of the functions on various sheets, nevertheless, we can produce this behavior. This freedom in the choice is equivalent to the introduction of the constant α in (9.9). We may conclude that this parametrized multivaluation of the two-chart Pöschl–Teller eigenfunctions is a feature associated with the branch cut of the Gauss hypergeometric function. The proper value of α was found in (9.10b) through the requirement of orthogonality between the discrete and continuous spectrum eigenfunctions.

These considerations will apply to all cases in the next section—involving three charts.

X. THE COUPLING $\mathscr{C} \times \mathscr{C}$

This section examines the coupling of two continuous series SAIR's to SAIR's belonging to the discrete, continuous and—if both factor SAIR's are exceptional—a single SAIR of the exceptional continuous series.

10.1: When two continuous-series SAIR's couple, as both σ_1 and σ_2 may be ± 1, all six charts of Fig. 2 enter into the picture; three of them in the upper-right half-plane for the total $\sigma = 1$, and three in the lower-left half-plane for total $\sigma = -1$. In each case, the $H_<$, P-, and $H_>$-charts involve strongly welled (and/or weakly barriered/welled) Pöschl–Teller potentials of the first and second kind, as shown in Fig. 6. The total Casimir operator eigenfunctions have three forms, one in each chart. Under the total inner product [Eqs. (6.18) with either $\sigma = +1$ or $\sigma = -1$] the solutions should be orthonormal, in the ordinary or generalized sense, according to whether the coupling is made into the $\mathscr{D}$ or $\mathscr{C}$ SAIR's, corresponding respectively to bound or free states of the H-chart potentials. Furthermore, they must be mutually orthogonal.

10.2: Again, we consider first the coupling $\mathscr{C} \times \mathscr{C} \to \mathscr{D}$ to the discrete series SAIR's, where $-(2k-1)^2 \leqslant 0$ correspond to the "bound" states of Fig. 6. The asymptotic behavior of a normalizable solution f_k must be the exponentially decreasing $w^{k_1,k_2,k}_{2(\infty)}(\theta_{H_>})$ in the $H_>$-chart, and similarly for the $H_<$-chart. Keeping (6.16c) in mind, we set

$$f^{H_>}_k(\theta) = C^{H_>}_k w^{k_1,k_2,k}_{2(\infty)}(\theta), \quad f^{H_<}_k(\theta) = C^{H_<}_k w^{k_2,k_1,k}_{2(\infty)}(\theta), \tag{10.1}$$

and leave $f^P_k(\theta)$ free and subject to adjustment so that, in

$$\begin{aligned}(f_k,f_l)_{s^2} &= (f^{H_<}_k, f^{H_<}_l)_h + (f^P_k, f^P_l)_p + (f^{H_>}_k, f^{H_>}_l)_h \\ &= [4(k+l-1)(k-l)]^{-1}\left[C^{H_<}_k {}^* C^{H_<}_l W(w^{k_2,k_1,k*}_{2(\infty)}, w^{k_2,k_1,l}_{2(\infty)})\,|_0^\infty \right. \\ &\quad \left. - W(f^{P*}_k, f^P_l)\,|_0^{\pi/2} + C^{H_>}_k {}^* C^{H_>}_l W(w^{k_1,k_2,k*}_{2(\infty)}, w^{k_1,k_2,l}_{2(\infty)})\,|_0^\infty\right],\end{aligned} \tag{10.2}$$

the "adjoining" Wronskian boundary values [at $\theta_{H_<} = 0$, $\theta_P = \pi/2$ and at $\theta_P = 0$, $\theta_{H_>} = 0$, as suggested by Fig. 6] cancel. At $\theta_{H_<} \to \infty$ and $\theta_{H_>} \to \infty$ they vanish exponentially due to the solution choice.

The condition of pairwise Wronskian cancellation should provide the expression for $f^P_k(\theta)$ in terms of $w^{k_1,k_2,k}_{n(0)}(\theta)$ or of $w^{k_1,k_2,k}_{m(1)}(\theta)$, $n,m = 1,2$, the relative coefficient ratio among the solutions in the three charts, and the quantization condition ($k - l$ integer). It will be observed that the branch cut of the hypergeometric functions in $w^{k_1,k_2,k}_{n(\infty)}(\Theta)$, $n = 1,2$, as given by (5.16) now falls on the P-chart θ range. For this reason we may expect, as in the $\mathscr{D} \times \mathscr{C} \to \mathscr{C}$ case of the last section, certain multivaluation features for the solution functions in this interval.

We proceed as in the last section, exploring the behavior of the fixed functions $w^{k_1,k_2,k}_{2(\infty)}(\theta)$ near $\theta_{H_>} \to 0^+$. This is given by $B^>_k \theta^{2k_2 - 1/2} + \tilde{B}^>_k \theta^{-2k_2 + 3/2}$, where $B^>_k$ and $\tilde{B}^>_k = B^>_k(k_2 \leftrightarrow 1 - k_2)$ are found from the $z \leftrightarrow 1/z$ linear transformation formulas (Ref. 19, Eq. 15.3.7) between Gauss functions, and given in (9.4b) as B_k. When $k_2 \in C$, the second summand is the complex conjugate of the first, while when $k_2 \in E$ they are both real. The boundary Wronskian is given in (9.5b). A similar expression holds for the function in the $H_<$-chart, with k_1 and k_2 exchanged, and coefficients which we shall denote by $B^<_k = B^>_k$ $(k_1 \leftrightarrow k_2)$ and $\tilde{B}^<_k = B^<_k$ $(k_1 \leftrightarrow 1 - k)$. We may now write the general solution in P-chart as

$$\begin{aligned} f^P_k(\theta) &= \beta_k B^>_k w^{k_1,k_2,k}_{1(0)}(\theta) + \tilde{\beta}_k \tilde{B}^>_k w^{k_1,k_2,k}_{2(0)}(\theta) \\ &= \gamma_k B^<_k w^{k_1,k_2,k}_{1(1)}(\theta) + \tilde{\gamma}_k \tilde{B}^<_k w^{k_1,k_2,k}_{2(1)}(\theta), \end{aligned} \tag{10.3}$$

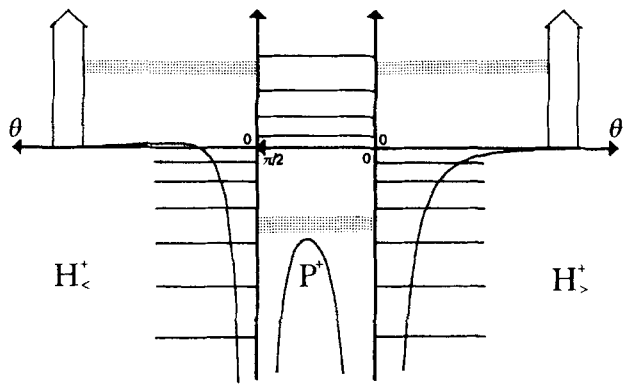


FIG. 6. The $\mathscr{C} \times \mathscr{C}$ coupling in the region composed of the $H^+_<$, P^+, and $H^+_>$ charts. Both continuous- and discrete-series summands are obtained.

which lead to real $f_k^{P}(\theta)$ when $\tilde\beta_k = \beta_k^*$ for $k_2 \in C$ or $\beta_k, \tilde\beta_k$ real for $k_2 \in E$. The expressions of $w^k_{m(0)}(\theta)$ in terms of the $w^{(k)}_{n(1)}(\theta)$ leads to the following relation between the γ's and the β's:

$$\gamma_k = [\sin 2\pi k_2]^{-1}[\beta_k \sin \pi(k_1 - k_2 + k) + \tilde\beta_k \sin \pi(k_1 + k_2 + k)], \tag{10.4}$$

and $\tilde\gamma_k$ given by the above expression exchanging $k_1 \leftrightarrow 1 - k_1$. For $k_1 \in C$, $\tilde\gamma_k = \gamma_k^*$ while for $k_1 \in E$ both γ_k and $\tilde\gamma_k$ are real, i.e., the γ's have the same properties as the β's only exchanging k_1 and k_2.

Now, the boundary value of $W(f_k^{P*}, f_l^{P})$ at $\theta_P \to 0^+$ may equal that of $W(w^{k*}_{2(\infty)}, w^{l}_{2(\infty)})$ at $\theta_{H_>} \to 0^+$ only when $\tilde\beta_k \beta_l = \beta_k \tilde\beta_l$ for all values of k and l in a discrete range. The normalization constants in (10.1) will then relate as $C_k^{H_>} = (\beta_k \tilde\beta_k)^{1/2}$. The same statement holds true at $\theta_P \to \pi/2^-$ and $\theta_{H_<} \to 0$: $\tilde\gamma_k \gamma_l = \gamma_k \tilde\gamma_l$ and $C_k^{H_<} = (\gamma_k \tilde\gamma_k)^{1/2}$. If the conditions on the β's and the γ's are to hold simultaneously, the relation (10.4) implies that $k - l$ must be an integer, i.e., $k = \epsilon + n$, $n \in Z^+$, are the allowed values of the spectrum, with $\epsilon \equiv \epsilon_1 + \epsilon_2 \pmod 1$ as given by (6.6). The overall normalization of the three-chart functions f_k, $k \in \mathscr{D}$, then follows from (10.1) upon letting $l \to k$, and reasoning in analogy with (3.14). It yields

$$(f_k, f_l)_{s^2} = [2(2k-1)]^{-1}[(2k_2 - 1)B_k^{\gtrless} \tilde B_k^{\gtrless} \beta_k \tilde\beta_k \partial_k \ln(\beta_k/\tilde\beta_k) + (2k_1 - 1)B_k^{\lessgtr} \tilde B_k^{\lessgtr} \gamma_k \tilde\gamma_k \partial_k \ln(\gamma_k/\tilde\gamma_k)]. \tag{10.5}$$

It should be noted that one choice of β's (and corresponding γ's) is

$$\left.\begin{aligned}\beta_k &= \sin \pi(k_2 + \epsilon_2)\\ \tilde\beta_k &= \sin \pi(k_2 - \epsilon_2)\end{aligned}\right\} \Leftrightarrow \begin{cases}\gamma_k = \sin \pi([k_1 + \epsilon_1] + [k - \epsilon])\\ \tilde\gamma_k = \sin \pi([k_1 - \epsilon_1] - [k - \epsilon]).\end{cases} \tag{10.6}$$

The required orthogonality of f_k, $k \in \mathscr{D}$, with f_l, $l \in \mathscr{C}$, as will be seen below, further restricts $\beta_k/\tilde\beta_k$ and $\gamma_k/\tilde\gamma_k$ to be constants independent of k (for $k = \epsilon + n$, $n \in Z^+$). The choice (10.6) turns out thus to be essentially unique—up to constants which are compensated by normalization—for the orthogonality of the full eigenfunction set.

Incorporating (10.6) into (10.5) and into the constants in (10.1a) and (10.3), we find the CGC's of the $\mathscr{C} \times \mathscr{C} \to D^+$ coupling through (6.15), (6.19), and (6.20). The results are, for the P-, $H_>$-, and $H_<$-charts,

$$C\begin{pmatrix}\epsilon_1, k_1, & \epsilon_2, k_2; & k\\ +1, \rho_1, & +1, \rho_2; & +1, \rho\end{pmatrix} = \delta(\rho_1^2/2 + \rho_2^2/2 - \rho^2/2) \times \left\{c_{k,P}^{k_1,k_2} \rho_1^{2k_1 - 1} \rho_2^{2k_2 - 1} \rho^{-2k_1 - 2k_2 + 1} F\left[\begin{matrix}k_1 + k_2 - k, k_1 + k_2 + k - 1\\ 2k_1\end{matrix}; \rho_1^2/\rho\right] + (k_1 \leftrightarrow 1 - k_1)\right\}, \tag{10.7a}$$

$$c_{k,P}^{k_1,k_2} = \frac{(-1)^{k-\epsilon}}{\pi} \sin \pi(k_1 + \epsilon_1) \Gamma(-2k_1 + 1)\left[\frac{2(2k-1)\Gamma(k_1 - k_2 + k)\Gamma(k_1 + k_2 + k - 1)}{\Gamma(-k_1 + k_2 + k)\Gamma(-k_1 - k_2 + k + 1)}\right]^{1/2}, \tag{10.7b}$$

$$C_>\begin{pmatrix}\epsilon_1, k_1, & \epsilon_2, k_2; & k\\ +1, \rho_1, & -1, \rho_2; & +1, \rho\end{pmatrix} = \delta(\rho_1^2/2 - \rho_2^2/2 - \rho^2/2) \times c_{k,H}^{k_1,k_2} \rho_1^{2k_1 - 1} \rho_2^{-2k+1} \rho^{2k-1} F\left[\begin{matrix}k_1 - k_2 + k, k_1 + k_2 + k - 1\\ 2k\end{matrix}; -\rho^2/\rho_2^2\right], \tag{10.8a}$$

$$c_{k,H}^{k_1,k_2} = [\pi\Gamma(2k)]^{-1}[2(2k-1)\sin\pi(k_2 + \epsilon_2)\sin\pi(k_2 - \epsilon_2) \times \Gamma(k_1 - k_2 + k)\Gamma(-k_1 + k_2 + k)\Gamma(k_1 + k_2 + k - 1)\Gamma(-k_1 - k_2 + k + 1)]^{1/2}, \tag{10.8b}$$

$$C_<\begin{pmatrix}\epsilon_1, k_1, & \epsilon_2, k_2; & k\\ -1, \rho_1, & +1, \rho_2; & +1, \rho\end{pmatrix} = C_>\begin{pmatrix}\epsilon_2, k_2, & \epsilon_1, k_1; & \epsilon, k\\ +1, \rho_2, & -1, \rho_1; & +1, \rho\end{pmatrix}, \tag{10.9}$$

where the subscript $\gtrless$ implies that $\rho_1 \gtrless \rho_2$.

The CGC's for the $\mathscr{C} \times \mathscr{C} \to D^-$ coupling are obtained through (6.21) from the above expressions.

10.3: There is one exceptional solution (10.1)–(10.3) which falls outside the limitations of our previous quantization of $k \in D$, and this is obtained when $B_k^{\gtrless} = 0 = B_k^{\lessgtr}$ due to a pole of one of the Γ functions in (9.4b) at the value

$$k = k_0 = k_1 + k_2 - 1. \tag{10.10}$$

This solution is hence present only when two exceptional continuous SAIR's are coupled to a total k. Since $\epsilon = \epsilon_1 + \epsilon_2$ is in general different from $\pm k_0$, we conclude that it belongs to the exceptional continuous type of SAIR's. It is thus only present when $\frac{1}{2} < k_0 < 1$, i.e., for $\frac{3}{2} < k_1 + k_2 < 2$ or, for $k_j = (1 + \kappa_j)/2$, $1 < \kappa_1 + \kappa_2 < 2$. In that case, $\tilde B_{k_0}^{\gtrless} = 1 = \tilde B_{k_0}^{\lessgtr}$ and $\tilde B_{k_0} = \tilde\gamma_{k_0}$ is an arbitrary constant. In fact, since one of the upper parameters of the Gauss function is zero, the nonnormalized solution is simply

$$f_{k_0}(\Theta) = C_{k_0,C}^{k_1,k_2} |\cos \Theta_C|^{-2k_1 + 3/2} |\sin \Theta_C|^{-2k_2 + 3/2}, \tag{10.11}$$

valid in the three charts $C = H_<, P, H_>$ with Θ_C as in (6.13). The ratio of the normalization coefficients $C_{k_0,C}^{k_1,k_2}$ in each pair of "adjoining" charts may be found through requiring that it be orthogonal to any other f_k, $k \in D$. This requirement leads to

$$C_{k_0,H_>}/C_{k_0,P} = (\beta_k/\tilde\beta_k)^{1/2}, \quad C_{k_0,H_<}/C_{k_0,P} = (\gamma_k/\tilde\gamma_k)^{1/2}, \tag{10.12}$$

which in turn demands that $\beta_k/\tilde\beta_k$, $k = \epsilon + n$, $n \in Z^+$, be independent of $k - \epsilon$, or a periodic function of k with period unity. This property is satisfied by (10.6). The norm of the

exceptional-series state may not be calculated through (10.5) since the process $l \to k_0$ is invalid here [it would give a fictitious infinite norm to the exceptional state (10.11)]. It may be obtained directly, however. Once the normalized solution is found, the rest of the program follows.

The CGC's for the coupling $E \times E \to E$ are thus found:

$$C\begin{pmatrix}\epsilon_1,k_1, & \epsilon_2,k_2;\epsilon,k_0\\ \tau_1,\rho_1, & \tau_2,\rho_2; +1,\rho\end{pmatrix} = \delta(\tau_1\rho_1^2/2 + \tau_2\rho_2^2/2 - \rho^2/2)$$
$$\times c_{k_0,E}^{k_1,k_2}[\sin\pi(k_1 - \tau_1\epsilon_1)\sin\pi(k_2 - \tau_2\epsilon_2)]^{1/2}$$
$$\times \rho_1^{-2k_1+1}\rho_2^{-2k_2+1}\rho^{2k_0-1}, \tag{10.13a}$$

where we have written an expression valid for all three charts; the normalization coefficient is

$$c_{k_0,E}^{k_1,k_2} = (2\pi)^{1/2}[\Gamma(-2k_1+2)\Gamma(-2k_2+2)\Gamma(2k_0-1)$$
$$\times\{\sin\pi(2k_1-1)\sin\pi(k_1-\epsilon_1)\sin\pi(k_2+\epsilon_2)$$
$$+\sin\pi(2k_2-1)\sin\pi(k_1+\epsilon_1)\sin\pi(k_2-\epsilon_2)$$
$$+\sin\pi(2k_0-1)\sin\pi(k_1-\epsilon_1)\sin\pi(k_2-\epsilon_2)\}]^{-1/2}. \tag{10.13b}$$

For total $\tau = -1$, (6.21) may be used.

10.4: The last case to be analyzed pertains to the coupling of two continuous-series SAIR's to SAIR's in the continuous series. Corresponding to each eigenvalue $-(2k-1)^2 > 0$, there are two independent generalized solutions to the eigenvalue problem. The solution functions f_k may be given the following forms in each of the three charts:

$$f_k^{H_>}(\theta) = G_{k,\varphi}^{>} w_{1(\infty)}^{k_1,k_2,k}(\theta) + \tilde{G}_{k,\varphi}^{>} w_{2(\infty)}^{k_1,k_2,k}(\theta)$$
$$= \alpha_> D_{k,\varphi} w_{1(0)}^{k_1,k_2,k}(\theta) + \alpha_>^{-1}\tilde{D}_{k,\varphi} w_{2(\infty)}^{k_1,k_2,k}(\theta), \tag{10.14a}$$

$$f_k^{P}(\theta) = D_{k,\varphi} w_{1(0)}^{k_1,k_2,k}(\theta) + \tilde{D}_{k,\varphi} w_{2(0)}^{k_1,k_2,k}(\theta)$$
$$= \varphi_k w_{1(1)}^{k_1,k_2,k}(\theta) + \varphi_k^{-1} w_{2(1)}^{k_1,k_2,k}(\theta), \tag{10.14b}$$

$$f_k^{H_<}(\theta) = \alpha_< \varphi_k w_{1(0)}^{k_2,k_1,k}(\theta) + \alpha_<^{-1}\varphi_k^{-1} w_{2(0)}^{k_2,k_1,k}(\theta),$$
$$= G_{k,\varphi}^{<} w_{1(\infty)}^{k_2,k_1,k}(\theta) + \tilde{G}_{k,\varphi}^{<} w_{2(\infty)}^{k_2,k_1,k}(\theta). \tag{10.14c}$$

Our starting point will be the "common" boundary between the $H_<$- and P-charts in Fig. 6, where we propose $f_k^P(\theta)$ in (9.14b) to be a linear combination of the two solutions $w_{n(1)}^{k_1,k_2,k}(\theta)$, $n = 1,2$, with a free parameter φ_k. The choice of real solutions requires that $|\varphi_k| = 1$ when $k_1 \in C$ and $\varphi_k = \varphi_k^*$ when $k_1 \in E$. All other coefficients in (9.14) except the α's are to depend on the choice of φ_k. When $k_1 \in E$ is such that $\pm\epsilon_1 \to k_1$, the $D \times \mathscr{C}$ coupling of the last section will be recovered in the limit $\varphi_k \to \infty$ (the overall normalization constant will ensure that only the $w_{2(1)}^k$ term vanishes, with no other singularity present). The $D_{k,\varphi}$ and $\tilde{D}_{k,\varphi}$ are uniquely determined in terms of φ_k through the $z \leftrightarrow 1-z$ transformations of the Gauss functions as in (9.3) and its $k_1 \leftrightarrow 1 - k_1$ replacement. Given $f_k^P(\theta)$ at $\theta = 0$ and $\pi/2$, the forms (10.14a)–(10.14c) of $f_k^{H_>}(\theta)$ and $F_k^{H_<}(\theta)$ at $f_k^{H_<}(\theta)$ at $\theta_{H_>} = 0$ and $\theta_{H_<} = 0$ are determined up to coefficients $\alpha_>$ and $\alpha_<$ If the boundary Wronskians of any two solutions f_k and f_l, $k,l \in C$ are to vanish, the α's must be independent of k and l. Their role is thus the analog of the α introduced in (9.9b). When $k_1 \in C$, $|\alpha_<| = 1$, while when $k_1 \in E$, $\alpha_< = \alpha_<^*$; similar conditions hold for $\alpha_>$ in terms of k_2. Now, orthogonality of f_k, $k \in C$ to f_l, $l \in D$, requires that $\alpha_> = (\tilde{\beta}_k/\beta_k)^{1/2}$ and $\alpha_< = (\gamma_k/\gamma_k)^{1/2}$, and hence the ratios $\tilde{\beta}_k/\beta_k$, γ_k/γ_k must be independent of k (when $k = \epsilon + n$, $n \in Z^+$). This condition is compatible with (10.12) and satisfied by (10.6), which also exhibits the corresponding reality properties. Orthogonality to the exceptional state f_{k_0} in (10.11) is automatically assured.

The last constants to be determined in terms of φ_k are the coefficients of the $w_{n(\infty)}^k$. The $z \leftrightarrow 1/z$ transformation of the Gauss functions—the inverse of (9.4) and its $k \leftrightarrow 1-k$ replacement—yields

$$G_{k,\varphi}^{>} = [\sin\pi(k_2+\epsilon_2)\sin\pi(k_2-\epsilon_2)]^{-1/2}$$
$$\times\{\sin\pi[-(k_1-\epsilon_1)+(k-\epsilon)]C(k_1,k_2,k)\varphi_k$$
$$+\sin\pi[-(k_1+\epsilon_1)-(k-\epsilon)]C(1-k_1,k_2,k)\varphi_k^{-1}\}, \tag{10.15a}$$

$$\tilde{G}_{k,\varphi}^{>} = G_{1-k,\varphi}^{>} = (G_{k,\varphi}^{>})^*, \tag{10.15b}$$

$$G_{k,\varphi}^{<} = [\sin\pi(k_1+\epsilon_1)\sin\pi(k_1-\epsilon_1)]^{-1/2}$$
$$\times[\sin\pi(k_1-\epsilon_1)C(k_1,k_2,k)\varphi_k$$
$$+\sin\pi(k_1+\epsilon_1)C(1-k_1,k_2,k)\varphi_k^{-1}], \tag{10.16a}$$

$$\tilde{G}_{k,\varphi}^{<} = G_{1-k,\varphi}^{<} = (G_{k,\varphi}^{<})^*, \tag{10.16b}$$

where

$$C(k_1,k_2,k) = \Gamma(2k_1)\Gamma(2k-1)/\Gamma(k_1-k_2+k)$$
$$\times\Gamma(k_1+k_2+k-1). \tag{10.17}$$

The equalities (10.15b) and (10.16b) [compare with (9.9a)] are the ones which insure that two solutions $f_{k,\varphi}$, $f_{l,\varphi}$, $k,l \in C$ may be subject to Dirac normalization though their cross-Wronskians at $\theta_{H_>} \to \infty$ in the manner of (4.9). The overall normalization of (10.14) will then be

$$c_{k,\varphi} = [2\pi(|G_{k,\varphi}^{>}|^2 + |G_{k,\varphi}^{<}|^2)]^{-1/2}. \tag{10.18}$$

Similarly, two solutions $f_{k,\varphi}$ and $f_{k,\psi}$ will be orthogonal when the two cross-Wronskians cancel. This happens when φ_k and ψ_k are such that

$$G_{k,\varphi}^{>}(G_{k,\psi}^{>})^* + (G_{k,\varphi}^{<})^* G_{k,\psi}^{<} = 0. \tag{10.19a}$$

A special choice of mutually orthogonal solutions, labeled by φ_+ and φ_-, may be built demanding

$$G_{k,\varphi_\pm}^{<} = \pm(G_{k,\varphi_\pm}^{>})^*. \tag{10.19b}$$

When the conjugation properties in (10.15) and (10.16) are used, (10.19b) yields an algebraic expression for $(\varphi_\pm)^2$. Alternatively, (10.19a) embodies the Schmidt orthogonalization process for generalized functions whereby a $f_{k,\psi}$ may be found which is orthogonal to any given $f_{k,\phi}$. For $k_1 \in E$, finally, the choice of $\varphi \to \infty$ yields a preferred $\mathscr{C} \times \mathscr{C} \to C$ CGC coefficient, which upon $\pm\epsilon_1 \to k_1$, yields the $D \times \mathscr{C} \to C$ CGC's seen in the last section.

Unfortunately, when (10.15)–(10.19) are substituted in (10.4), no significant algebraic reductions seem to take place in the final expressions, which would merit their explicit display. For the $\mathscr{C} \times \mathscr{C} \to C$ coupling we may thus write

$$C^{\varphi}_{>}\begin{pmatrix}\epsilon_1,k_1, & \epsilon_2,k_2; & \epsilon,k\\ 1+,\rho_1, & -1,\rho_2; & +1,\rho\end{pmatrix}=\delta(\rho_1^2/2-\rho_2^2/2-\rho^2/2)c_{k,\varphi}$$
$$\times\left\{G^{>}_{k,\varphi}\rho_1^{2k_1-1}\rho_2^{-2k_1+2k-1}\rho^{-2k+1}F\left[\begin{matrix}k_1+k_2-k,k_1-k_2-k+1\\ 2-2k\end{matrix};-\rho^2/\rho_2^2\right]+(k\leftrightarrow 1-k)\right\}, \tag{10.20a}$$

$$C^{\varphi}\begin{pmatrix}\epsilon_1,k_1, & \epsilon_2,k_2; & \epsilon,k\\ +1,\rho_1, & +1,\rho_2; & +1,\rho\end{pmatrix}=\delta(\rho_1^2/2+\rho_2^2/2-\rho^2/2)c_{k,\varphi}$$
$$\times\left\{\varphi\left(\rho_1^{2k_1-1}\rho_2^{2k_2-1}\rho^{-2k_1-2k_2+1}F\left[\begin{matrix}k_1+k_2-k,k_1+k_2+k-1\\ 2k_1\end{matrix};\rho^2/\rho_1^2\right]\right)+\varphi^{-1}(k_1\leftrightarrow 1-k_1)\right\}, \tag{10.20b}$$

$$C^{\varphi}_{<}\begin{pmatrix}\epsilon_1,k_1, & \epsilon_2,k_2; & \epsilon,k\\ -1,\rho_1, & +1,\rho_2; & +1,\rho\end{pmatrix}=\delta(-\rho_1^2/2+\rho_2^2/2-\rho^2/2)c_{k,\varphi}$$
$$\times\left\{G^{<}_{k,\varphi}\rho_1^{-2k_2+2k-1}\rho_2^{2k_2-1}\rho^{-2k+1}F\left[\begin{matrix}k_1+k_2-k,-k_1+k_2-k+1\\ 2-2k\end{matrix};-\rho^2/\rho_1^2\right]+(k\leftrightarrow 1-k)\right\}. \tag{10.20c}$$

For $\tau=-1$, (6.21) may be used.

XI. CONCLUDING REMARKS

The explicit construction of the Clebsch–Gordan coefficients for an algebra is but one of a set of related problems whose complete elucidation for so(2,1) is of interest in mathematical physics. This algebra is, after all, the simplest of all noncompact semisimple Lie algebras. The symmetry, composition, and asymptotic properties of the so(3) CGC's have been fruitfully exploited, but their so(2,1) counterparts have not yet received a comparable coverage. Composition properties such as (1.1) entail special-function relations between confluent and Gauss hypergeometric functions; the Plancherel measure[22] and the SO(2,1) UIR matrix elements[23] in the parabolic and other bases are known, and now the CGC's are available. The same remarks apply for the orthogonality and completeness relations for the CGC's, as well as a deeper study of their analyticity properties[3] with respect to the SAIR indices k_1,k_2,k, extended to indecomposable, finite-dimensional, and other non-self-adjoint representations of the algebra.

CGC's in bases other than the parabolic basis may be computed and compared with the existing results. This involves integrating the coefficients with the overlap functions between the iso(1) basis used here, and the so(1,1) or so(2) bases.[23] The former requires a triple Mellin transform of our results over ρ_1,ρ_2, and ρ and summation over the range of τ's to compare with the results of Mukunda and Radhakrishnan[7]; the latter a triple Laguerre and/or Whittaker transform to compare with Holman and Biedenharn's work.[4] In relation with the former—a still practical calculation—one may obtain independently the matrix elements of SO(3,1) and SO(2,2) in various bases.

Perhaps the nearest task of interest is the analog of Sec. V, the study of all common self-adjoint extensions of a second-order differential operator algebra, as applied to the N-dimensional symplectic algebra sp($2N,R$). The oscillator representation of the latter is well known,[12] and certain special cases of the "radial/hyperbolic" reductions $\mathrm{sp}(2rs)\supset \mathrm{Sp}(2r)+\mathrm{so}(s-t,t)$ have been subject to scrutiny,[24] in particular $\mathrm{sp}(4)\approx\mathrm{so}(3,2)$. Finally, one may realistically hope that the CG and representation problems for the four-dimensional Lorentz, Poincaré, and conformal algebras may simplify in their analog of the parabolic basis, as it did here, for so(2,1).

ACKNOWLEDGMENTS

We would like to thank Dr. Alberto Alonso y Coria and Dr. Ricardo Weder (IIMAS—UNAM) for their interest in our work, as well as Dr. H. Kalb, of Universität Darmstadt, for a crucial illuminating discussion on the properties of strongly singular potentials.

APPENDIX

The self-adjoint irreducible representations of the three-dimensional Lorentz algebra so(2,1), with generators J_1, J_2, J_0 satisfying the commutation relations (2.1f), may be classified[1,25] through the eigenvalue q of the Casimir operator $Q=J_1^2+J_2^2-J_0^2$, and either a "multivaluation" index $\epsilon\in(-\frac{1}{2},\frac{1}{2}]$ or a sign $\tau=\pm 1$. The latter two indicate the common self-adjoint extension of the algebra generators as detailed in Sec. V; ϵ is given by the spectrum $\{\mu\}$ of the elliptic subalgebra generator J_0, modulo unity. When unique, τ specifies the sign of the parabolic subalgebra generator $J_-=J_0-J_1$. The representations are conveniently labeled also by a parameter k which relates to q through $q=k(1-k)$. The equivalence under $k\leftrightarrow 1-k$ is used in the $\mathscr{C}$-series (below) to reduce its range to $\mathrm{Re}k\geqslant\frac{1}{2}$.

The classification of the SAIR's of so(2,1) is as follows [$R^+=[0,\infty)$, $Z^+=\{0,1,2,\cdots\}$]:

Continuous series ($\mathscr{C}$):

C: Nonexceptional (C^{ϵ}_{q}):

$q\geqslant\frac{1}{4}$ [i.e., $k=(1+i\kappa)/2$, $\kappa\in R^+$], $\epsilon\in(1-\frac{1}{2},\frac{1}{2}]$

with the exception of $q=\frac{1}{4}$ [i.e., $k=\frac{1}{2}$, $\kappa=0$], $\epsilon=\frac{1}{2}$.

E: Exceptional (C^{ϵ}_{q}):

$0<q<\frac{1}{4}$ [i.e., $k=(1+\kappa)/2$, $\kappa\in(0,1)$], $|\epsilon|<1-k$.

In both subcases: $\mu\in\{\epsilon+n, n\in Z\}$, $\tau\in\{-1,+1\}$.

Discrete series ($\mathscr{D}$)

D^+: *Positive discrete series* (D_k^+):

$k>0$ [i.e., $q\leqslant\frac{1}{4}$], $\epsilon\equiv k \pmod 1$,

$\mu\in\{k+n, n\in Z^+\}$, $\tau=+1$.

D^-: *Negative discrete series* (D_k^-):

$k>0$ [i.e., $q\leqslant\frac{1}{4}$], $\epsilon\equiv -k \pmod 1$,

$\mu\in\{-k-n, n\in Z^+\}$, $\tau=-1$.

The $k\leftrightarrow 1-k$ symmetry can be upheld in the exceptional interval -$\mathscr{D}$, $0<k<1$ if we ascribe $\epsilon=k\in(0,\frac{1}{2}]$ to $1-k\in[\frac{1}{2},1)$ and keep $0<\epsilon\leqslant\frac{1}{2}$, while for $k\in(\frac{1}{2},1]$ we take $0>\epsilon=1-k>-\frac{1}{2}$.

The Clebsch–Gordan series[2,4] are written below. We use the following conventions:

(i) All sums are direct and range over integer-spaced k's.

(ii) Direct integrals, as written, do not exhibit the Plancherel measure.[22]

(iii) In all cases $\epsilon=\epsilon_1+\epsilon_2 \pmod 1$, with $\epsilon_j\equiv\pm k_j \pmod 1$ for $k_j\in\mathscr{D}$.

(iv) The numerals in brackets indicate the formulae in this article where the corresponding CGC's are given.

$\mathscr{D}\times\mathscr{D}$:

$$D_{k_1}^{\pm}\times D_{k_2}^{\pm}=\sum_{k=k_1+k_2}^{\infty} D_k^{\pm}, \qquad [(7.1)]$$

$$D_{k_1}^{\pm}\times D_{k_2}^{\mp}=\sum_{1/2<k}^{|k_1-k_2|} D_k^{\mathrm{sgn}(k_1-k_2)}+\int_{k\in C} C_q^{\epsilon} \quad [(8.4),(8.1)]$$
$$+(D_{1/2}^{\mathrm{sgn}(k_1-k_2)} \text{ when } |k_1-k_2|=\tfrac{1}{2}); \qquad [(8.1)]$$

$\mathscr{D}\times\mathscr{C}$:

$$D_{k_1}^{\pm}\times C_{q_2}^{\epsilon_2}=\sum_{k=\epsilon}^{\infty} D_k^{\pm}+\int_{k\in C} C_q^{\epsilon}; \qquad [(9.7),(9.1)\text{–}(9.8)]$$

$\mathscr{C}\times\mathscr{C}$:

$$C_{q_1}^{\epsilon_1}\times C_{q_2}^{\epsilon_2}=\sum_{k=\epsilon}^{\infty} D_k^{+}+\sum_{k=\epsilon}^{\infty} D_k^{-}+2\int_{k\in C} C_q^{\epsilon}$$
$$[(10.7)\text{–}(10.9),(10.20)]$$
$$+(C_q^{\epsilon}[k=k_1+k_2-1] \quad \text{for } k_1,k_2,k\in E). \qquad [(10.13)]$$

[1] V. Bargmann, Ann. Math. **48**, 568 (1947).
[2] L. Pukánszky, Trans. Am. Math. Soc. **100**, 116 (1961).
[3] W. J. Holman and L. C. Biedenharn, Ann. Phys. (N.Y). **39**, 1 (1966).
[4] W. J. Holman and L. C. Biedenharn, Ann. Phys. (N.Y.) **47**, 205 (1968).
[5] I. Ferretti and M. Verde, Nuovo Cimento **55**, 110 (1968); M. Verde, *Proceedings of the Mendeleevian Conference*, Torino, 1969 (Torino Academy of Science, Torino, Italy, 1971), p. 305.
[6] K. H. Wang, J. Math. Phys. **11**, 2077 (1970).
[7] N. Mukunda and B. Radhakrishnan, J. Math. Phys. **15**, 1320, 1322, 1643, 1656 (1974).
[8] M. Andrews and J. Gunson, J. Math. Phys. **5**, 1391 (1964); S. S. Sannikov, Dokl. Akad. Nauk SSSR **171**, 1058 (1967) [Sov. Phys. Dokl. **11**, 1045 (1967)]; H. Ui, Ann. Phys. (N.Y.) **49**, 69 (1969); A. O. Barut and R. Wilson, J. Math. Phys. **17**, 900 (1976); E. Chaćon, D. Levi, and M. Moshinsky, J. Math. Phys. **17**, 1919 (1976); S. D. Majumdar, J. Math. Phys. **17**, 193 (1976); D. Basu and S. D. Majumdar, J. Math. Phys. **20**, 492 (1979); D. Basu, "The Master Analytic Function and the Lorentz Group," Kinam (to be published).
[9] J. L. Burchnall and T. W. Chaundy, Proc. London Math. Soc. **50**, 56 (1944).
[10] J. Schwinger, AEC Report, 1952 (unpublished), reprinted in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. van Dam (Academic, New York, 1965); A. Malkin and V. I. Man'ko, Pisma Zh. Eksp. Teor. Fiz. **2**, 230 (1965) [JETP Lett. **2**, 146 (1966)]; A. O. Barut and A. Bohm, J. Math. Phys. **11**, 2938 (1970).
[11] A. Weil, Acta Math. **11**, 143 (1963).
[12] M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772, 1780 (1971); M. Moshinsky, SIAM J. Appl. Math. **25**, 193 (1973).
[13] (a,b) K. B. Wolf, J. Math. Phys. **15**, 1295, 2102 (1974); (c) C. P. Boyer and K. B. Wolf, J. Math. Phys. **16**, 1493 (1975); (d) K. B. Wolf, J. Math. Phys. **21**, 680 (1980).
[14] K. B. Wolf, *Integral Transforms in Science and Engineering* (Plenum, New York, 1979).
[15] N. Mukunda and B. Radhakrishnan, J. Math. Phys. **14**, 254 (1973).
[16] G. Pöschl and E. Teller, Z. Phys. **83**, 143 (1933).
[17] M. Rosen and P. M. Morse, Phys. Rev. **42**, 210 (1932); W. Lotmar, Z. Phys. **93**, 528 (1935); L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951), Sec. 4.8.
[18] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. 2.
[19] M. Abramowitz and I. A. Stegun (Eds.), *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1964).
[20] K. Meetz, Nuovo Cimento **34**, 690 (1964).
[21] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Sums and Products* (Academic, New York, 1965).
[22] L. Pukánszky, Math. Ann. **156**, 96 (1964).
[23] D. Basu and K. B. Wolf, "The Unitary Irreducible Representations of SL (2,R) in All Subgroup Decompositions," J. Math. Phys. **23**, 189 (1982).
[24] M. Moshinsky, T. H. Seligman, and K. B. Wolf, J. Math. Phys. **13**, 901 (1972); P. Kramer, M. Moshinsky, and T. H. Seligman, in *Group Theory and Its Applications*, Vol. III, edited by E. M. Loebl (Academic, New York, 1975); M. Kashiwara and M. Vergne, Invest. Math. **44**, 1 (1978).
[25] P. J. Sally, Jr., Bull. Am. Math. Soc. **72**, 269 (1966).

Upper and lower bounds for eigenvalues of nonlinear elliptic equations: I. The lowest eigenvalue[a)]

Rafael Benguria
Departamento Física, Universidad de Chile, Casilla 5487, Santiago, Chile

M. Cristina Depassier
Instituto de Física, Universidad Católica de Chile, Casilla 114-D, Santiago, Chile

(Received 7 May 1982; accepted for publication 1 October 1982)

We give a method for finding bounds for the lowest eigenvalue of nonlinear elliptic equations with monotone, local, nonlinearities. This is an extension to nonlinear problems of Barta's method for linear elliptic operators.

PACS numbers: 02.30. + g, 02.60. + y

1. INTRODUCTION

There are two main variational characterizations for the fundamental eigenvalue of linear elliptic problems, namely the (integral) Rayleigh–Ritz principle and the (local) Barta method.[1] The Rayleigh principle provides upper bounds for the fundamental eigenvalue whereas Barta's principle gives both upper and lower bounds.[2] The Rayleigh principle is a special case, for the fundamental eigenvalue, of the Courant–Fisher variational characterization[3] for all the eigenvalues. As for Barta's method, it also extends to the rest of the eigenvalues but only for one dimensional problems.[4] When going into nonlinear problems the Courant–Fisher principle goes into the Ljusternik–Schnirelman category theory,[5] where the solutions to nonlinear (eigenvalue) elliptic equations are critical points of a given functional subject to certain constraints. However, the numerical values of this functional at the critical points do not coincide with the eigenvalues corresponding to the solutions aforementioned. Thus the Ljusternik–Schnirelman theory does not directly give bounds for the eigenvalues of nonlinear elliptic problems. In this article we show that Barta's principle for the eigenvalue of linear problems remains practically unchanged when going into nonlinear elliptic eigenvalue equations with local, monotone, nonlinearities. Therefore, this method directly provides with bounds for the eigenvalues of these nonlinear problems. More precisely, it gives bounds for the graphs of the eigenvalues as functions of the norm of the corresponding solutions. We believe the requirement on the nonlinearities to be local can somehow be relaxed and thus, we conjecture that Barta's method should extend to equations such as the Hartree equation,[6] the Thomas–Fermi–von Weizsäcker equation,[7] some equations connected with non-Boussinesq convection,[8] etc. The monotonicity requirement is more stringent as we show in the Ex. 2 below.There are other ways of getting bounds for the eigenvalue of the equations considered here; in particular Amann's method[9] of proving existence of solutions to equations of this type by constructing upper and lower solutions yields as a byproduct bounds on the eigenvalues. The advantage of the extended Barta's method is that the conditions on the trial functions used in the variational inequalities are easier to satisfy. In Sec. 2 we prove our main theorem, i.e., the bounds for the lowest eigenvalue, and we give some examples and applications.

[a)] Supported in part (M.C.D.) by Dirección de Investigación de la Universidad Católica de Chile through DIUC Grant 210/81, and in part (R.B.) by Servicio Desarrollo Cientifico de la Universidad de Chile.

2. BOUNDS ON THE LOWEST EIGENVALUE

Let us consider the problem[10]

$$\begin{aligned} &\mathcal{L}u + f(x,u) = \lambda a u \quad &&\text{in } \Omega, \\ &u = 0 &&\text{on } \partial\Omega, \end{aligned} \tag{1}$$

where $\Omega \subset \mathcal{R}^N$ is a bounded domain with smooth boundary. Here $\mathcal{L}$ denotes a self-adjoint elliptic operator defined by

$$\mathcal{L}u = -\sum_{i,j=1}^{N} \frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial u}{\partial x_j}\right) + c(x)u, \tag{2}$$

with $a_{ij} = a_{ji} \in C^{1,\alpha}(\bar{\Omega})$, $c \in C^{0,\alpha}(\bar{\Omega})$, $c \geqslant 0$ and

$$\sum_{i,j=1}^{N} a_{ij}(x)\xi_i\xi_j \geqslant \eta|\xi|^2,$$

all $x \in \bar{\Omega}$, all $\xi \in \mathcal{R}^N$ with $\eta > 0$, the ellipticity constant. Moreover, $a \in C^0(\bar{\Omega})$ and $a > 0$ in $\bar{\Omega}$. We assume $f: \bar{\Omega} \times \mathcal{R}^+ \to \mathcal{R}$ is continuous, $f(x,0) = 0$, $f(x,s) = o(s)$ in the neighborhood of $s = 0$, uniformly with respect to $x \in \bar{\Omega}$, and $s \to (f(x,s)/s)$ (defined to be 0 in $s = 0$) is a strictly increasing function on $\mathcal{R}^+$, all $x \in \Omega$. Moreover, $\lim_{s\to+\infty}(1/s)f(x,s) = +\infty$, uniformly with respect to $x \in \bar{\Omega}$. Let λ_1 denote the lowest eigenvalue of the linear problem $\mathcal{L}\phi = \lambda a\phi$. It is known[10] that under the above hypothesis on f, there is a unique positive solution u_λ to Eq. (1) for every $\lambda > \lambda_1$. Furthermore, the mapping $\lambda \to u_\lambda$ is continuous from $(\lambda_1, +\infty)$ into $C^{1,\alpha}(\bar{\Omega})$. We are interested in getting bounds on the graph $\lambda = \lambda(\|u_\lambda\|)$, where $\|u_\lambda\|$ denotes the norm of u_λ [without loss of generality we will work with the L^2-norm, i.e., $\|u\| \equiv (\int |u|^2\,dx)^{1/2}$]. Our main result is the following:

Theorem 1: Fix $r > 0$ and let $u \in C^2(\bar{\Omega})$ be any function on Ω such that $u > 0$ almost everywhere in Ω, $u \geqslant 0$ on $\partial\Omega$ and $\|u\| = r$. Then,

$$\lambda(r) \geqslant \inf_{\bar{\Omega}} \{(\mathcal{L}u + f(x,u))/au\}, \tag{3}$$

where $\lambda(r)$ is the eigenvalue corresponding to the positive solution u_λ of Eq. (1) with norm $\|u_\lambda\| = r$.

Proof: We need only consider the case $u \neq u_\lambda$, for if $u = u_\lambda$, all $x \in \Omega$, then Eq. (1) implies

$$[\mathcal{L}u + f(x,u)]/au = \lambda(r) \tag{4}$$

 0022-2488/83/030501-03$02.50

for all $x\in\Omega$ and (3) follows from here. So let $u\in C^2(\bar{\Omega})$, $u\not\equiv u_\lambda$ and define

$$A=\min_{\bar{\Omega}}(\mathcal{L}u+f(x,u)-\lambda(r)au). \tag{5}$$

Since $u\in C^2(\bar{\Omega})$ and $\bar{\Omega}$ is compact A is finite and there exists $y\in\bar{\Omega}$ such that

$$A=\mathcal{L}u(y)+f(y,u(y))-\lambda(r)a(y)u(y).$$

We consider separately the two possibilities i) $A<0$ and ii) $A\geq 0$ (as in Ref. 4). If i) holds, the continuity of $\mathcal{L}u+f(x,u)-\lambda(r)au$ implies that there is an open neighborhood U of y such that

$$\mathcal{L}u+f(x,u)-\lambda(r)au<0, \tag{6}$$

all $x\in U\cap\bar{\Omega}$, and in particular u cannot vanish identically on any open neighborhood contained in $U\cap\bar{\Omega}$ so it follows that

$$u(z)>0 \tag{7}$$

for some $z\in U\cap\bar{\Omega}$. Since $a>0$ in $\bar{\Omega}$, Eqs. (6) and (7) imply

$$\lambda(r)>\{\mathcal{L}u(z)+f(z,u(z))\}/a(z)u(z),$$

which proves the proposition in case i). Now we conclude the theorem by showing that ii) cannot hold unless $u\equiv u_\lambda$. If ii) holds we have

$$\mathcal{L}u+f(x,u)-\lambda(r)au\geq 0, \tag{8}$$

all $x\in\bar{\Omega}$. Eqs. (1) and (8) together with the maximum principle imply

$$u\geq u_\lambda,\quad \text{all } x\in\bar{\Omega}. \tag{9}$$

In fact let $D=\{x\in\Omega\,|\,u(x)<u_\lambda(x)\}$; since u and u_λ are continuous in $\bar{\Omega}$, D is an open subset of Ω. Let $W\equiv(f(x,u_\lambda)/u_\lambda)-\lambda(r)a$ and $V\equiv(f(x,u)/u)-\lambda(r)a$. By hypothesis $s\to(1/s)f(x,s)$ is strictly increasing so $V<W$ in D. Moreover, $u>0$ a.e. in D (in fact in Ω), $u=u_\lambda$ in ∂D, $D\cap\partial\Omega=\Phi$, $\mathcal{L}u+Vu\geq 0$ and $\mathcal{L}u_\lambda+Wu_\lambda=0$ in D. Therefore, by the comparison theorem 2.1 in Ref. 11 (see also the remark below), $u\geq u_\lambda$ in D, hence D is empty and (9) follows. Since $\|u\|=\|u_\lambda\|=r$, (9) is impossible unless $u\equiv u_\lambda$ and the theorem follows from here.

Remark: In order to better understand how (9) follows from (1) and (8) we give a heuristic argument. For simplicity we take the following particular case: $\mathcal{L}=-\Delta$, $a=1$, $f(x,u)=u^2$. Then (1) and (8) read,

$$-\Delta u_\lambda+u_\lambda^2-\lambda u_\lambda=0 \tag{10}$$

and

$$-\Delta u+u^2-\lambda u\geq 0. \tag{11}$$

Multiplying (10) by u, (11) by u_λ and subtracting we get

$$-\nabla(u_\lambda^2\nabla g)+u_\lambda^2 ug\geq 0, \tag{12}$$

where

$$g=(u/u_\lambda)-1. \tag{13}$$

Since u_λ^2 and uu_λ^2 are positive, the operator $-\nabla(u_\lambda^2\nabla\cdot)+u_\lambda^2 u$. is elliptic. Thus (12) implies $g\geq 0$ and (13) implies $u\geq u_\lambda$.

A proof completely analogous to that of Theorem 1 gives the following upper bound

Theorem 2: Fix $r>0$ and let $u\in C^2(\bar{\Omega})$ be any function on Ω such that $u>0$ almost everywhere in Ω, $u=0$ on $\partial\Omega$ and $\|u\|=r$. Then,

$$\lambda(r)\leq\sup_{\bar{\Omega}}\{(\mathcal{L}u+f(x,u))/au\}, \tag{14}$$

where $\lambda(r)$ is the eigenvalue corresponding to the positive solution u_λ of Eq. (1) with norm $\|u_\lambda\|=r$.

We now give some applications of Theorems 1 and 2.

Example 1: Consider the equation

$$-\frac{d^2u}{dx^2}+u^3=\lambda u, \tag{15}$$

defined on $\Omega=(o,b)\subset\mathcal{R}$, with $u(o)=u(b)=0$. The positive solution to Eq. (15), with the above boundary conditions, is given parametrically by

$$u_\lambda(x)=\frac{2\sqrt{2}}{b}kK(k)\,\mathrm{sn}\left(\frac{2x}{b}K(k),k\right), \tag{16}$$

with

$$\lambda=(4/b^2)(1+k^2)K(k)^2,\quad 0\leq k<1. \tag{17}$$

Here and below K and E denote the complete elliptic integrals. From (16) we get the L^2-norm of u_λ,

$$\|u_\lambda\|=\{(8/b)K(k)[K(k)-E(k)]\}^{1/2}. \tag{18}$$

Thus, the graph $\lambda(\|u_\lambda\|)$ for the positive solution of (15) is given parametrically by equations (17) and (18) with $0\leq k<1$. λ is an increasing and convex function of $\|u_\lambda\|$, $\lambda(0)=(\pi/b)^2$. Moreover, as $\|u_\lambda\|\to 0 (k\to 0)$, $\lambda\approx(\pi/b)^2+(3\|u_\lambda\|^2/2b)$, whereas if $\|u_\lambda\|\to\infty$ $(k\to 1)$, $\lambda\approx\|u_\lambda\|^2/b$. In order to get a lower bound for $\lambda(\|u_\lambda\|)$ we use in Theorem 1 the trial function $u(x)=rb^{-1/2}$, all $x\in\Omega$. Note that $\|u\|=r$, so by Eq. (3) we get $\lambda(r)\geq r^2/b$. Also, trying $u(x)=(2/b)^{1/2}r\sin(\pi x/b)$ in Eq. (3) gives $\lambda(r)\geq(\pi/b)^2$. Therefore, $\lambda(r)\geq\max\{(\pi/b)^2, r^2/b\}$. As for upper bounds, trying $u(x)=(2/b)^{1/2}r\sin(\pi x/b)$ in Eq. (14) yields $\lambda(r)\leq(\pi/b)^2+2(r^2/b)$. So the curve $\lambda(\|u_\lambda\|)$, given parametrically by (17) and (18), lies between the curves $\lambda_-(r)=\max\{(\pi/b)^2, r^2/b\}$ and $\lambda_+(r)=(\pi/b)^2+2(r^2/b)$.

As the next example shows, if the nonlinearity in problem (1) is not monotone, Theorems 1 and 2 are not valid anymore.

Example 2: Consider the equation

$$-\frac{d^2u}{dx^2}-u^3=\lambda u \tag{19}$$

defined on $\Omega=(o,b)\subset\mathcal{R}$, with $u(o)=u(b)=0$. The problem (19) violates one of the hypothesis of Theorem 1 and 2, namely, $f(x,u)/u=-u^2$ is not strictly increasing (in fact, $-u^2$ is strictly decreasing). The positive solution to Eq. (19) is given parametrically by

$$u_\lambda(x)=(kK(k)/b)\cdot(8(1-k^2))^{1/2}\,\mathrm{sn}(2xK(k)/b)/dn(2xK(k)/b), \tag{20}$$

with

$$\lambda=4K(k)^2(1-2k^2)/b^2,\quad 0\leq k<1. \tag{21}$$

Therefore, the graph $\lambda(\|u_\lambda\|)$ is given parametrically by (21) and

$$\|u_\lambda\|=\{(8/b)K(k)[E(k)-(1-k^2)K(k)]\}^{1/2}. \tag{22}$$

λ is a decreasing and concave function of $\|u_\lambda\|$, $\lambda(o)=(\pi/$

$b)^2$. Moreover, as $\|u_\lambda\| \to 0\ (k \to 0)$, $\lambda \approx (\pi/b)^2 - (2\|u_\lambda\|^2/b)$, whereas if $\|u_\lambda\| \to \infty\ (k \to 1)$, $\lambda \approx -\frac{1}{16}\|u_\lambda\|^4$. Our purpose here is to show that Theorem 1 is not valid for this problem. In fact, let us assume (3) holds and take the trial function (with $\|u\| = r$) $u(x) = (2/b)^{1/2} r \sin(\pi x/b)$; then Eq. (3) gives $\lambda(r) \geqslant (\pi/b)^2 - (2r^2/b)$ which contradicts the asymptotic behavior $\lambda(r) \approx -\frac{1}{16}r^4$, for large r, found above.

Example 3: (Monotonicity of λ in $\|u_\lambda\|$). It is well known[10] that u_λ is strictly increasing with λ, i.e., if $\lambda < \nu$, then $u_\lambda < u_\nu$ in Ω. This, of course, implies that $\|u_\lambda\|$ is strictly increasing with λ. This last fact can also be obtained directly from Theorem 1: we now show that $s > r$ implies $\lambda(s) \geqslant \lambda(r)$. Denote by u_λ the positive solution of (1) with norm $\|u_\lambda\| = r$ [and thus, eigenvalue $\lambda(r)$], and choose $u = (s/r)u_\lambda$ as a trial function on Eq. (3). Note that $\|u\| = s$, and $u > u_\lambda$ in Ω, which implies $f(x, u)/u \geqslant f(x, u_\lambda)/u_\lambda$ in Ω. So,

$$[\mathscr{L}u + f(x,u)]\frac{1}{au} = \frac{\mathscr{L}u_\lambda}{au_\lambda} + \frac{f(x,u)}{au}$$

$$= \lambda(r) - \frac{f(x,u_\lambda)}{au_\lambda} + \frac{f(x,u)}{au}$$

$$\geqslant \lambda(r), \tag{23}$$

where the second equality follows from Eq. (1). Introducing (23) in Eq. (3) we find $\lambda(s) \geqslant \lambda(r)$, hence λ is monotone nondecreasing in $\|u_\lambda\|$.

In the Ex. 1 above, λ is not only an increasing function of $\|u_\lambda\|$ but it is also convex. We conjecture that λ is a convex function of $\|u_\lambda\|$ for the general problem (1), at least for convex domains Ω.

Note added in proof: Equation (8) above says that u is an upper solution for Eq. (1) with $\lambda = \lambda(r)$. Form here (9) follows.

ACKNOWLEDGMENT

We thank Dr. Jorge Krause for encouragement and interest in our work.

[1] M. H. Protter and H. F. Weinberger, *Maximum Principles in Differential Equations* (Prentice-Hall, Englewood Cliffs, N.J., 1967), Chap. 2, Sec. 8, p. 92.

[2] Strictly speaking there is a lower bound of the Rayleigh–Ritz type, namely, Temple's inequality. See M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV: Analysis of Operators* (Academic, New York, 1978), Theorem XIII.5, p. 84.

[3] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. I.

[4] M. F. Barnsley, "Lower Bounds for Quantum Mechanical Energy Levels," J. Phys. A: Math. Gen. **11**, 55 (1978).

[5] M. S. Berger, *Nonlinearity and Functional Analysis* (Academic, New York, 1977).

[6] P. L. Lions, "Some Remarks on Hartree Equation," Nonlinear Anal. Theory, Methods Applications **5**, 1245 (1981).

[7] R. Benguria, H. Brezis, and E. H. Lieb, "The Thomas–Fermi–von Weizsäcker Theory of Atoms and Molecules," Commun. Math. Phys. **79**, 167 (1981).

[8] M. C. Depassier and E. A. Spiegel, "The Large Scale Structure of Compressible Convection," Astron. J. **86**, 496 (1981).

[9] H. Amann, "Fixed Point Theorems and Nonlinear Eigenvalue Problems in Ordered Banach Spaces," SIAM Rev. **18**, 620 (1976). See also, D. H. Sattinger, *Topics in Stability and Bifurcation Theory, Lecture Notes in Mathematics 309* (Springer-Verlag, New York, 1973).

[10] H. Berestycki, "Le Nombre de Solutions de Certains Problemes Semilinéaires Elliptiques," J. Func. Anal. **40**, 1 (1981).

[11] T. Hoffmann-Ostenhof, "A Comparison Theorm for Differential Inequalities with Applications in Quantum Mechanics," J. Phys. A: Math. Gen. **13**, 417 (1980).

Evolution equations with high order conservation laws

L. Abellanas and A. Galindo
Department of Mathematical Methods in Physics and Department of Theoretical Physics, Universidad Complutense, Madrid-3, Spain

(Received 31 December 1980; accepted for publication 15 October 1982)

An asymptotic algorithm presented in a previous paper is applied to investigate the possible structures of evolution equations $u_t = u_M + K(u,...,u_{M-1})$, $M = 3,5$, which could be compatible with the existence of a conserved density $\rho_0(u)$, depending only on u, and with the existence as well of conserved densities with arbitrarily high-order derivatives. For $M = 3$ it is shown that the Calogero–Degasperis–Fokas equation is essentially the only nonpolynomial equation of that type. For the case $u_t = D[u_4 + Q]$, with $Q(u,...,u_3)$ a polynomial, we find a very narrow class of admissible structures for Q, typified by the few particular examples known up to date. Actually, there is in this case an essentially unique structure, modulo a Miura transformation.

PACS numbers: 02.30.Jr

INTRODUCTION

While the existence of some particular evolution equations (e.e.'s) having an infinite number of conservation laws (c.l.'s) was well established a long time ago,[1] it is only very recently that serious attempts have been made to classify those e.e.'s sharing that relevant and striking property. Restricting ourselves to the third- and fifth-order e.e.'s the situation seems to be as follows.

Third-order polynomial e.e.'s include of course the celebrated KdV equation together with its modified and potential versions. In fact it was proven in Ref. 2 under rather mild hypothesis that these two polynomial nonlinear e.e.'s are unique. More explicitly there are not other polynomial nonlinear third-order e.e.'s $u_t = u_3 + f(u,u_1,u_2)$ with polynomial conserved densities of both low $[\rho_0(u)]$ and high $[\rho(u,...,u_N), N \geqslant 1]$ order. Our theorem can be generalized to e.e.'s of the form $u_t = P(u,u_1,u_2,u_3)$, P polynomial, with the additional assumption that it admits at least one nontrivial, i.e., different from $\alpha u_x + \beta u_t$, $\alpha,\beta \in \mathbb{R}$, symmetry. This is an obvious consequence of Theorem A in Ref. 3.

On the other hand, nonpolynomial third order e.e.'s have received little attention both because of the greater difficulty involved and the fact that no relevant examples motivated such a general analysis. However, the situation has changed with the discovery by Calogero, Degasperis, and Fokas[4,5] of a new type of equation

$$u_t = u_3 + (\mu e^{au} - \nu e^{-au})u_1 - (a^2/8)u_1^3 + cu_1,$$
$$\mu,\nu,a,c \in \mathbb{R} \qquad \text{(CDF)}$$

which can be written as $u_t = \Phi(u)u_x$, with an adequate hereditary symmetry Φ. Such an equation also has an infinite number of conservation laws, is related to the modified KdV equation through a Bäcklund transformation,[5b] and reduces to the potential modified KdV for $\mu = \nu = c = 0$.

In Sec. 2 we prove that the $\mathrm{KdV}_{0,1,2}$ and (CDF) equations are essentially the unique e.e.'s of the form $u_t = u_3 + F(u,u_1,u_2)$ which are *rare* in the sense specified in Sec. 1 and which conserve some nonconstant $\rho_0(u)$ (Theorem 1). This last restriction can be dropped out for the family $u_t = u_3 + f(u,u_1)$, and the resulting collection of e.e.'s turns out to be the same, except for the additional inclusion of the PKdV (Theorem 1').

Concerning fifth-order e.e.'s the situation reduces to the knowledge of a few polynomial e.e.'s which have been shown to admit an infinite number of c.l.'s.

This seems to be true for equations like

$$\text{(E1)}\quad u_t = u_5 + \alpha u_3 + \beta u u_1, \quad \text{Ref. 6;}$$

$$\text{(E2)}\quad u_t = u_5 + \alpha u u_3 + \beta u_1 u_2 + \gamma u^2 u_1, \quad \text{Refs. 7–10 and 5b;}$$

$$\text{(E3)}\quad u_t = u_5 - 5(u_1u_3 + u^2u_3 + u_2^2 + 4uu_1u_2 + u_1^3 - u^4u_1), \quad \text{Ref. 11;}$$

$$\text{(E4)}\quad u_t = u_5 + \pi(u)u_3 + 2\pi'(u)u_1u_2 + (\pi''/2)u_1^3 + h(u)u_1, \quad \text{Ref. 12;}$$

$$\pi(u) \equiv (\alpha/2)u^2 + \beta u + \gamma, \quad h(u) = \tfrac{3}{10}\pi^2(u) + \text{const}.$$

This has been proved only for some particular values of the parameters α,β,γ in case (E2) and in full generality for (E4). The last type of equation (E4) is obtained from the hereditary symmetry Φ_4 in Ref. 12 as $u_t = \Phi_4^2 u_1$.

Equations (E1)–(E4) are polynomial e.e.'s of the form $u_t = u_5 + D[Q(u,u_1,u_2,u_3)]$. In Sec. 3 we analyze those e.e.'s of this type which are rare and admit some polynomial c.d. $\rho(u,...,u_N)$, with $\rho_{u_N u_N} \neq 0$, $N \geqslant 7$. Theorem 2 classifies them into three families which reduce to a single one under a Miura transformation.

The singularization[13] of the Gardner family of equations $u_t = u_3 + (\alpha u^2 + \beta u + \gamma)u_1$, via the prolongation structure techniques of Wahlquist–Estabrook,[13–16] among the e.e.'s $u_t = u_3 + f(u)u_1$ admitting a simple pseudopotential, could lead one to suspect that results like ours might also be recovered by a suitable application of those techniques. An alternative isolation of the Gardner family and the (CDF) equation among the e.e.'s $u_t = u_3 + g(u,u_1)$ has been independently obtained by requiring the existence of one fifth-order infinitesimal symmetry.[5a]

The question of whether or not (i) the existence of simple pseudopotentials,[14] (ii) the existence of some generalized symmetry, and (iii) the existence of an infinite number of nontrivial conserved densities $\rho(u,...,u_N)$ of arbitrarily high order, are equivalent in some sense is an open, deep and intriguing problem, which deserves further investigation. However, concerning (i) versus (iii), the evolution equation

 0022-2488/83/030504-06$02.50

$u_t = u_3 + uu_2 + u_2 + u_1^2 + uu_1$ represents a case which admits a simple pseudopotential (namely, $q_x = u - q$, $q_t = u_2 + uu_1$), conserves $\rho_0(u) = u$, but according to Theorem 1 has not an infinite number of nontrivial c.d.'s with arbitrarily high-order derivatives. Therefore (i) and (iii) are actually inequivalent in a strict sense. Up to this moment, our Theorems 1 and 2 largely surpass in the size of the sample family all characterization results so far obtained for e.e.'s $u_t = u_M + K(u,...,u_{M-1})$, $M = 3,5$ by other methods.

The proofs of Theorems 1, 1', and 2 make extensive use of an asymptotic method whose main formulas are recalled in the preliminary Sec. 1. Auxiliary information is relegated to Appendices A, B, and C.

1. THE ASYMPTOTIC ALGORITHM

Let us briefly summarize the essentials of the asymptotic algorithm first presented in Ref. 2 and expanded in Ref. 17, which will be used in the rest of this paper.

Given an e.e. for $u = u(x,t) \in \mathbb{R}$, $(x,t) \in \mathbb{R}^2$,

$$u_t = f(u,...,u_M), \quad M \text{ odd} \geqslant 3, \tag{1}$$

where $u_k \equiv D^k u$, $D \equiv x$-total derivative, we shall say that f is *rare* when Eq. (1) admits c.d.'s $\rho(u,...,u_N)$, with $\rho_{u_N u_N} \neq 0$, for infinitely many values of N.

It is known[2] that such ρ's have the asymptotic expression

$$\rho \sim \frac{1}{2} \sum_{j=0}^{J} a^{(j)}(u,...,u_{M+2j}; N) u_{N-j}^2 + \tilde{\rho}(u,...,u_{N-J-1}; N) \tag{2}$$

for all $J \ll N$ and $N \gg 1$. (The notation $a \sim b$ has the usual meaning of $a - b \in \operatorname{Ran} D$).

The analysis in Ref. 2 leads to the following conditions for $a^{(j)}$:

$$\sum_{j=0}^{n} \sum_{m=0}^{n-j} c_{j,m}^{(n)} \Bigg[\Bigg\{ \sum_{i=0}^{j} a^{(i)} u_{N-i} \times \sum_{k=0}^{j-i} \binom{N-i}{k} \left[f_{u_{M+i-j+k}} \right]_k \Bigg\}_{u_{N+j+m-n}} \Bigg]_m = h^{(n)}, \quad n = 1,2,\cdots \tag{3}_n$$

with

$$c_{j,m}^{(n)} \equiv (-1)^j \left[\binom{M+N-j}{m} + (-1)^{M+n+m} \binom{N-n+j+m}{m} \right], \tag{4}$$

$$h^{(n)} \equiv \sum_{\substack{r,s \geqslant 0 \\ M+2r+s=n}} (-1)^r \binom{N-r}{s} D^s Y a^{(r)}, \tag{5}$$

where

$$Y \equiv \sum_{l \geqslant 0} [D^l f] \frac{\partial}{\partial u_l}. \tag{6}$$

We remark that (5) contains some new terms overlooked in Ref. 2, and which actually intervened only for high level obstructions ($n \geqslant 9$), unaffecting thus Proposition 3 in Ref. 2.

In particular, when $f_{u_M} = 1$, $f_{u_{M-1}} = 0$, the coefficients $a^{(j)}$ satisfy the recursive relations

$$a^{(n)} = \sum_{j=0}^{n-1} \Gamma^{nj} a^{(j)}, \quad n \geqslant 1, \quad a^{(0)} = \text{const} \neq 0, \tag{7}$$

in terms of the differential operators

$$\Gamma^{nj} = (-1)^{n-j} M^{-1} \Bigg\{ \sum_{m=0}^{2(n-j)+1} \times \sum_{k=0}^{2(n-j)+1-m} \alpha_{njmk} D^{m-1} \left[D^k f_{u_{M-(2n+1-m-2j-k)}} \right] + \sum_{s>0} \beta_{njs} D^{s-1} Y \Bigg\} \tag{8}$$

with

$$\alpha_{njmk} \equiv \left[\binom{N+M-(2n+1-m-j)}{m} (-1)^m + \binom{N-j}{m} \right] \binom{N-j}{k}, \tag{9}$$

$$\beta_{njs} \equiv \delta_{2(n-j),M-1+s} \binom{N-j}{s}. \tag{10}$$

Some of these Γ^{nj} are explicitly given in Appendix A.

As indicated in Ref. 17, for f (with $f_{u_M} = 1, f_{u_{M-1}} = 0$) to be rare, the system (3) should not present any obstruction to integrability, which amounts to

$$\sum_{j=0}^{n-1} (-1)^j \Bigg\{ \sum_{k=0}^{2(n-j)+1} \alpha_{nj0k} \left[D^k f_{u_{M-(2n+1-2j-k)}} \right] + \beta_{nj0} Y \Bigg\} a^{(j)} \sim 0 \tag{11}_n$$

for all $n \geqslant 1$, and infinitely many N's.

Appendices B and C contain the lowest ($n \leqslant 4$) obstructions for $M = 3,5$, respectively, in explicit form.

2. THIRD-ORDER RARE EVOLUTION EQUATIONS

The purpose of this section is to isolate the rare e.e.'s of third order, without restriction to polynomialness. This seems relevant due to the existence of the (CDF) equation.

The question is particularly important since one might suspect that the polynomial requirement could be responsible for the uniqueness of the KdV and modified KdV in results like Proposition 3 in Ref. 2. The next theorem provides an adequate answer to this type of question.

Theorem 1: Let

$$u_t = u_3 + F(u,u_1,u_2) \tag{12}$$

be a rare evolution equation, with some nontrivial c.d. $\rho_0(u)$. Then F can be transformed (by a suitable change of the dependent variable if necessary) into one of the following forms:

$$\text{(a) } \mathrm{KdV}_{0,1,2}: F = (\alpha u^2 + \beta u + \gamma) u_1, \tag{13}$$

$$\text{(b) CDF: } F = (\beta e^{\alpha u} - \gamma e^{-\alpha u}) u_1 - (\alpha^2/8) u_1^3 + \delta u_1, \tag{14}$$

with $\alpha, \beta, \gamma, \delta \in \mathbb{R}$.

Proof: Condition $(3)_1$ leads to $F_{u_2} = (3/2) D \ln a^{(0)}$, whence $a^{(0)} = a^{(0)}(u,u_1)$, and

$$F = A(u,u_1) u_2^2 + B(u,u_1) u_2 + C(u,u_1), \quad A \equiv (3/4)(\ln a^{(0)})_{u_1}. \tag{15}$$

Now, conservation of $\rho_0(u)$ forces $a^{(0)} = a^{(0)}(u)$ and

$$F = (\ln a^{(0)})_u u_1 u_2 + C(u,u_1). \tag{16}$$

It is a simple exercise to check that the u_2 term in $u_t = u_3 + F$ disappears under the rarity-preserving change $u \to v = \psi(u)$, with $\psi'(u) = [a^{(0)}(u)]^{1/3}$. Therefore we can restrict our considerations to

$$u_t = u_3 + F(u,u_1) \tag{17}$$

for which the obstructions (Bn) in Appendix B apply. For this e.e., $a^{(0)} = \text{const}(\neq 0)$, and without loss of generality we choose $a^{(0)} = 1$.

(B1) $\Rightarrow \exists H(u)$ such that $F_u = H'(u)u_1$, i.e.,

$$F = H(u)u_1 + G(u_1), \tag{18}$$

$$\text{(B2)} \Rightarrow \begin{cases} G^{\text{IV}} = 0, \text{ i.e., } G = \dfrac{\alpha}{6}u_1^3 + \dfrac{\beta}{2}u_1^2 + \gamma u_1 + \delta, \\ H''' = -\dfrac{4\alpha}{3}H', \\ \beta H' = \delta H' = 0, \end{cases} \tag{19}$$

whose only solutions are $\text{KdV}_{0,1,2}$, CDF, and $F = G(u_1)$, with $G^{\text{IV}} = 0$.

As to the last possibility, note that the equation $u_t = u_3 + (\alpha/6)u_1^3 + (\beta/2)u_1^2 + \gamma u_1 + \delta$ will conserve some $\rho_0(u)$ only if $\beta = \delta = 0$, leaving thus a special case of CDF. Q.E.D.

Remark: After some increasingly tedious calculations, the reader can convince himself that (13) and (14) automatically satisfy (B2b)–(B4b), owing to some rather miraculous cancellations. However, this is not surprising at all, since it is well known that the $\text{KdV}_{0,1,2}$ and CDF equations are rare. That these are essentially the only rare equations of the form (12) with a conserved $\rho_0(u)$ (for instance, conservative equations), is the main conclusion of Theorem 1.

Furthermore, the imposition of a conserved $\rho_0(u)$ can be dispensed with provided that the attention is focused on the important family $u_t = u_3 + f(u,u_1)$. The previous arguments from (17) on lead immediately to the following result.

Theorem 1': The only rare evolution equations of the form

$$u_t = u_3 + f(u,u_1), \quad f(0,0) = 0 \tag{20}$$

are the $\text{KdV}_{0,1,2}$, $\text{PKdV}_{0,1,2}$, and CDF equations.

3. POLYNOMIAL FIFTH-ORDER RARE EVOLUTION EQUATIONS

Let us now characterize those rare e.e.'s

$$u_t = u_5 + D[Q(u,u_1,u_2,u_3)] \equiv P \tag{21}$$

with the properties: (i) P is a polynomial, (ii) (21) has some polynomial c.d. $\rho(u,\ldots,u_N)$, with $\rho_{u_N u_N} \neq 0$, $N \geqslant 7$.

Theorem 2: An e.e. with the above properties has necessarily the form

$$\begin{aligned} u_t = u_5 &+ [\lambda u_1 + q(u)]u_3 + \lambda u_2^2 \\ &+ [2q'(u) + k\delta_{\lambda,0}\delta_{\tau,0}]u_1 u_2 + (q''(u)/2)u_1^3 \\ &+ [(1 - \delta_{\lambda,0}\delta_{\tau,0})\chi q^2(u) + h(u)]u_1, \end{aligned} \tag{22}$$

where

(1) $q(u) = (\tau/2)u^2 + \tau' u + \tau''$, $\tau, \tau', \tau'' \in \mathbb{R}$,

(2) $h(u)$ is (at most) quadratic in u,

(3) $\lambda, \chi \in \mathbb{R}$.

Moreover, only three different possibilities are admissible:

(a) $\lambda \neq 0$, $\tau = -\frac{2}{3}\lambda^2$, $\chi = \frac{1}{3}$,

(b) $\lambda = 0$, $\tau \neq 0$, $\chi = \frac{3}{10}$, $h(u) = \text{const}$,

(c) $\lambda = \tau = 0$, $(1 - \delta_{k,0})[h''(u) - \frac{1}{25}(2\tau' + k)(6\tau' - 4k)] = 0$.

Remark: An example of type (a) is given in Ref. 11 with $\lambda = -5$, $\tau' = \tau'' = 0$, $h = 0$. On the other hand the equation $u_t = \Phi^2 u_x$ obtained by means of the hereditary symmetry[5b] $\Phi = D^2 + (\alpha u^2 + 2\beta u + \gamma) + u_1 D^{-1}(\alpha u + \beta)$, belongs to class (b) with $\tau = 5\alpha$, $\tau' = 5\beta$, $\tau'' = 2\gamma$. Finally class (c) includes a number of fifth-order equations which appear scattered in the work of many people. References 5–12 contain a partial collection of this type.

Proof: It is a straightforward application of (Cn) in Appendix C. We only give those partial steps which are useful to follow the details of the argument. First of all, we know from Ref. 2 that (ii) $\Rightarrow Q = Q(u,u_1,u_2,)$.

Now (C1) forces the existence of some $F = F(u,u_1)$ such that

$$Q_{u_1} = D[F(u,u_1)]. \tag{23}$$

We investigate separately the cases (I) $F_{u_1} \neq 0$ and (II) $F = F(u)$.

Case I ($F_{u_1} \neq 0$): Differentiating (23) with respect to u_2 one gets $Q_{u_1 u_2} = F_{u_1}$, i.e.,

$$Q_{u_2} = F(u,u_1) + G(u,u_2). \tag{24}$$

Therefore (C2) implies

$$G(u,u_2)DF \sim 0. \tag{25}$$

A close look at (25) shows that G must be constant. Therefore $Q = F(u,u_1)u_2 + A(u,u_1)$. Next condition (C3a) requires $P(F_u - DF_{u_1}) \sim 0$, and hence

$$F = \lambda(u)u_1 + q(u), \quad q(u) = (\tau/2)u^2 + \tau' u + \tau'', \quad \tau, \tau', \tau'' \in \mathbb{R}. \tag{26}$$

It is easy to check that (C3b) is trivially fulfilled. On the other side (C4a) reads as follows:

$$\begin{aligned} P[&2qq' - 2\lambda\lambda' u_1^2 - 5\tau u_2 - 2\lambda^2 u_2 \\ &- 5\lambda'' u_1 u_2 - 5A_{uu}(u,u_1)] \sim 0. \end{aligned} \tag{27}$$

The vanishing of the successive dominant terms (after integration by parts) in $u_2 u_3^2$ and $u_1 u_3^2$ leads to

$$\left.\begin{aligned} &\lambda(u) = \text{const} \equiv \lambda, \\ &A = \frac{\tau u + \tau'}{2}u_1^2 + a(u). \end{aligned}\right\} \tag{28}$$

The u_2^3 terms in (27) provide us with the relation $\lambda(\lambda^2 + \frac{5}{2}\tau)u_2^3 \sim 0$. Since $\lambda \neq 0$ (assumption $F_{u_1} \neq 0$) we get

$$\tau = -\tfrac{2}{3}\lambda^2. \tag{29}$$

Next dominant term is $u_1 u_2^2$, which forces $a'(u)$ to be

$$a'(u) = \frac{\tau^2}{20}u^4 + \frac{\tau\tau'}{5}u^3 + \frac{\alpha}{2}u^2 + \beta u + \gamma . \tag{30}$$

The remaining of (27) reduces to

$$(5\alpha - 2\tau'^2 - 2\tau\tau'')uP \sim 0 . \tag{31}$$

Two unique solutions are possible:

(I_1) $\tau = \tau' = 0 \Rightarrow u_t = u_5 + \tau'' u_3 + a'(u)u_1$, with quadratic $a'(u)$. This is a particular case of class (c) in the Theorem.

(I_2) $\tau\tau' \neq 0$, $\alpha = \frac{2}{5}(\tau'^2 + \tau\tau'')$. This gives class (a) in our Theorem.

A rather tedious computation shows that both I_1 and I_2 families of equations automatically fulfill the condition (C4b).

Case II [F = F(u)]: From (23) and (C2) we conclude

$$Q = G(u,u_2) + \tfrac{1}{2}F'(u)u_1^2, \quad G_{u_2}F'(u)u_1 \sim 0 . \tag{32}$$

As a consequence G must be (at most) quadratic in u_2 [this is obvious from (32) if $F'(u) \neq 0$; otherwise (C3a) implies $G_{u_2u_2u_2}u_4^2 \sim 0$, i.e., $G_{u_2u_2u_2} = 0$].

Therefore we put

$$G(u,u_2) = A(u)u_2^2/2 + B(u)u_2 + C(u) . \tag{33}$$

Condition (C3a) is dominated by $5A''u_1u_3^2$ and $(A')^2u_1u_2^3$. In consequence $A = \text{const}$. Two different alternatives are open: (II_1) $A \neq 0$, (II_2) $A = 0$.

Subcase $(II)_1$, $A \neq 0$: When explicitly written (32) implies $AF'(u) = \text{const}$, hence $F'' = 0$. Having this in mind (C3a) yields $B''(u) = 0$, and moreover from (C3b) it follows that $F' = -B'$, $C''' = -\frac{4}{5}(B')^2$. Finally (C4a) requires

$$[4AB'u_3 + 2(B')^2u_1 - 5C'''u_1 + 2A^2u_5]\,Q \sim 0 \tag{34}$$

which is dominated by $A^3u_3^3$. As a consequence $A \neq 0$ is untenable.

Subcase $(II)_2$, $A = 0$: A simple calculation based on (C3a) provides us with the following restrictions:

$$\begin{aligned} &B''' = 0, \quad \text{i.e.,} \quad B(u) = \beta u^2 + \gamma u + \delta, \\ &\beta(F' - B') = 0. \end{aligned} \tag{35}$$

There are two alternative solutions $\beta = 0$ and $F' = B'$, which we pass to investigate.

Analysis of $(II)_2$ with $\beta = 0$: With these hypothesis the candidate equation reads

$$\begin{aligned} u_t = {} & u_5 + (\gamma u + \delta)u_3 + (F'(u) + \gamma)u_1u_2 \\ & + (F''/2)u_1^3 + C'(u)u_1 . \end{aligned} \tag{36}$$

Now (C3b) and (C4a) force $F(u)$ to be linear in u, and $C'(u)$ quadratic (at most) in u. Let us write

$$F' \equiv \rho - \gamma, \quad C'(u) \equiv (\rho_0/2)u^2 + \rho_1 u + \rho_2 . \tag{37}$$

Then we get from (C4b),

$$125\rho_0(\gamma - \rho/2) = \rho(10\rho^2 + 70\gamma^2 - 55\rho\gamma) \tag{38}$$

which is trivially satisfied if $\rho = 2\gamma$. Otherwise (38) yields

$$\rho_0 = (\rho/25)(14\gamma - 4\rho) . \tag{39}$$

Both alternatives fall into class (c).

Analysis of $(II)_2$ with $F' = B'$: Condition (C3b) is automatically fulfilled. Now, (C4a) reduces to

$$[10\beta u_2 + 5C''(u) - 2B(u)B'(u)]P \sim 0 \tag{40}$$

which is dominated by $u_1u_2^2$ terms. The vanishing of the associated coefficient forces $C^{IV}(u) = \frac{18}{5}\beta(2\beta u + \gamma)$; hence

$$C'(u) = \tfrac{3}{10}\beta^2u^4 + \tfrac{3}{5}\beta\gamma u^3 + (\sigma/2)u^2 + \sigma' u + \sigma'', \quad \sigma, \sigma', \sigma'' \in \mathbf{R} . \tag{41}$$

The rest of (40) gives a polynomial of first degree in u, which is identically zero for $\beta = 0$. If, on the contrary, $\beta \neq 0$ then we conclude from (40) that

$$\begin{aligned} \sigma &= \tfrac{6}{5}\beta\delta + \tfrac{3}{5}\gamma^2 , \\ \sigma' &= \tfrac{3}{5}\gamma\delta , \end{aligned} \tag{42}$$

or, in other words, $C'(u) = \frac{3}{10}F^2(u) + \text{const}$.

We see that $\beta = 0$ corresponds to equations within class (c) with $k = 0$, while $\beta \neq 0$ determines the Fuchssteiner-like equations within class (b).

It is an easy exercise to check that both resultant equations verify the additional restriction imposed by (C4b).

The proof of Theorem 2 is now complete. Q.E.D.

Recalling that KdV_1 and KdV_2 are related through the remarkable Miura transformation[18] and thus KdV_1 stands up as the essentially unique (nonlinear) polynomial rare conservative e.e. of third order, one might wonder whether the three families (a), (b), and (c) distinguished by Theorem 2 can be suitable linked. In fact, the Miura transformation $u \to u^2 + \alpha u_x + \beta u$ brings family (c),

$$\begin{aligned} &-u_t + u_5 + (\tau' u + \tau'')u_3 + (2\tau' + k)u_1u_2 + ((\rho_0/2)u^2 \\ &\quad + \rho_1 u + \rho_2)u_1 = 0 , \end{aligned} \tag{43}$$

into

$$\begin{aligned} &(2u + \alpha D_x + \beta)[-u_t + u_5 + (\alpha k u_1 + \tau' u^2 \\ &\quad + \beta\tau' u + \tau'')u_3 + \alpha k u_2^2 + (4\tau' u + 2\beta\tau')u_1u_2 + \tau' u_1^3 \\ &\quad + \{(\rho_0/2)u^4 + \rho_0\beta u^3 + (\rho_1 + (\rho_0/2)\beta^2)u^2 \\ &\quad + \rho_1\beta u + \rho_2\}u_1] = 0 \end{aligned} \tag{44}$$

provided that

$$\begin{aligned} &\alpha^2(k - \tau') = 10 , \\ &5\rho_0 = (\tau' - k)(3\tau' + 2k) , \\ &10\rho_1 = (\tau' - k)(6\tau'' + k\beta^2) . \end{aligned} \tag{45}$$

Equation (44), with the restrictions (45), leads either to (a) (if $k \neq 0$) or (b) (if $k = 0$). Once more an essential uniqueness emerges for the polynomial conservative rare e.e.'s of fifth order, and this is again accomplished via the Miura transform. (A particular example of this method was exhibited in Ref. 11.)

ACKNOWLEDGMENTS

We thank Professor Degasperis for calling our attention to Ref. 4 prior to its publication, and Professor Fokas and the anonymous referee for some interesting remarks. In the algebraic computation of which the Appendices contain just a small sample, the generous help of Professor J. Martínez has been invaluable. The financial support of the Comisión Asesora de Investigación Científica y Técnica is finally acknowledged.

APPENDIX A

The lowest differential operators Γ^{nj} for the cases $M = 3,5$, written in terms of the auxiliary operators

$$\overline{\Gamma}^{nj} \equiv (-M)^{n-j}\Gamma^{nj}$$

are the following:

Case $M = 3$

$$\overline{\Gamma}^{10} \simeq [-1+2N] f_{u_{M-2}} + 2D^{-1} f_{u_{M-3}},$$

$$\overline{\Gamma}^{20} \simeq \left[12\binom{N}{2} + 24\binom{N}{3}\right] D^2 f_{u_{M-2}} + \left[3N + 12\binom{N}{2}\right] D f_{u_{M-3}},$$

$$\overline{\Gamma}^{21} = D^{-1}Y + 2D^{-1} f_{u_{M-3}} + (-2+2N) D^{-1}(D f_{u_{M-2}}) - f_{u_{M-2}} - \left[1 + 3\binom{N}{2}\right] D^2,$$

$$\overline{\Gamma}^{30} \simeq \left[54\binom{N}{3} + 288\binom{N}{4} + 288\binom{N}{5}\right] D^4 f_{u_{M-2}} + \left[9\binom{N}{2} + 108\binom{N}{3} + 144\binom{N}{4}\right] D^3 f_{u_{M-3}},$$

$$\begin{aligned}\overline{\Gamma}^{31} = {} & \left[3 - 3N + 3\binom{N}{2}\right]\{DY + 2D^{-1}(D^2 f_{u_{M-3}}) + (D^2 f_{u_{M-2}}) \\ & + 2D f_{u_{M-3}} + D^2 f_{u_{M-2}} - 2D^{-1}(D^3 f_{u_{M-2}}) - 2D(D f_{u_{M-2}}) - D^4\} \\ & + 6\binom{N}{3} D^{-1}(D^3 f_{u_{M-2}}) + 18\binom{N}{3} D(D f_{u_{M-2}}) - 9\binom{N}{4} D^4 + [-3+3N](D f_{u_{M-3}}),\end{aligned}$$

$$\overline{\Gamma}^{32} = D^{-1}Y + 2D^{-1} f_{u_{M-3}} + (-4+2N) D^{-1}(D f_{u_{M-2}}) - f_{u_{M-2}} + \left[-4 + 3N - 3\binom{N}{2}\right] D^2.$$

(The symbol $\simeq$ stands for equality when applied to $a^{(0)} = \text{const.}$)

Case $M = 5$

$$\overline{\Gamma}^{10} \simeq [-3+2N] f_{u_{M-2}} + 2D^{-1} f_{u_{M-3}},$$

$$\begin{aligned}\overline{\Gamma}^{20} \simeq {} & \left[-5 + 20\binom{N}{2} + 40\binom{N}{3}\right] D^2 f_{u_{M-2}} + \left[5 + 5N + 20\binom{N}{2}\right] D f_{u_{M-3}} \\ & + [-5+10N] f_{u_{M-4}} + 10D^{-1} f_{u_{M-5}},\end{aligned}$$

$$\overline{\Gamma}^{21} = 2D^{-1} f_{u_{M-3}} + [-2+2N] D^{-1}(D f_{u_{M-2}}) - 3 f_{u_{M-2}} - \left[5 + 5N + 5\binom{N}{2}\right] D^2,$$

$$\begin{aligned}\overline{\Gamma}^{30} \simeq {} & \left[150\binom{N}{3} + 800\binom{N}{4} + 800\binom{N}{5}\right] D^4 f_{u_{M-2}} + \left[25\binom{N}{2} + 300\binom{N}{3} + 400\binom{N}{4}\right] D^3 f_{u_{M-3}} \\ & + \left[100\binom{N}{2} + 200\binom{N}{3}\right] D^2 f_{u_{M-4}} + \left[25N + 100\binom{N}{2}\right] D f_{u_{M-5}},\end{aligned}$$

$$\begin{aligned}\overline{\Gamma}^{31} = {} & 5\{D^{-1}Y + 2D^{-1} f_{u_{M-5}} - f_{u_{M-4}}\} + [5-5N]\{(D f_{u_{M-3}}) - 2D^{-1}(D f_{u_{M-4}})\} \\ & + \left[5 - 5N + 5\binom{N}{2}\right]\{2D^{-1}(D^2 f_{u_{M-3}}) - 2D^{-1}(D^3 f_{u_{M-2}}) - (D^2 f_{u_{M-2}})\} \\ & + 10\binom{N}{3} D^{-1}(D^3 f_{u_{M-2}}) + \left[5 + 10\binom{N}{2}\right] D f_{u_{M-3}} + \left[-5 + 5N + 10\binom{N}{2} + 30\binom{N}{3}\right] D(D f_{u_{M-2}}) \\ & - \left[5 + 15\binom{N}{2}\right] D^2 f_{u_{M-2}} - \left[5 + 25\binom{N}{2} + 25\binom{N}{3} + 25\binom{N}{4}\right] D^4,\end{aligned}$$

$$\overline{\Gamma}^{32} = 2D^{-1} f_{u_{M-3}} + [-4+2N] D^{-1}(D f_{u_{M-2}}) - 3 f_{u_{M-2}} - \left[5 + 5\binom{N}{2}\right] D^2.$$

APPENDIX B

When $(11)_n$, $1 \leqslant n \leqslant 4$, is developed in detail for $M = 3$, and its fulfillment is required identically in N [note that the left-hand side of $(11)_n$ is a polynomial in N], it gives rise to the obstructions (Bn) given below:

$$f_u \sim 0, \tag{B1}$$

$$Yf_{u_1} \sim 0\,, \tag{B2a}$$

$$YD^{-1}f_u + f_u f_{u_1} \sim 0\,, \tag{B2b}$$

$$YD^{-1}Yf_{u_1} \sim 0\,, \tag{B3a}$$

$$YD^{-1}YD^{-1}f_u + YD^{-1}f_u f_{u_1} + f_{u_1} YD^{-1}f_u - 2(D^{-1}f_u)Yf_{u_1} \sim 0\,, \tag{B3b}$$

$$YD^{-1}YD^{-1}Yf_{u_1} + f_{u_1} YD^{-1}Yf_{u_1} - 2f_{u_1} YD^2 f_{u_1} + 3Y(f_u)^2 + 3Yf_{u_1} Df_u - (f_{u_1})^2 Yf_{u_1} \sim 0\,, \tag{B4a}$$

$$\begin{aligned} &-2f_{u_1} D^4 f_u - 10f_{u_1} f_u D^2 f_{u_1} - 15(f_u)^2 Df_{u_1} + 10(f_u)^3 \\ &\quad - \tfrac{10}{3}(f_{u_1})^3 f_u + 6f_{u_1} YDf_u + 4f_u DYf_{u_1} + 3Y(f_u)^2 \\ &\quad + 8f_u f_{u_1} D^{-1}Yf_{u_1} + 2YD^{-1}[YD^{-1}YD^{-1}f_u \\ &\quad + YD^{-1}f_u f_{u_1} + f_{u_1} YD^{-1}f_u \\ &\quad - 2(D^{-1}f_u)Yf_{u_1}] + 2f_{u_1} YD^{-1}[YD^{-1}f_u + f_u f_{u_1}] \\ &\quad + 4(D^{-1}Yf_{u_1})YD^{-1}f_u \\ &\quad - 4(D^{-1}f_u)YD^{-1}Yf_{u_1} \sim 0\,. \end{aligned} \tag{B4b}$$

APPENDIX C

Similarly, for $M = 5$:

$$f_{u_2} \sim 0\,, \tag{C1}$$

$$5f_u - f_{u_2} f_{u_3} \sim 0\,, \tag{C2}$$

$$Yf_{u_3} \sim 0\,, \tag{C3a}$$

$$\tfrac{2}{5} f_{u_2}(f_{u_3})^2 - YD^{-1}f_{u_2} + f_{u_3}[D^2 f_{u_2} - Df_{u_1} - f_u] - f_{u_1} f_{u_2} \sim 0\,, \tag{C3b}$$

$$Y[(f_{u_3})^2 - 5f_{u_1}] \sim 0\,, \tag{C4a}$$

$$\begin{aligned} &150 f_u f_{u_1} - 25Yf_{u_1} + f_{u_3}[70D^4 f_{u_2} - 100D^3 f_{u_1} \\ &\quad + 50D^2 f_u] + f_{u_2}[150Df_u - 100D^2 f_{u_1}] \\ &\quad + 10YD^{-1}[5f_u - f_{u_2} f_{u_3}] + 15Y(f_{u_3})^2 \\ &\quad + 10f_{u_3} YD^{-1}f_{u_2} - 20(D^{-1}f_{u_2})Yf_{u_3} - 50f_{u_2}(Df_{u_3})^2 \\ &\quad + (f_{u_3})^2[45D^2 f_{u_2} - 30Df_{u_1} - 30f_u] + (f_{u_2})^2 \\ &\quad \times [15Df_{u_3} - 10f_{u_2}] + f_{u_2} f_{u_3}[14(f_{u_3})^2 - 60f_{u_1}] \sim 0\,. \end{aligned} \tag{C4b}$$

[1] R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys **9**, 1204–1209 (1968).
[2] L. Abellanas and A. Galindo, J. Math. Phys. **22**, 445–448 (1981).
[3] G. Z. Tu, Comm. Math. Phys. **77**, 289–297 (1980).
[4] F. Calogero and A. Degasperis, J. Math. Phys. **22**, 23–31 (1981).
[5a] A. S. Fokas, J. Math. Phys. **21**, 1318–1325 (1980).
[5b] A. S. Fokas and B. Fuchssteiner, "Symplectic structures, their Bäcklund transformations and hereditary symmetries" (preprint, Caltech, 1980).
[6] T. Kawahara, J. Phys. Soc. Jap. **33**, 260 (1972).
[7] P. D. Lax, Comm. Pure Appl. Math. **21**, 467–490 (1968); M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **31**, 125–127 (1973).
[8] P. J. Caudrey, R. K. Dodd, and J. D. Gibbons, Proc. R. Soc. London Ser. A **351**, 407–422 (1976).
[9] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Comm. Pure Appl. Math. **27**, 97–133 (1974).
[10] D. J. Kaup, Stud. Appl. Math. **62**, 189–216 (1980).
[11] A. Fordy and J. Gibbons, J. Math. Phys. **21**, 2508–2510 (1980).
[12] B. Fuchssteiner, Nonlinear Analysis. Theory, Methods Appl. **3**, 849–862 (1979).
[13] F. B. Estabrook and H. D. Wahlquist, J. Math. Phys. **17**, 1293–1297 (1976).
[14] J. Corones, J. Math. Phys. **17**, 756–759 (1976).
[15] D. J. Kaup, "Lectures on the Estabrook–Wahlquist Method, with Examples of Application" [Plasma Research Report, PRR 79/6 (1979)].
[16] D. J. Kaup, Physica D **1**, 391 (1980).
[17] L. Abellanas and A. Galindo, Lett. Math. Phys. **6**, 391–394 (1982).
[18] R. M. Miura, J. Math. Phys. **9**, 1202 (1968).

Geometric construction and properties of some families of solutions of nonlinear partial differential equations. I[a)]

Robert Hermann
Association for Physical and Systems Mathematics, Brookline, Massachusetts 02146

(Received 24 May 1982; accepted for publication 5 November 1982)

This series of papers deals with "The 19th Century Theory of Partial Differential Equations from an Advanced Standpoint." In the treatises of Darboux, Goursat, and Forsythe one finds methods for classifying nonlinear differential equations according to the geometric properties of families of solutions. This work was used by Elie Cartan in his theory of exterior differential systems, and is involved in an indirect way in today's work on "nonlinear physics." I plan to present several major themes of the classical work (e.g., "general," "singular," and "complete" solutions, "intermediate integral") using geometric methods developed by Cartan, Vessiot, Ehresmann, and Spencer. My aim is to develop this material from a point of view that is both fundamental and directed toward its ultimate application. Another possible utilization is in the development of the symbolic computation computer systems such as MACSYMA and REDUCE. This first paper concentrates on a description of what is meant by a "general solution," and presents the classical Lagrange–Charpit method for constructing "general" solutions by means of a "complete solution." The Monge–Ampère equation is also treated in modernized form. A new geometric concept, that of a *Lagrange–Vessiot submersion*, is isolated from the classical literature.

PACS numbers: 02.30.Jr

I. INTRODUCTION

This paper is oriented both toward the revival in a geometric context of ideas of the 19th century theory of partial differential equations[1–6] and the development of certain material in the modern geometric theory of nonlinear partial differential equations that might be useful in today's "nonlinear" physics and engineering.

In the classical literature, one finds extensive discussion of various special families of solutions of nonlinear partial differential equations—"general," "singular," "complete,"... solutions, and various quasialgorithmic procedures for generating such classes of solutions. (For example, the methods of Lagrange–Charpit, Jacobi, Ampere, Darboux, Goursat, etc.) Often each of the solutions in these families is parametrized by submanifolds of auxillary manifolds assigned to the data. These auxillary manifolds (which are often only "generalized manifolds," in some appropriate sense, particularly as studied in the work of Ehresmann)[7] also carry some definite sorts of geometric structure.

One important example is the space of geodesics of a Riemannian metric on a manifold M: It is the quotient space of a foliation with two-dimensional leaves on the tangent bundle $T(M)$, and admits a symplectic structure.[8] This symplectic structure plays a basic role in many areas of mathematics and physics, e.g., ergodic theory, relativity, and twistor theory. In general, single integral calculus of variations problems (which, of course, involve ordinary differential equations) lead to foliations whose base space has a symplectic structure.

In this paper (and those to follow) I will be mainly concerned with the intrinsically more complicated case of partial differential equations. Any serious work on the modern geometric theory of partial differential equations inevitably involves an elaborate differential-geometric formalism, usually based on the theory of jets of mappings, as originated by C. Ehresmann.[7] I will begin this work with an exposition of some aspects of this theory, beginning with a treatment of the basic operation of *prolongation* of differential equations, as described in the work of E. Cartan, E. Vessiot, M. Kuranshi, and D. Spencer.

I also have in mind developing the "differential algebra" aspects of the formalism, going beyond the 19th century with the aid of symbolic computer systems such as MACSYMA and REDUCE. Just as progress in digital computers in the 1960's stimulated the revival of 19th century numerical analysis as a vital part of contemporary mathematical science, so, I believe, will the development of these computer systems lead to a revival of the 19th century theory of differential equations. The attractive feature of the classical work from the point of view of computer science is its emphasis on "algorithmic" methods for thinking about solvability and "integrability" of differential equations and its surprisingly sophisticated development of algebraic methods for description of the solutions.

This work has been incubating for over 25 years, and attempts to unify diverse material I have published in various forms in the intervening years. In 1953–1955 I studied with Ehresmann and Spencer. Ehresmann had a brilliant picture in his mind, which he could only partially communicate in print to his colleagues, of how the classical work (particularly that of Lie, Vessiot, and Cartan) could be integrated into modern mathematics. Unfortunately, Ehresmann never wrote an authoritative version of his ideas as they might apply to differential geometry in the conventional sense: His mathematical intuition was mainly topological and algebraic. Spencer,[9] coming from a background in analysis, saw

 0022-2488/83/030510-12$02.50

most profoundly how to utilize these ideas (plus others from the theory of sheaves and the theory of elliptic differential operators) into a machinery which could carry out the Lie, Vessiot, and Cartan ideas (particularly in the context of "pseudogroups") in a modern framework. I have also greatly benefited from the ideas and papers of many others. I will mention the following people here, and thank them: B. Cenkl, F. Estabrook, R. Gardner, H. Goldschmidt, H. Johnson, V. Guillemin, M. Kruskal, B. Kuperschmidt, W. Shadwick, J. Stachel, S. Sternberg, A. M. Vinogradov, and H. Wahlquist.

In this first paper of a series, I plan to lay foundations for applications to follow. I will also discuss two more specific topics; an elaboration of the treatment in Ref. 10 of the Monge–Ampère equations and their "intermediate integral," and a discussion of the 19th century ideas of "integrability" of a differential equation in terms of analytic functions.

This latter topic has recently and indirectly come to the foreground in mathematical physics in an amazing way. Painlèvé[11] analyzed which second-order nonlinear ordinary differential equations possessed general solutions in this form. He wrote down a list of such equations defining a new class of functions extending the elliptic functions, which are known to be associated with the soluble problems of classical analytical mechanics.[12]

Another fertile field for a geometric approach based on the material to be developed in this series is that of *bifurcation theory*, i.e., the dependence on external parameters of the solutions of differential equations. This is an important applied subject that clearly falls within the province of *deformation of geometric structures*, as developed by Ehresmann, Kodaira and Spencer, Kuranishi, and many others. I have presented some elementary material about this in Vol. XII of Interdisciplinary Mathematics[13] and plan a further development in this series.

II. THE EHRESMANN JET SPACES AND THE PROLONGATION CALCULUS

All manifolds will be C^∞, finite-dimensional, and paracompact, unless mentioned otherwise. All maps between manifolds will also be "smooth," i.e., C^∞.

Let X and Y be such manifolds. A *local map from X to Y* is a pair (U,ϕ) consisting of an open subset U of X and a C^∞ map $\phi: U \to Y$. Let $\mathscr{M}(X,Y)$ be the space of all triples of the form

$$(x,U,\phi)$$

where

$$x \in X, \quad U \text{ is an open subset of } X,$$

$$\phi: U \to Y \text{ is a map}, \quad x \in U.$$

Construct maps

$$\pi: \mathscr{M}(X,Y) \to X$$

$$\tau: \mathscr{M}(X,Y) \to Y$$

as follows:

$$\pi(x,\phi,U) = x, \tag{2.1}$$

$$\pi(x,\phi,U) = \phi(x). \tag{2.2}$$

π is called the *source map*, and τ the *target map*. We will construct the spaces that are basic to the Ehresmann theory as quotients of $\mathscr{M}(X,Y)$ under equivalence relations. The source and target maps π and τ will be constant on these equivalence classes, and hence will pass to the quotient to define maps from the space of equivalence classes to X and Y, respectively. For the sake of notational simplicity, we will use the same letter to denote these maps on the equivalence classes.

First, there is the following equivalence relation:

$$(x,\phi,U) \sim (x',\phi',U')$$

$$\text{iff} \quad x = x', \ \phi = \phi' \text{ on an open subset of } U \cap U' \text{ containing } x. \tag{2.3}$$

The quotient of $\mathscr{M}(X,Y)$ under the equivalence relation (2.3) (i.e., the space of all equivalence classes) is alled the *sheaf* of all maps: $X \to Y$, denoted as $\mathscr{S}(X,Y)$.

For each integer r, we define an equivalence relation as follows:

$$(x,\phi,U) \sim (x',\phi',U')$$

$$\text{iff} \quad x = x', \ \phi \text{ and } \phi' \textit{ meet to order } r \textit{ at } x, \text{ in the sense that, in a local coordinate system for } X \text{ about } x, \text{ the partial derivatives at } x \text{ of } \phi \text{ and } \phi' \text{ agree to order } r. \tag{2.4}$$

The quotient of $\mathscr{M}(X,Y)$ by this equivalence relation is called the space of *rth-order jets* of $X \to Y$ denoted as

$$J^r(X,Y).$$

It is easily proved that the $J^r(X,Y)$, $r = 0,1,2,\ldots$, are C^∞, finite-dimensional paracompact manifolds. The source and target maps

$$\pi: J^r(X,Y) \to X,$$

$$\tau: J^r(X,Y) \to Y$$

are submersions (i.e., the induced linear mapping on tangent bundles is onto), and hence define the Ehresmann jet spaces as fiber spaces.

Consider the following relation:

$$(x,\phi,U) \sim (x',\phi',U')$$

$$\text{iff} \quad x = x', \ \phi \text{ and } \phi' \text{ meet to } \textit{infinite order} \text{ at } x, \text{ in the sense tha the Taylor expansions about } x \text{ of } \phi \text{ and } \phi' \text{ in a local coordinate system agree.}$$

The quotient of $\mathscr{M}(X,Y)$ by this equivalence relation is denoted by $J^\infty(X,Y)$, and called the space of *infinite-order jets*.

There are obvious inclusions among the equivalence classes of these equivalence relations. (For example, two maps ϕ and ϕ' which agree to rth order at x, agree to order s, where $s < r$.) They lead to natural maps between the various quotient spaces:

$$\mathscr{S}(X,Y) \to J^\infty(X,Y)$$

$$\cdots \to J^r(X,Y) \to J^{r-1}(X,Y) \to \cdots \to J^0(X,Y) = X \times Y. \tag{2.5}$$

Using the source and target maps, respectively, we ob-

tain fiber space maps

$$J^r(X,Y)\to X, \quad J^r(X,Y)\to Y. \tag{2.6}$$

Differential forms on X and Y can be lifted up to $J^r(X,Y)$ via the pullback of the maps (2.6). For simplicity, *we will often make no notational distinction between differential forms on X and Y and their pullback to $J^r(X,Y)$.*

If $\phi: U\to Y$ is a local map from X to Y, its *r-jet* or *prolongation* is the map

$$j^r(\phi): U\to J^r(X,Y) \tag{2.7}$$

obtained as follows:

For $x\in U$, construct the element (x,ϕ) of $\mathscr{M}(X,Y)$

and then assign it the equivalence class on $J^r(X,Y)$

to which it belongs. This is defined as $j^r(\phi)(x)$.

As x varies over U, the map (2.7) is obtained. (2.8)

There is a natural construction which leads from a pair of bases of differential forms on X and Y to basis of differential forms on $J^r(X,Y)$. We will now describe it after a brief recapitulation of the algebraic properties of vector fields on differential forms.

$\mathscr{F}(Z)$ denotes the ring of C^∞, real-valued functions on Z. For each integer $j\geqslant 0$, let $\mathscr{D}^j(Z)$ denote the C^∞ differential form of degree j. $\mathscr{D}^j(Z)$ is an $\mathscr{F}(Z)$-module. Exterior multiplication

$$\Lambda: \mathscr{D}^r(Z)\times\mathscr{D}^k(Z)\to\mathscr{D}^{j+k}(Z)$$

makes the collection

$$\mathscr{D}(Z)=\{\mathscr{D}^j(Z): j=0,1,2,\cdots\}$$

into a graded, associative algebra. The exterior differentiation operation

$$d: \mathscr{D}^j(Z)\to\mathscr{D}^{j+1}(Z)$$

is defined, with the usual properties.

A one-differential form θ on $J^r(X,Y)$ is a *contact form* if

$$(j^r\phi)^*(\theta)=0 \quad \text{for } \textit{each} \text{ local map } \phi: U\to Y. \tag{2.9}$$

Let $\mathscr{C}^r(X,Y)$ be the $\mathscr{F}(J^r(X,Y))$-module of contact forms.

Suppose

$$\dim X=n, \quad \dim Y=m. \tag{2.10}$$

Choose the following ranges of indices and summation convention:

$$1\leqslant i,j\leqslant n, \quad 1\leqslant a,b\leqslant m.$$

Suppose that (θ^a) is a basis for the one-forms $\mathscr{D}^1(Y)$ ("basis" as an $\mathscr{F}(Y)$-module) and that (ω^i) is a basis for $\mathscr{D}^1(X)$. Consider these as 1-forms on $J^r(X,Y)$ pulled back via the projection map. Then, there are functions

$$(y^a_i, y^a_{i_1 i_2},\ldots, y^a_{i_1\cdots i_r})$$

on $J^r(X,Y)$ such that the forms

$$\begin{aligned}
&\eta^a=\theta^a-y^a_i\omega^i,\\
&\eta^a_{i_1}=dy^a_{i_1}-y^a_{i_1 i_2}\omega^{i_2},\\
&\quad\vdots\\
&\eta^a_{i_1\cdots i_{r-1}}=dy_{i_1\cdots i_{r-1}}-y^a_{i_1\cdots i_r}\omega^{i_r}
\end{aligned} \tag{2.11}$$

are a basis for $\mathscr{C}^r(X,Y)$. Also, the forms

$$\omega^i,\theta^a,dy^a_i,\ldots,dy^a_{i_1\cdots i_r} \tag{2.12}$$

form a basis for 1-forms on $J^r(X,Y)$.

In this paper, we will assume that the "moving frames" (ω^i), (θ^a) are differentials of coordinate systems. We suppose that X and Y have a coordinate system, labeled

$$x^i, y^a,$$

with

$$\omega^i=dx^i, \quad \theta^a=dy^a. \tag{2.13}$$

One then proves readily that the functions

$$(x^i, y^a, y^a_i,\ldots,y^a_{i_1\cdots i_r}) \tag{2.14}$$

associated with these bases form a coordinate system for

$$J^r(X,Y).$$

A symbolic notation such as

$$(x,y,\partial_x y), \ (x,y,\partial_{xx}y,\cdots) \tag{2.15}$$

is often useful.

III. DIFFERENTIAL EQUATIONS, OPERATORS, AND SYMBOLS

Let X, Y, $J^r(X,Y)$, $r=0,1,2,\cdots$, be as before. Work with a fixed coordinate system (x^i), (y^a) for X, Y, and the associated coordinates $(x^i, y^a, y^a_i,\cdots)$ for $J^r(X,Y)$.

Definition: Let V be a real vector space, assumed finite-dimensional. An *rth-order differential operator symbol* with values in V is a (C^∞) map

$$\sigma: J^r(X,Y)\to V. \tag{3.1}$$

Given such a symbol, we will define a map

$$D_\sigma: \mathscr{M}(X,Y)\to\mathscr{M}(X,V), \tag{3.2}$$

which is the associated *differential operator*. For a map

$$\phi: U\to Y,$$

U an open subset of X, $x\in U$, D_σ is defined by

$$D_\sigma(\phi)(x)=\sigma(j^r(\phi)(x)). \tag{3.3}$$

The essence of this "geometric" approach to the theory of partial differential equations is the distinction between the "symbol," as a "*geometric object*," i.e., as a C^∞ mapping on a finite-dimensional manifold sitting above the domain and range manifolds X and Y ("independent" and "dependent" variables), and as a *mapping* between the *infinite*-dimensional space of local mappings. This correspondence is also implicit (and, in Lie's work, often quite explicit) in the 19th century literature on partial differential equations.

Definition: The *differential equation* associated with such a differential operator is the subset

$$\mathrm{DE}\equiv\sigma^{-1}(0) \tag{3.4}$$

of $J^r(X,Y)$. The *solutions* of the differential equations are the $\phi\in\mathscr{M}(X,Y)$ such that

$$D_\sigma(\phi)=0, \tag{3.5}$$

or, equivalently,

$$j^r(\phi)(U)\subset\mathrm{DE}(\sigma). \tag{3.6}$$

Remark: This geometric way of defining the *geometric concept* of "differential equation" can be defined in many different "categories," e.g., the category of C^∞ maps, or "algebraic" maps, or, even with a little ingenuity, over "characteristic p." For the purpose of differential geometry and physics, the most important categories seem, at the present time at least, to be the C^∞ and the real or complex analytic maps.

IV. PROLONGATIONS OF DIFFERENTIAL OPERATORS AND DIFFERENTIAL EQUATIONS

In the classical literature on nonlinear partial differential equation the process of *prolongation* is central, although not precisely defined. In Goldschmidt's and Spencer's work[9,14,15] one finds that it as an operator taking objects that live on a jet space of given order to one of next higher order. In this section I will describe a variant of their ideas.

With the notation explained in Sec. IV, let σ: $J^r(X,Y)\to V$ be a symbol map. A local mapping $\phi: X\to Y$, expressed in local coordinates $(x,y(x))$ for X, Y, is then a *solution* of the associated differential equation if

$$\sigma(x,y(x),\partial y(x)) = 0 \quad \text{for all } x\in U. \tag{4.1}$$

Now, for fixed i, $1\leqslant i\leqslant n$, differentiate (4.1) with respect to x^i. An $(r+1)$th-order differential equation is obtained for which ϕ is a solution. The symbol of these differential equations, which will be a map denoted as

$$\delta_i\sigma: \quad J^{r+1}(X,Y)\to V, \tag{4.2}$$

will be a map on $J^{r+1}(X,Y)$ taking values in the vector space V.

For example, if $r=1$, and if σ is a function of the coordinates

$$(x^i, y^a, y^a_i) \tag{4.3}$$

on $J^1(X,Y)$, then

$$\delta_i\sigma(x^i, y^a, y^a_i, y^a_{ij}) = \frac{\partial\sigma}{\partial x^i} + \frac{\partial\sigma}{\partial y^a}y^a_i + \frac{\partial\sigma}{\partial y^a_j}y^a_{ij}. \tag{4.4}$$

Remark: Notice in formula (4.4) that the "affine" structure of the jet spaces described and utilized by Goldschmidt[15] makes its appearance. As described partially in Ref. 16, it also plays a role in analytical mechanics.

We can now describe this in a more algebraic way by introducing the vector fields $\mathscr{V}(X)$ on X. Suppose

$$\mathscr{L}_V = A^i(x)\frac{\partial}{\partial x^i} \tag{4.5}$$

is such a vector field. Then, set

$$\mathscr{L}_V(\sigma) = A^i\delta_i(\sigma). \tag{4.6}$$

Let us sum up as follows:

Theorem 4.1: For each integer $r\geqslant 0$, let $\mathscr{F}(J^r(X,Y),V)$ be the space of C^∞ mapsL $J^r(X,Y)\to V$. [Algebraically, $\mathscr{F}(J^r(X,Y),V)$ is $\mathscr{F}(J^r(X,Y))\otimes V$.] Thus, the prolongation described above is an $\mathscr{R}$-linear map

$$\mathscr{L}_V: \ \mathscr{F}(J^r(X,Y),V)\to\mathscr{F}(J^{r+1}(X,Y),V). \tag{4.7}$$

$\mathscr{L}_V$, as the notation indicates, is a generalization of the usual Lie derivative. The following formula holds:

For each map $\gamma: X\to Y$

$$\mathscr{L}_V(j^r(\gamma)^*(\sigma)) = j^{r+1}(\gamma)^*(\mathscr{L}_V(\sigma)). \tag{4.8}$$

Let $t\to\exp(tV)$ be the one-parameter pseudogroup on X generated by the vector field V. Then, for $x\in X$,

$$\mathscr{L}_V(j^r(\gamma)^*(\sigma))(x) = \frac{\partial}{\partial t}(j^r(\gamma)^*(\sigma))[\exp(-tV)(x)]\,|_{t=0}. \tag{4.9}$$

V. PROLONGATION OF DIFFERENTIAL EQUATIONS

Continue with the notation of Sec. IV. Let V be a vector space, and

$$\sigma: \ J^r(X,Y)\to V$$

be an rth-order symbol map. The subspace

$$\sigma^{-1}(0)\equiv \mathrm{DE}(\sigma) \tag{5.1}$$

of $J^r(X,Y)$ defines a *differential equation*. A map $\phi: X\to Y$ is a *solution* if the following condition is satisfied.

$$j^r(\phi)(X)\subset \mathrm{DE}(\sigma). \tag{5.2}$$

An alternate way of putting this is that

$$j^r(\phi)^*(\sigma) = 0. \tag{5.3}$$

For each integer $s>0$, and $V_1,\ldots,V_s\in\mathscr{V}(X)$, consider the generalized Lie derivative

$$\mathscr{L}_{V_1}\cdots\mathscr{L}_{V_s}(\sigma). \tag{5.4}$$

It is a map $J^{r+s}(X,Y)\to W$. For each integer m, $m\geqslant s$, we can pull back the functions (5.4) to live on $J^{r+m}(X,Y)$, via the "forgetting" map: $J^{r+m}\to J^{r+s}$. In this way, we get a collection

$$\{\sigma, \mathscr{L}_{V_1}(\sigma),\ldots,\mathscr{L}_{V_1}\cdots\mathscr{L}_{V_m}(\sigma): \ V_1,\ldots,V_m\in\mathscr{V}(X)\} \tag{5.5}$$

of maps $J^{r+m}(X,Y)\to W$. Let $\mathrm{DE}^m(\sigma)$ be the subset of points of $J^{r+m}(X,Y)$ on which *all* the maps (5.5) *vanish*. Of course, as a differential equation, $\mathrm{DE}^m(\sigma)$ has precisely the same set of solutions as $\mathrm{DE}(\sigma)$. However, it is the geometric properties of each $\mathrm{DE}^m(\sigma)$ that are of greatest interest.

In particular, one is interested in knowing something of the algebrogeometric-topological properties of $\mathrm{DE}^m(\sigma)$ as subset of manifolds. Similar questions about subsets of jet spaces defined by equations occur in the theory of singularity of mappings.[17] In this discipline, work can be done effectively in the category of C^∞ mapping; however, at the partial differential equation level, in the current state of mathematical development, serious work will require the assumption that the manifolds X, Y and the symbol map σ be *real analytic*. When discussing the notion of *general* and *singular* solutions, later on in this paper, we shall make this assumption, since it will make available the powerful mathematical theorems to study these subsets that are not (yet) available in the C^∞ case. In fact, it is well known in the theory of partial differential equations that there are major qualitative differences between the C^∞ and real analytic category. If the coefficients of these equations were assumed analytic, the classic Cauchy–Kowalewsky theorem would guarantee the existence at least of local solutions.

VI. GENERAL SOLUTIONS OF SYSTEMS OF REAL-ANALYTIC PARTIAL DIFFERENTIAL EQUATIONS

Continue with the set up described in previous sections: $X, Y, J^r(X,Y), \sigma: J^r(X,Y) \to V$, $DE(\sigma) \subset J^r(X,Y)$, the mth prolongation $DE^m(\sigma) \subset J^{r+m}(X,Y)$, for $m \geqslant 1$. Now, assume that the data is *real analytic*. $DE^m(\sigma)$ is then a real analytic subvariety of the real analytic manifold $J^{r+m}(X,Y)$. It is then a union of a finite number of irreducible subvarieties.[18] We can then consider the analytic real-valued functions $f: U \to R$ defined on open subsets U of $J^{r+m}(X,Y)$, which are identically zero on at least one irreducible branch of $DE^m(\sigma)$. Denote these functions by $DE^m(\sigma, \mathscr{F})$. (They form a sheaf.) Let GS be a family of maps: $X \to Y$, such that each $\phi \in GS$ is a solution of $DE(\sigma)$,

$$j^r(\phi)^*(\sigma) = 0. \tag{6.1}$$

For each integer $m \geqslant 0$, each open subset $U \subset J^{r+m}(X,Y)$, we can consider the set of maps f: $U \to R$ such that

$$J^r(\phi)^*(f) = 0 \quad \text{for all } \phi \in GS. \tag{6.2}$$

Definition: GS is a *general solution* of the differential equations associated with the symbol map $\sigma: J^r(X,Y) \to V$ if each function f satisfying (6.2), for an integer $m \geqslant 0$, belongs to $DE^m(\sigma, \mathscr{F})$.

In words, this condition means that each differential equation satisfied by every map ϕ in the family GS is contained in the differential equation obtained by prolonging σ by differentiation.

This definition of "general solution" is one of two given and used in the 19th century literature[1,5]; it is due to Ampère. The other is given Darboux' name, but we will not go into it here.

These generalities are best comprehended by thinking about certain simple examples.

VII. THE GENERAL SOLUTION OF THE ONE-DIMENSIONAL WAVE EQUATION

Here,

$X = R^2$, with coordinates (x^1, x^2),

$Y = R$, with coordinates y.

$J^2(X,Y)$ has coordinates

$$z = (x, y, t_1, y_2, y_{11}, y_{12}, y_{22}).$$

Define $\sigma: J^2(X,Y) \to R$ as follows:

$$\sigma(z) = y_{12}. \tag{7.1}$$

The solutions of the differential equations are the maps

$$x \to y(z)$$

such that

$$\frac{\partial y}{\partial x_1 \partial x_2} = 0. \tag{7.2}$$

This is (in "light cone" or "characteristic" coordinates) just the wave equations, in one space, one time variable.

To define the prolongations of (7.1), let $1 \leqslant i, j, i_1, \ldots, \leqslant 2$ be indices

$$(x, y, y_i, y_{ij}, \cdots)$$

are coordinates of $J^\infty(X,Y)$. The prolonged varieties are the subsets of $J^{r+m}(X,Y)$ obtained by setting equal to zero the functions

$$y_{i \cdots i_s 12}, \tag{7.3}$$

where $1 \leqslant i_1, \ldots, i_s \leqslant 2$, s an arbitrary nonnegative integer.

The classical "general solution" of (7.3) is the family of maps $R^2 \to R$, of the form

$$\phi(x_1, x_2) = g_1(x_1) + g_2(x_2), \tag{7.4}$$

where g_1, g_2 are arbitrary maps: $R \to R$.

It is now obvious that any map $f: J^{r+m} \to R$ that satisfies (6.2) is a function *only* of the variables (7.3), which is the condition required to verify that the family of maps (7.4) is a general solution of the DE (7.2).

VIII. THE CLAIRAUT EQUATION: ITS "GENERAL" AND "SINGULAR" SOLUTION

The classical Clairaut equation is a first-order ordinary differential equation of the form

$$y(x) - x\frac{dy}{dx} - f\left(\frac{dy}{dx}\right) = 0 \tag{8.1}$$

to be solved for a map $x \to y(x)$ of $R \to R$. f is givn C^∞ map: $R \to R$.

To put this into the standard framework described above, let

$$X = Y = R. \tag{8.2}$$

Denoting the coordinates of $J^1(X,Y)$ by (x,y,y'), the symbol σ is the map: $J^1(X,Y) \to R$ defined as follows:

$$\sigma(x,y,y') = y - xy' - f(y'). \tag{8.3}$$

In order to define the prolonged systems, let us differentiate (8.1):

$$\begin{aligned} 0 &= \frac{dy}{dx} - \frac{dy}{dx} - x\frac{d^2y}{dx} - f'\left(\frac{dy}{dx}\right)\frac{d^2y}{dx^2} \\ &= -\frac{d^2y}{dx^2}\left(x + f'\left(\frac{dy}{dx}\right)\right). \end{aligned} \tag{8.4}$$

Thus,

$$\delta\sigma((x,y,y',y'')) = y''(x + f'(y')). \tag{8.5}$$

The variety in $J^2(X,Y)$ obtained by setting

$$0 = \sigma = \delta$$

is *reducible*.

At the differential equation level, C^∞ solutions of (8.4) are of two sorts: Solutions of

$$\frac{d^2y}{dx^2} = 0 \tag{8.6}$$

or

$$f'\left(\frac{dy}{dx}\right) = -x. \tag{8.7}$$

The "general solution" of (8.6) is

$$y = ax + b, \tag{8.8}$$

where a and b are real constants. The condition that it be a

solution of (8.1) introduces a relation between a and b:

$$ax + b - xa - f(a) = 0,$$

or

$$f(a) = b. \tag{8.9}$$

Theorem 8.1: The family

$$\phi_a(x) = ax + f(a), \tag{8.10}$$

parametrized by $a \in R$, is a general solution, in the Ampère sense, of the differential equation (8.1).

Proof: Suppose

$$\alpha(x, y, y', \ldots, y^{(r)})$$

in an rth-order differential operator symbol, for which $y(x) = \phi_a(x)$ is a solution for each a. Now,

$$\frac{d\phi_a}{dx} = a,$$

$$\frac{d^2\phi_a}{dx^2} = 0,$$

$$\vdots$$

Hence,

$$\alpha(x, ax + f(a), a, \ldots, 0) = 0, \tag{8.11}$$

and then

$$\alpha(x, y, z, \ldots, 0) = 0 \quad \text{for all } x, y. \tag{8.12}$$

We conclude that α must be for the form:

$$\alpha = \beta_1 y'' + \cdots \beta_{r-2} y^{(r)},$$

where $\beta_1, \ldots, \beta_{r-2}$ are arbitrary functions on $J^r(X, Y)$. This is what is required to prove the ϕ_a are "general solution."

The Clairaut equation is the example traditionally used in the classical treatises to introduce the notion of "singular solution." They are the solutions of the original differential equation (8.1), which are not part of the general solution (8.10). They must then be solutions of (8.7). The general solution of (8.7) depends on one real parameter, say

$$x \to y^s(x; c) \tag{8.13}$$

with

$$f'(y^s_x(x; c)) = -x. \tag{8.14}$$

Now, the condition that the family of functions of x, indicated in (8.13), *also* satisfy the original differential equation, is

$$y^s(x, c) - x y^s_x - f(y^s_x) = 0. \tag{8.15}$$

In general, (8.14) and (8.15) will have solutions for only a discrete set of c's. These are the *singular solutions* of the Clairaut differential equation (8.1).

Now, the traditional geometric interpretation of these singular solutions of the Clairaut equation is that they are *the envelopes* of the family of curves in R^2 given by the general solution

$$x \to (x, y(x; a)) = (x, ax + f(a)). \tag{8.16}$$

For each a, (8.16) is a straight line. They are the tangent lines to the curves $x \to (x, y^s(x))$, where y^s are the solutions of (8.14) and (8.15).

IX. COMPLETE SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS IN THE SENSE OF LAGRANGE AND VESSIOT. LAGRANGE-VESSIOT FOLIATIONS AND FIBRATIONS

In the classical literature, the concept of a "complete solution" of a partial differential equation is traditionally defined in the context of nonlinear first-order partial differential equations for one unknown function. The most important example for physical purposes (and the typical one from the mathematical point of view) is the Hamilton–Jacobi equation:

$$H\left(q, \frac{\partial S}{\partial q}\right) = 0, \tag{9.1}$$

where H is a function of n "position" variables $(q^1, \ldots, q^n)$, and "momentum" variables $(p_1, \ldots, p_n)$. [In modern language, H is a real-valued function on the contangent bundle $T^*(Q)$ to a manifold Q.] A *complete solution* of (9.1) is then a set of functions

$$q \to S(q; a),$$

depending on additional parameters $a = (a_1, \ldots, a_n)$ such that

$$H\left(q, \frac{\partial S}{\partial q}(q; a)\right) = 0, \tag{9.2}$$

$$\det\left(\frac{\partial S}{\partial a^i \partial a^j}\right) \neq 0. \tag{9.3}$$

Let M be the space of variables (q, p). Let

$$\omega = dp_1 \wedge dq^1 + \cdots + dp_n \wedge dq^n. \tag{9.4}$$

Let $\mathscr{E}$ be the exterior differential system (i.e, the ideal in the Grassmann algebra $\mathscr{D}(M)$ of differential forms on M, which is closed under exterior differentiation d) generated by the 2-form ω and the 0-form H. Thus, for each $a \in R^n$, the map

$$\phi_a : q \to q, p\frac{\partial S}{\partial q}$$

is a submanifold map of $Q \to M$. It defines an n-dimensional integral submanifold of $\mathscr{E}$, i.e.,

$$\phi_a^*(\mathscr{E}) = 0. \tag{9.5}$$

As a varies, we obtain a foliation of M by n-dimensional submanifolds [that ϕ_a is a submanifold map is guaranteed by condition (9.3)]. In the contemporary literature on symplectic manifolds, this is often called a *Lagrangian foliation* (with historical justification, since it was Lagrange who introduced the concept of "complete solution" for first-order nonlinear partial differential equations).

To introduce the notion of "singular solution of a first-order partial differential equation" (as developed by Goursat[1] and Forsythe[5]), suppose that (9.3) is only satisfied in certain regions, but not identically. The solutions of (9.2) of the form

$$q \to S(q; a(q)) \tag{9.6}$$

such that

$$\det\left(\frac{\partial^2 S}{\partial a^i \partial a^j}(q; a(q))\right) = 0 \tag{9.7}$$

for all q, are called *singular solutions*. In this way of regard-

ing them, they are dependent on the preliminary choice of "complete solution."

E. Vessiot, in 1924,[19] proposed a generalization of this classical concept of "complete solution." We will now rephrase Vessiot's idea in terms of the Ehresmann jet spaces.

Definition: Let $\sigma: J^r(X,Y) \to V$ be the symbol of an rth-order differential operator with domain manifold X and range manifold Y. Let $\mathrm{DE}(\sigma) = \sigma^{-1}(0)$ be the subset of the jet space, which defines the "differential equation." Let γ: $J^r(X,Y) \to Z$ be a submersion mapping of $J^r(X,Y)$ onto a manifold Z. ("Submersion means that the induced linear mapping on tangent bundles is *onto*.) For each $z \in Z$, the fiber $\gamma^{-1}(z)$ is then a submanifold of $J^r(X,Y)$. γ is said to define a *Lagrange–Vessiot fibration* of $\mathrm{DE}(\sigma)$ if the following conditions are satisfied:

(a) $\dim \gamma^{-1}(z) = \dim X$ for all $z \in Z$, (9.8)

(b) $\gamma^{-1}(z) \subset \mathrm{DE}(\sigma)$ for all $z \in Z$, (9.9)

(c) the projection map $\gamma^{-1}(z) \to X$ is a local diffeomorphism, (9.10)

(d) the contact forms on $J^r(X,Z)$ vanish when restricted to $\gamma^{-1}(z)$. (9.11)

In his fundamental paper[19] Vessiot gives sufficient conditions for the local existence of such complete solutions that we will not go into it at this point. [The conditions are that the exterior differential system generated by the contact form, restricted to $\mathrm{DE}(\sigma)$, assumed to be a submanifold, are in *involution* in Cartan's sense.[10,20,21]]

In this paper, I will assume that Lagrange–Vessiot fibrations exist, and examine the consequences. I will be particularly concerned with conditions that *one* such fibration determines the "general solution" of $\mathrm{DE}(\sigma)$. This is what happens in the case of a nonlinear first-order partial differential equation: Such a description is called the Method of Lagrange and Charpit[1,5] in the classical literature.

X. GENERAL SOLUTIONS THAT ARE GENERATED FROM ONE PARTICULAR COMPLETE SOLUTION. LAGRANGE–VESSIOT FIBRATION BY THE METHOD OF LAGRANGE AND CHARPIT

Continue with the notation of Sec. IX. Let γ: $J^r(X,Y) \to Z$ be a submersion map that defines a Lagrange–Vessiot fibration [i.e., satisfies conditions (9.8)–(9.11)] for the partial differential equation associated with the symbol map $\sigma: J^r(X,Y) \to Z$.

Let U be an open subset of X, and let $\phi: U \to Y$ be a map which is a solution of the partial differential equation generated by σ, i.e.,

$$j^r(\phi)(U) \subset \mathrm{DE}(\sigma). \tag{10.1}$$

We can then construct the map

$$\gamma j^r(\phi): U \to Z.$$

Thus, to the solution ϕ we assign the subset

$$\gamma j^r(\phi)(U)$$

of Z. For certain ϕ's, this will be a *submanifold* of Z. The classical method of Lagrange and Charpit attempts to reverse this process, i.e., assigning a solution ϕ to the partial differential equation to certain submanifolds of Z. In certain favorable cases, one can generate a general solution to the partial differential equation in this way.

In order to understand better when the situation is "favorable" to the Lagrange–Charpit method, it is most convenient to introduce more explicitly the geometric ideas of Cartan's theory of exterior differential systems.

XI. THE METHOD OF LAGRANGE–CHARPIT IN THE CONTEXT OF CARTAN'S THEORY OF EXTERIOR DIFFERENTIAL SYSTEMS

Let M be a manifold with a set $\mathscr{E}$ of differential forms on M which define an *exterior differential system*, i.e., $\mathscr{E}$ is an ideal of algebraic sense for the Grassmann algebra structure on differential forms, plus $\mathscr{E}$ is closed under the exterior derivative operation d. For $p \in M$, let M_p be the tangent space to M at p; let M_p^d be its dual space. Let $E(M_p^d)$ be the exterior algebra of the vector space M_p^d. Let $\mathscr{E}(p)$ be the set of values at p of the forms on $\mathscr{E}$. $\mathscr{E}(p)$ is then an ideal in $E(M_p^d)$.

Recall that a submanifold map $\phi: P \to M$ is said to be an *integral submanifold* of $\mathscr{E}$ if

$$\phi^*(\mathscr{E}) = 0. \tag{11.1}$$

A linear subspace α of the tangent space M_p is said to be an *integral element* of $\mathscr{E}$ if

$$\theta(\alpha) = 0 \quad \text{for all } \theta \in \mathscr{E}. \tag{11.2}$$

Thus, the condition (11.1) that the map $\phi: P \to M$ be an integral submanifold is that $\phi_*(P_p)$ be an integral element, for each $p \in P$.

α is said to be a *maximal integral element* if it is contained in no larger integral element.

Definition: A tangent vector $v \in M_p$ is *Cauchy characteristic* for the ideal $\mathscr{E}$ if the following condition is satisfied:

$$v \perp \mathscr{E} \subset \mathscr{E}(p). \tag{11.3}$$

$\mathrm{CC}(\mathscr{E})$ denotes the set of such Cauchy characteristic tangent vectors. It defines a linear subbundle

$$p \to \mathrm{CC}(\mathscr{E})(p)$$

of the tangent vector bundle $T(M)$ (but one which may have varying dimensions of the fibers, i.e., some sort of "singular" vector bundle). A vector field $V \in \mathscr{V}(M)$ is said to be *Cauchy characteristic* if the values $V(p)$, for each p, are Cauchy characteristic.

Theorem 11.1: If $\alpha \subset M_p$ is a maximal integral element of the exterior system $\mathscr{E}$ at p, then

$$\mathrm{CC}(\mathscr{E})(p) \subset \alpha. \tag{11.4}$$

Proof: Let $v \in M_p$ be Cauchy characteristic for $\mathscr{E}$, i.e., (11.3) is satisfied. Then, for $v_1, \ldots, v_m \in \alpha$, $\theta \in \mathscr{E}^{m+1}$,

$$\begin{aligned} \theta(v, v_1, \ldots, v_m) &= (v \perp \theta)(v_1, \ldots, v_m) \\ &= 0, \end{aligned}$$

using (11.4) and the hypothesis that α is an integral element. This shows that

$$\alpha + (v)$$

is also an integral element; by maximality,

$\alpha + (v) \subset \alpha$

whence $v \in \alpha$. Q.E.D.

Definition: A submersion map $\gamma: M \to A$, with A a manifold, is said to be a *Lagrange–Vessiot submersion* (relative to the exterior system $\mathscr{E}$ on M) if the following conditions are satisfied:

(a) For each $a \in A$, the fiber $\gamma^{-1}(a)$ is an integral submanifold of $\mathscr{E}$.

(b) For each $p \in M$, the tangent subspace $\gamma_*^{-1}(0)$ is a *maximal* integral element of $\mathscr{E}$.

Theorem 11.2: Suppose $\gamma: M \to A$ is a Lagrange–Vessiot submersion, in the sense of the above definition. Then, the Cauchy characteristic tangent vectors and vector fields are tangent to the fibers of γ.

Proof: Follows from Theorem 11.1.

Now, let $\gamma: M \to A$ be a fixed Lagrange–Vessiot submersion and let $\phi: P \to M$ be an integral submanifold of $\mathscr{E}$ such that

$$\dim P = \text{dim fibers of } \gamma; \quad \text{for each } p \in P, \text{ the integral element } \phi_*(P_p) \text{ is maximal.} \tag{11.5}$$

Consider the composition map

$$\gamma\phi: P \to A. \tag{11.6}$$

Theorem 11.3: For each $p \in P$, the Cauchy characteristic vectors $CC(\mathscr{E})(\phi(p))$ are tangent to the submanifold $\phi(P)$ of M, and are mapped onto zero under γ_*.

Proof: This follows from Theorem 11.1, i.e., that the Cauchy characteristic vectors are contained in *every* maximal integral element.

Definition: The integral submanifold $\phi: P \to M$ is said to be of *Lagrange–Charpit* type relative to the Vessiot submersion $\gamma: M \to M$ if, for each $p \in P$, $\phi_*(P_p)$ is a maximal integral element of $\mathscr{E}$, and if the dimension of the kernel $(\gamma\phi)_*^{-1}(0)$ of the map $(\gamma\phi): P \to A$ is equal to the dimension of the Cauchy characteristic vectors of $\mathscr{E}$.

Theorem 11.4: If the dimension of the Cauchy characteristic vectors of $\mathscr{E}$ is constant over M, and if $\phi: P \to M$ is an integral submanifold of Lagrange–Charpit type, then $\gamma\phi$ is a submersion mapping from P to a submanifold of A. The dimension of this manifold of A is equal to the dimension of P minus the dimension of the Cauchy characteristic vectors.

Proof: Again, this follows from our hypotheses and the implicit function theorem, since the map $\gamma\phi: P \to A$ is of constant rank.

Thus, we have presented the Lagrange–Charpit method in modern language. In favorable cases, it can be *inverted*, assigning to each submanifold of certain dimension on A an integral manifold of $\mathscr{E}$. In certain cases described in the classical literature, *this will* give a "general solution" in the Ampère sense described above. However, there are many partial differential equations, for which the construction of the "general solution" is more complicated geometrically.

XII. REMARKS ON ANALYTICITY OF THE GENERAL SOLUTION AND THE RELATION TO SOLUBILITY INTEGRABILITY

Both for theoretical and practical reasons (since they were concerned with physics and eingineering problems that required extensive hand computation), a main topic in the 19th century theory of differential equations was the feasibility of finding solutions of differential equations by algorithms. In the work of Kowalewsky and Painlèvé on the analytical structure of ordinary nonlinear differential equations, one finds what is perhaps the most geometrically sophisticated and interesting work in this direction. I will now present a few remarks which attempt to put these ideas in a more contemporary setting, and suggest further possibilities of research.

The jet-space formalism of Ehresmann and Spencer, as described in previous sections, can serve to formulate in a coordinate-free way what is meant by a system of *complex-analytic* differential equations. Let X, Y be complex analytic manifolds, and $J_c^r(X,Y)$ be the manifold of complex-analytic jets of mapping: $X \to Y$. Now, we deal with complex-analytic symbol maps

$$\sigma: J_c^r(X,Y) \to V,$$

which determine, classically, differential equations with complex analytic coefficients.

The local existence theories of Cauchy and Kowalewsky, Riquier, Cartan, and Köhler, give general sufficient conditions for the existence of *local* solutions $\phi: U \to Y$, which are complex analytic. One can attempt to analytically continue these solutions. However, only in the case complex $\dim X = 1$ was there enough technique available to say anything nontrivial about such possibiliies. (Of course, quite a bit was known or speculated about analytic continuation, possible existence of algebraic solutions, etc., of special solutions of special equations.) In this case, which of course involves ordinary differential equations, the key applied input came from celestial and analytical mechanics, where exact or appropriate solution by power series was the preferred practical technique. Thus, a natural question was the following one:

In the case where X is a one-dimensional complex manifold, is there an open, dense subset U and an analytic coordinate x in U such that a *general solution* of the differential equation could be exhibited as quotients of functions defined in U depending on the required number of parameters, which could be written as power series in x which were convergent in U and a fixed domain of the auxillary parameter?

(12.1)

There was, of course, other historical motivation for posing this sort of question. Many of the "soluble" mechanics problems could be solved in terms of *elliptic functions*: Jacobi had introduced the *theta functions* precisely in order to exhibit these functions as quotients of functions which were described by convergent power series.

XIII. THE DIFFERENTIAL ALGEBRA OF DIFFERENTIAL FORMS ON THE INFINITE ORDER JET SPACE

Now return to the usual situation where X and Y are real, C^∞, finite-dimensional manifolds. We have defined the space $J^\infty(X,Y)$ of *infinite order jets* of mappings; $X \to Y$. We have the tower of maps

$$J^\infty(X,Y) \to J^{r+1}(X,Y) \to J^r(X,Y) \to \cdots. \tag{13.1}$$

$J^\infty(X,Y)$ is not a manifold in the usual sense; however, the tower (13.1) can be used to attach a *differential form algebra* to it.

For each integer $r \geqslant 0$, let

$$\mathscr{D}(J^r(X,Y)) = \{\mathscr{D}^n(J^r(X,Y)):\ n = 0,1,\cdots;\wedge\} \tag{13.2}$$

be the graded associative differential algebra of differential forms on the manifold $J^r(X,Y)$. The dual of maps on forms to the tower (13.1) leads to a set of *inclusion* relations

$$\mathscr{D}(J^0(X,Y)) \subset \mathscr{D}(J^1(X,Y)) \subset \cdots \tag{13.3}$$

among the differential forms on the tower of jet spaces. We can then define a differential form algebra as the *union* of this sequence of sets. We denote it as $\mathscr{D}(J^\infty(X,Y))$, and call it the *differential form algebra* of the infinite order jet space. Thus, a differential form of degree n on the $J^\infty(X,Y)$ is, by definition, an element of $\mathscr{D}^n(J^r(X,Y))$ for r sufficiently large. The operations $d, \wedge$ on these forms are just those inherited from the corresponding operations on the finite-order jet manifolds, where they are defined by the usual calculus-on-manifold apparatus.

However, $\mathscr{D}(J^\infty(X,Y))$ has an additional algebraic structure that it inherits from the *prolongation* operation on functions defined in Sec. IV. Thus, to every vector field V on X we have assigned an operation

$$\delta(V):\ \mathscr{F}(J^r(X,Y)) \to \mathscr{F}(J^{r+1}(X,Y)). \tag{13.4}$$

$\delta(V)$ is a *derivation* with respect to the algebraic structure on $\mathscr{F}(J^r(X,Y))$:

$$\delta(V)(f_1 f_2) = \delta(V)(f_1) f_2 + f_1 \delta(V)(f_2) \quad \text{for} \quad f_1, f_2 \in \mathscr{F}(J^1(X,Y)). \tag{13.5}$$

There is a convention inherent in the right-hand side of (13.5). $\delta(V)(f_1)$ in an element of $\mathscr{F}(J^{r+1}(X,Y))$, whereas f_2 is an element of $\mathscr{F}(J^r(X,Y)$(. In order to multiply them, $\mathscr{F}(J^r(X,Y))$ must be identified with a subset [as it is via the dual of the projection map

$$J^{r+1}(X,Y) \to J^r(X,Y)$$

of $\mathscr{F}(J^{r+1}(X,Y))$].

Now, $\mathscr{F}(J^{r+1}(X,Y))$ is $\mathscr{D}^0(J^r(X,Y))$, and $\mathscr{D}^0(J^\infty(X,Y))$ is, by definition, the union of all the $\mathscr{D}^0(J^r(X,Y))$, for $r = 0,1,\cdots$. $\delta(V)(f)$ is the same whether it comes from J^r, $J^{r+1},\cdots$ (of course, this is inherent in the geometric meaning of $\delta(V)$!). Hence, $\delta(V)$ can be defined in a uniform way as a *derivation* of the algebra

$$\mathscr{F}(J^\infty(X,Y)) \equiv \mathscr{D}^0(J^\infty(X,Y)),$$

such that

$$\delta(V)(\mathscr{D}^0(J^r(X,Y))) \subset \mathscr{D}^0(J^{r+1}(X,Y)) \quad \text{for} \quad r = 0,1,\cdots. \tag{13.6}$$

We can now extend $\delta(V)$ to act on higher degree differential forms, in a way which commutes with exterior derivation and acts as a derivation on exterior multiplication. Suppose, for example, that (x) denotes coordinates on X, (y) coordinates on Y,

$$(x, y, \partial_x y, \partial_{xx} y, \cdots)$$

denote coordinates on $J^\infty(X,Y)$. A differential form on $J^\infty(X,Y)$ can be written as follows:

$$\omega = a df_1 \wedge \cdots\ df_m + \cdots,$$

with $a, f_1, \ldots, f_m \in \mathscr{D}^0(J^\infty(X,Y))$. Thus,

$$\delta(V)(\omega) = \delta(V)(a) df_1 \wedge \cdots \wedge df_m + \mathrm{ad}[\delta(V)(f_1)] \wedge \cdots \wedge df_m + \cdots. \tag{13.7}$$

We have

$$d\delta(V) = \delta(V) d \quad \text{for} \quad V \in \mathscr{V}(X). \tag{13.8}$$

Thus, we can translate "geometry" to "algebra" by assigning to (X,Y) the associative differential form algebra

$$\mathscr{D}(J^\infty(X,Y))$$

with its graded, differential structure. Many of the jewels which lie hidden in the 19th century theory of differential equations and the calculus of variations will be found when this algebraic structure is completely analyzed and understood.

Remark: A special case of this construction was given in Chap. 10 of Interdisciplinary Mathematics, Vol. XII.[13] Here, I have only touched on an area that is extensively developed in the work of Kupershmidt and Vinogradov.[22,23]

XIV. THE CLASSICAL TECHNIQUES OF THE "INTERMEDIATE INTEGRAL" FOR MONGE–AMPÈRE SECOND-ORDER PARTIAL DIFFERENTIAL EQUATIONS

In the 19th century, there was extensive work on finding conditions that given second-order partial differential equations could exhibit families of solutions that were also solutions of a first order partial differential equation. This was known as the *method of the intermediate integral*, and is discussed by Goursat[1] and Forsythe.[5] In this section, I will investigate in modern terms certain aspects of this method, elaborating on a brief discussion in Ref. 10.

Let X and Y be manifolds.

Definition: Let $\sigma: J^2(X,Y) \to V$ be the symbol of a second-order differential operator. A first-order, scalar-valued symbol $f: J^1(X,Y) \to R$ is said to be an *intermediate integral* of the system defined by σ if every solution $\phi: X \to Y$ of the differential equation

$$j^1(\phi)^*(f) = 0 \tag{14.1}$$

also satisfies

$$j^2(\phi)^*(\sigma) = 0. \tag{14.2}$$

There is an obvious practical interest in finding such "intermediate integrals." For example, physical systems often lead to second order partial differential equations. Solutions of (14.1) can be found using ordinary differential equations. Thus, the method, if it works, will generate familes of

solutions of the physical equations by means of ordinary differential equations.

An example of such a situation is provided by factoring of second order operators. Suppose $Y = R$ and that

$$\Delta\colon \mathscr{F}(Y) \to \mathscr{F}(Y)$$

is a second-order differential operator such that

$$\Delta(f) = \sigma j^2(f). \tag{14.3}$$

Suppose that

$$\Delta = \Delta_2 \Delta_1, \tag{14.4}$$

where Δ_2, Δ_1 are finite-order operators. Let

$$f\colon J^1(X,R) \to R$$

be the symbol of Δ_1. Then, f is an "intermediate integral" of σ.

In this section, we shall pursue the "Monge–Ampère" method[2,5,10] of constructing second-order equations which admit such "intermediate integrals."

Let $\mathscr{C} \subset \mathscr{D}(J^1(X,Y))$ be the Grassmann algebra ideal generated by the contact forms on $J^1(X,Y)$. Suppose $\dim Y = m$, $\dim X = n$.

Definition: The second-order differential equation defined by the symbol map $\sigma\colon J^2(X,Y) \to V$ is of *Monge–Ampère type* if there is a 2-form ω on $J^1(X,Y)$ such that the following conditions are satisfied:

$$\omega^n \wedge \theta^1 \wedge \cdots \wedge \theta^m = 0 \quad \text{for all} \quad \theta^1,\ldots,\theta^m \in \mathscr{C}. \tag{14.5}$$

($\omega^n = \omega \wedge \cdots \wedge \omega$ denotes the exterior product of n copies of ω).

Each map $\phi\colon X \to Y$ such that

$$j^1(\phi^*)(\omega) = 0 \tag{14.6}$$

also satisfies

$$j^2(\phi)^*(\sigma) = 0.$$

In this paper, we will deal only with the case

$$\dim X = n = 2, \quad \dim Y = m = 1, \tag{14.7}$$

i.e., we consider *partial differential equations in two independent and one dependent variables.*

Let $\mathscr{I}$ be the Grassmann algebra ideal of forms on $J^1(X,Y)$ generated by $\mathscr{C}$ and ω. For $x \in J^1(X,Y)$, let $E(\mathscr{I})(z)$ be the characteristic tangent vectors of this Grassmann ideal $\mathscr{I}$, i.e.,

$$E(\mathscr{I})(z) = \{v \in J^1(X,Y)_z\colon v \perp \mathscr{I} \subset \mathscr{I}\}. \tag{14.8}$$

Let us make the following regularity assumption:

$$\dim E(\mathscr{I})(z) \text{ is constant as } z \text{ ranges over } J^1(X,Y). \tag{14.9}$$

(Note that $\mathscr{I}$ is not closed under d, so that it is not an "exterior differential system." Of course, it can be closed up to generate such a system.)

The dimension of $J^1(X,Y)$ is then five. $\mathscr{C}$ is then generated by a single 1-form. Let $E(\mathscr{C}) \subset T(J^1(X,y))$ be the submodule of the tangent bundle to $J^1(K,Y)$ defined by the annihilation of the forms in $\mathscr{C}$. (Since $\mathscr{C}$ is generated by 1-forms $E(\mathscr{C})$ can be defined as the Cauchy characteristic vector of $\mathscr{C}$, which again is a Grassmann ideal, but not closed under d.) The fibers of $E(\mathscr{C})$ are then *four*-dimensional.

Condition (13.5) then means that ω, restricted to $E(\mathscr{C})$, is a skew-symmetry form of rank 2. Thus it can be written locally as follows:

$$\omega = \theta_1 \wedge \theta_2 + \theta_3 \wedge \theta, \tag{14.10}$$

where

θ is a 1-form generating $\mathscr{C}$.

The vectors in $E(\mathscr{I})$ are then defined by the following Pfaffian equations:

$$\theta_1 = \theta_2 = \theta = 0. \tag{14.11}$$

To avoid degenerate cases, let us assume that

$$\omega \notin \mathscr{C}. \tag{14.12}$$

Then forms (13.11) are linearly independent, i.e.,

$$\dim E(\mathscr{I}) = 2. \tag{14.13}$$

Definition: Let $\phi\colon X \to Y$ be a map which is a solution of the differential equation

$$j^1(\phi)^*(\omega) = 0. \tag{14.14}$$

The curves $t \to x(t)$ on X such that

$$j^1(\phi)^*(\theta_1)\left(\frac{dx}{dt}\right) = 0 \tag{14.15}$$

or

$$j^1(\phi)^*(\theta_2)\left(\frac{dx}{dt}\right) = 0 \tag{14.16}$$

are the *characteristics*.

Let (x^i, y, y_i), $i = 1,2$, be the natural jet coordinates on $J^1(X,Z)$ with respect to coordinates (x^i) for X and a coordinate y for Y. Then

$$\theta = dy_i - y_i\, dx^i. \tag{14.17}$$

Then, mod θ, θ_1, and θ_2 can be written as

$$\theta_1 = A^i\, dy + A_i\, dx^i, \tag{14.18}$$

$$\theta_2 = B^i\, dy_i + B_i\, dx^i, \tag{14.19}$$

where the A's and B's are functions of (s^i, y, y_i). Thus,

$$\begin{aligned}\theta_1 \wedge \theta_2 &= (A^1 B^2 - A^2 B^1)\, dy_1 \wedge y_2 \\ &\quad + (A^i B_j - A^j B_i)\, dy_i \wedge dx^j \\ &\quad + (A_1 B_2 - A_2 B_1) dx^1 \wedge dx^2 \\ &\equiv \alpha\, dy_1 \wedge dy_2 + \alpha_j^i\, dy_i \wedge dx^j + \beta\, dx^1 \wedge dx^2,\end{aligned} \tag{14.20}$$

$$j^1(\phi)^*(y) = y(x^1, x^2),$$

$$j^1(\phi)^*(y_i) = \frac{\partial y}{\partial x^i}.$$

The differential equation (14.14), in these coordinates, takes on the Monge–Ampère form:

$$\alpha\left(x, y, \frac{\partial y}{\partial x}\right)\left(\frac{\partial^2 y}{\partial x^1 \partial x^1}\frac{\partial^2 y}{\partial x^2 \partial x^2} - \left(\frac{\partial^2 y}{\partial x^1 \partial x^2}\right)^2\right)$$

$$+ F\left(x, y, \frac{\partial^2 y}{\partial x^2}\right)\frac{\partial y}{\partial x} = 0. \tag{14.21}$$

Theorem 14.1: Let $f\colon J^1(X,Y) \to R$ be a function such that

$$df(E(\mathscr{I})) = 0, \quad df \notin \mathscr{C}. \tag{14.22}$$

Then, any $\phi: x \to Y$ satisfying the first-order differential equation

$$j^1(\phi)^*(f) = \text{const} \tag{14.23}$$

also satisfies the Monge–Ampère equation (14.21), i.e., f is an "intermediate integral" of (14.21).

Proof: (14.23) means that df is a linear combination of θ, θ_1, θ_2, which implies that ω can be written as follows:

$$\omega = df \wedge \theta_1' + (\text{forms i } \mathscr{C}); \tag{14.24}$$

hence

$$j^1(\phi)^*(\omega) = 0,$$

whence, as we have seen, (14.21).

XV. THE FAMILIES OF SOLUTIONS OF THE MONGE–AMPÈRE EQUATION GENERATED BY ALL INTERMEDIATE INTEGRALS

I now want to make more precise some remarks on pp. 250–1 of Ref. 10 concerning when the intermediate integrals generate general solutions of Monge–Ampère equations. Again, let X and Y be manifolds with

$$\dim X = 2, \quad \dim Y = 1.$$

Let

$$Z = J^1(X,Y). \tag{15.1}$$

Let $\mathscr{P} \subset \mathscr{D}^1(Z)$ be a nonsingular Pfaffian system on Z, i.e., $\mathscr{P}$ is an $\mathscr{F}(Z)$-module of 1-differential forms on M such that the following condition is satisfied:

$$\dim(\mathscr{P}(z)) = \text{const}, \tag{15.2}$$

where z runs over all points of z. [$\mathscr{P}(z)$ denotes the set of values at z of all forms in $\mathscr{P}$. $\mathscr{P}(z)$ is a linear subspace of Z_z^d, the space of one-covectors to Z at z.] Set:

$$\mathscr{P}^1 = \{\theta \in \mathscr{P} : d\theta \in \mathscr{D}^1(Z) \wedge \mathscr{P}\} \tag{15.3}$$

$\mathscr{P}^1$ is another Pfaffian system, called the *first derived system to* $\mathscr{P}$. We can iterate the construction, obtaining a decreasing sequence

$$\mathscr{P} \supset \mathscr{P}^1 \supset \cdots \tag{15.4}$$

of Pfaffian systems. We will suppose these are *nonsingular* in the sense that their values at points z are constant as z ranges over Z. Thus, there is a first integer r such that

$$\mathscr{P}^r = \mathscr{P}^{r+1} = \mathscr{P}^{r+2} = \cdots. \tag{15.5}$$

Now, suppose that ω is a 2-form on Z such that

$$\omega \in \mathscr{P} \wedge \mathscr{P}, \tag{15.6}$$

$$\omega \wedge \omega = 0. \tag{15.7}$$

(14.5) and (14.6) guarantees that, locally, ω can be written as

$$\omega = \theta_1 \wedge \theta_2 \quad \text{with } \theta_1, \theta_1 \in \mathscr{P}. \tag{15.8}$$

We can now define a partial differential equation of Monge-Ampère type for mappings (defined locally) $\phi: X \to Y$ such that

$$j^1(\phi)^*(\omega) = 0, \tag{15.9}$$

i.e., the maps ϕ satisfying (15.9) are the *solutions*. Linear algebra, plus (15.7), implies the following result:

Theorem 15.1: Suppose ϕ is a map: $X \to Y$ satisfying (15.7). Then,

$$\dim(j^1(\phi)^*(\mathscr{P}))(x) = 0 \quad \text{or} \quad 1 \quad \text{for } x \in X. \tag{15.10}$$

Let us consider points $x \in X$ where the dimension in (15.8) is one. At such points (which form an open subset of X, whose complement is a union of varieties of dimension 0 or 1, if the data is real analytic), the integral curves of $j^1(\phi)^*(\mathscr{P})$ are the *characteristics*. (In the Monge–Ampère case, the two-parameter family of characteristics for general second-order partial differential equations "degenerate" to a one-parameter family.)

Definition: A function $f: Z \to R$ such that

$$df \in \mathscr{P}, \tag{15.11}$$

is an *intermediate integral* for the Monge–Ampère equation (14.8).

Theorem 15.2: If $f \in \mathscr{F}(Z)$ is an intermediate integral for the Monge–Ampère equation (14.8), then

$$df \in \mathscr{P}^r. \tag{15.12}$$

By (15.4), $\mathscr{P}^r$ is completely integrable in the Frobenius sense and hence defines a foliation on Z. f is constant on the leaves of this foliation.

Proof: Again, this follows from (15.3) and the fact that $d(df) = 0$ for all $f \in \mathscr{F}(Z)$.

One can analyze how many intermediate integrals there are locally for the Monge–Ampère equation (15.9). Let us suppose that

$$\dim \mathscr{P}^2(z) \text{ is constant as } z \text{ varies over } Z. \tag{15.13}$$

$\mathscr{P}^2$ then defines a foliation on Z. Let us suppose that this foliation is *regular* in the sense that it has a quotient space $\overline{Z}$, which is a manifold. Thus, we can parametrize the f's satisfying (15.11) by elements of $\mathscr{F}(\overline{Z})$.

To such an f satisfying (15.12), we consider

$$f: J^1(X,Y) \equiv Z \to R$$

as a scalar-valued symbol of a *first-order* partial differential operator. The solutions of the corresponding first-order partial differential equation will then also be solutions of the Monge–Ampère equation. A general solution for this equation can be found by the Lagrange–Charpit method described earlier; they will be parametrized by curves in the quotient space $\overline{Z}$ of a Lagrange–Vessiot submersion. Two f's can also define the submanifold

$$\overline{f} = 0$$

on $\overline{Z}$. All together we have defined a manifold $W = \overline{Z} \times \overline{\overline{Z}}$ such that the families of solutions of the Monge–Ampère equation are parametrized by *submanifolds* of a certain dimension of W. In case $\mathscr{P} = \mathscr{P}^1$, i.e., the characteristics are Forbenius integrable, this family of solutions will be a *general solution*, in the Ampère sense.

Much of the 19th century geometric theory of partial differential equations is motivated by the goal of generalizing this Monge–Ampère method. From the modern point of view, it is important since

(a) it is "algorithmic,"

(b) it carries through in the C^∞ case.

This manifold W also carries a geometric structure,

which defines various groups. These groups act then on solutions of the Monge–Ampère equations in ways that go "beyond" the action of groups on the jet spaces associated with X and Y.

XVI. FINAL REMARKS

This development of the geometric features of the 19th century theory of nonlinear partial differential equations suggests certain approaches to modern problems of "nonlinear" physics and engineering. First, the Ehresmann theory, combined with Cartan's and Vessiot's methods, assigns certain geometric structures and associated "invariants" to such equations that are not at all obvious when the equations are described in terms of coordinates.

Second, and perhaps more important in terms of computers that can do symbol manipulation more efficiently than humans, it suggests an analysis of the problem of "solubility" of "integrability" of such equations along the lines of "complexity." Thus, a first-order partial differential equation in one dependent variable should be regarded as "less complex" than a second-order one, since there is a geometric "algorithm" for constructing families of solutions of the fomer (and even in a C^∞ way). Certain second-order equations are "less complex" than others, because there is an algorithm for constructing certain of their solutions out of solutions of the "lower-order" or less "complex" systems, and so on.

Another important point to emphasize is the role that this geometric theory can play in the studies of *global* and *singularity properties* of solutions or partial differential equations. For example, the modern theory of linear pseudodifferential-Fourier integral operators emphasize a geometric invariant called the *Maslov index*. This is a characteristic class associated with the fiber bundle structures naturally associated with the jet spaces. A. M. Vinogradov has already[23] presented some general remarks pointing toward possibilities of generalization. The calculus of variations, particularly the second-variation formulae and associated "Morse indices," is also a happy hunting ground for this formalism.[24–26]

[1]E. Goursat, *Lecons sur l'intégration des équations aux derivés partielles du second ordre* (Gauthier-Villars, Paris, 1890), Vols. I and II.
[2]G. Darboux, *Lecons sur la théorie générale des surfaces* (Gauthier-Villars, Paris, 1894, 1915).
[3]E. von Weber, G. Floquet, and E. Goursat, "Proprietés générales des systems d'equations aux derivées partielle," in *Encyclopédie des Sciences Mathématiques* (Gauthier-Villars, Paris, 1913), Tome II, Vol. 4.
[4]C. Riquier, *Les systémes d'equations aux derivées partielles* (Gauthier-Villars, Paris, 1910).
[5]A. Forsyth, *Theory of Differential Equations* (Dover, New York, 1959).
[6]E. Picard, *Traité d'analyse* (Gauthier-Villars, Paris, 1922).
[7]C. Ehresmann: (a) "Sur les structures infinitésimales régularières," Congrés Int. Math. Amsterdam **1**, 479–80 (1954); (b) "Connexions infinitésimales," Colloq. Top. Alg. Bruxelles, 29–55 (1950); (c) "Structures infinitésimales et pseudogroupes de Lie," Colloq. Int. C.N.R.S. Géom. Diff. Strasbourg, 97–110 (1953); (d) C. R. Acad. Sci. Paris **240**, **246**, 360 (1958) (1954); **241**, 397, 1755 (1955); (e) "Connexions d'ordre supérieur," *Atti 5 Cong. dell'Unione Mat. Italiana, 1955* (Cremonese, Rome, 1956), pp. 326–8; (f) "Categories topologiques et catégories différentiables," Colloq. Géom. Diff. Globale Bruxelles, C.B.R.M., 137–150 (1958); (g) "Groupoides differenciales," Rev. Un. Mat. Argentina Buenos-Aires **XIX**, 48 (1960).
[8]R. Hermann, "Some differential geometric aspects of the Lagrange variational problem," Ill. J. Math. **6**, 634–73 (1962).
[9]A. Kumpera and D. C. Spencer, *Lie Equations*, Ann. of Math. Studies **73** (Princeton U.P., Princeton, N.J., 1972).
[10]R. Hermann, *Geometry Physics and Systems* (Marcel Dekker, New York, 1973).
[11]E. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[12]E. T. Whittaker, *A Treatise on Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U.P., London and New York, 1959).
[13]R. Hermann, *Geometry of Non-Linear Differential Equations, Bäcklund Transformations and Solitons, Part A*, Vol. XII of Interdisciplinary Mathematics (Math. Sci. Press, Brookline,Mass. 1976).
[14]D. C. Spencer, "Overdetermined systems of linear partial differential equations," Bull. Am. Math. Soc. **75**, 179–239 (1969).
[15]H. Goldschmidt, (a) "Nonlinear partial differential equations," J. Diff. Geom. **1**, 269–307 (1967); (b) "Sur la structure des équations de Lie," J. Diff. Geom. **6**, 357–73 (1972); **7**, 67–95 (1972).
[16]R. Hermann, "The differential geometry of general mechanical systems from the Lagrangian point of view," J. Math. Phys. **23**, 2077 (1982).
[17]M. Golubitsky and V. Guillemin, *Stable Mappings and Their Singularities* (Springer-Verlag, New York, 1973).
[18]H. Whitney, *Complex Analytic Varieties* (Addison-Wesley, Reading, Mass., 1972).
[19]E. Vessiot, "Sur une théorie nouvelle des problèmes généraux d'integration," Bull. Soc. Math. Fr. **52**, 336–96 (1924).
[20]E. Cartan, *Les systèmes éxterieures et leurs applications géométriques* (Herman, Paris, 1946).
[21]R. Hermann, "E. Cartan's geometric theory of partial differential equations," Adv. in Math. **1**, 265–317 (1965).
[22]B. A. Kupershmidt, in *Lecture Notes in Mathematics* **775** (Springer-Verlag, New York, 1980), pp. 162–218.
[23]A. M. Vinogradov, "Many-valued solutions, and a principle for the classification of nonlinear differential equations," Dokl. Akad. Nauk SSSR **210**, 11–14 (1973) [Sov. Math. Dokl. **14**, 661–5 (1973)], MR 50 No. 1294.
[24]R. Hermann, "The second variation for variational problems in canonical form," Bull. Am. Math. Soc. **71**, 145–8 (1965).
[25]R. Hermann, "The second variation for minimal submanifolds," J. Math. Mech. **16**, 473–92 (1966).
[26]R. Hermann, *Differential Geometry and the Calculus of Variations* (Math. Sci. Press, Brookline, Mass., 1977), 2nd ed.

The Painlevé property for partial differential equations[a)]

John Weiss, M. Tabor, and George Carnevale
Center for Studies of Nonlinear Dynamics, La Jolla Institute, La Jolla, California 92038

(Received 6 May 1982; accepted for publication 22 October 1982)

In this paper we define the Painlevé property for partial differential equations and show how it determines, in a remarkably simple manner, the integrability, the Bäcklund transforms, the linearizing transforms, and the Lax pairs of three well-known partial differential equations (Burgers' equation, KdV equation, and the modified KdV equation). This indicates that the Painlevé property may provide a unified description of integrable behavior in dynamical systems (ordinary and partial differential equations), while, at the same time, providing an efficient method for determining the integrability of particular systems.

PACS numbers: 02.30.Jr

I. INTRODUCTION

The Painlevé property for ordinary differential equations is defined as follows. The solutions of a system of ordinary differential equations are regarded as (analytic) functions of a complex (time) variable. The "movable" singularities[1] of the solution are the singularities of the solution (as a function of complex t) whose location depends on the initial conditions and are, hence, movable. (Fixed singularities occur at points where the coefficients of the equation are singular.) The (ordinary differential) system is said to possess the Painlevé property when all the movable singularities are single-valved (simple poles).[1]

Experience has shown that, when an (ode) system possesses the Painlevé property, the system will be "integrable."[2–4] Albowitz *et al.*[5] have proven that when a partial differential equation is solvable by the inverse scattering transform[6] and a system of ordinary differential equations is obtained from this pde by an exact similarity reduction then the solution (of this system of ode's) associated with the Gel'fand–Levitan–Marchenko equation will possess the Painlevé property. Furthermore, they conjecture that, when *all* the ordinary differential equations obtained by exact similarity transforms from a given partial differential equation have the Painlevé property, perhaps after a change of variables, then the partial differential equation will be "integrable." In another context Chudnovsky[7] has observed that a certain partial differential equation which arises as a compatability condition in a generalization of the theory of Isomonodromy Deformations is meromorphic [in (x,t)]. It is somewhat remarkable that a connection between integrability and the Painlevé property has been noted since the work of Kowalevskaya,[2] and, as of yet, the precise equivalence of these concepts remains to be determined.

In this paper, we define a "Painlevé property" for partial differential equations that does *not* refer to that for ordinary differential equations. Indeed, we believe that by extending the definition of the Painlevé property it is possible to treat the phenomenon of integrable behavior in a unified manner. Furthermore, as a consequence of the procedure defined in the next section, the Cole–Hopf transform for Burgers' equation and the Lax pair for the Korteweg–de Vries equation are found in a remarkably simple and natural manner. Bäcklund transforms for both of these equations and the modified kdV equation are also found. Additional results for the sine–Gordon, Boussinesq, and the Kadomstev–Petviashvili equations are presented in the Appendix.

II. THE PAINLEVÉ PROPERTY

One major difference between analytic functions of one complex variable and analytic functions of several complex variables is that, in general, the singularities of a function of several complex variables cannot be isolated.[8] If $f = f(z_1,...,z_n)$ is a meromorphic function of N complex variable ($2N$ real variables), the singularities of f occur along analytic manifolds of (real) dimension $2N$-2. These manifolds are determined by conditions of the form

$$\phi(z_1,...,z_n) = 0, \tag{2.1}$$

where ϕ is an analytic function of $(z_1,...,z_n)$ in a neighborhood of the manifold.[8]

Therefore, we say that a partial differential equation has the Painlevé property when the solutions of the pde are "single-valued" about the movable, singularity manifolds. To be precise, if the singularity manifold is determined by (2.1) and $u = u(z_1,...,z_n)$ is a solution of the partial differential equation, then we assume that

$$u = u(z_1,...,z_n) = \phi^{\alpha} \sum_{j=0}^{\infty} u_j \phi^j, \tag{2.2}$$

where

$$\phi = \phi(z_1,...,z_n)$$

and

$$u_j = u_j(z_1,...,z_n)$$

are analytic functions of $(z_1,...,z_n)$ in a neighborhood of the manifold (2.1), and α is an integer. Substitution (2.2) into the partial differential equations determines the possible values of α and defines the recursion relations for $u_j, j = 0,1,2,...$. The procedure is analogous to that for ordinary differential equations.[3] In the next three sections we apply this procedure to Burgers' equation, the Korteweg–de Vries equation, and the modified KdV equation.

[a)] This work supported by Department of Energy Contract DOE 10923 and ONR Contract No. N00014-79-C-0537.

 0022-2488/83/030522-05$02.50

III. BURGERS' EQUATION

For Burgers' equation

$$u_t + uu_x = \sigma u_{xx}\,, \tag{3.1}$$

we assume that

$$u = \phi^{\alpha} \sum_{j=0}^{\infty} u_j \phi^j\,, \tag{3.2}$$

where

$$\phi = \phi(x,t)\,, \quad u_j = u_j(x,t)$$

are analytic functions of (x,t) near $M = \{(x,t)\colon \phi(x,t) = 0\}$.

If M is a singularity manifold, it is readily found that

$$\alpha = -1 \tag{3.3}$$

by a leading order analysis. The recursion relations for u_j are found to be

$$\begin{aligned} &u_{j-2,t} + (j-2)u_{j-1}\phi_t \\ &\quad + \sum_{m=0}^{j} u_{j-m}\left[u_{m-1,x} + (m-1)\phi_x u_m\right] \\ &\quad = \sigma\left[u_{j-2,xx} + 2(j-2)u_{j-1,x}\phi_x\right. \\ &\quad \left. + (j-2)u_{j-1}\phi_{xx} + (j-1)(j-2)u_j\phi_x^2\right]\,. \end{aligned} \tag{3.4}$$

Collecting terms involving u_j, it is found that

$$\sigma\phi_x^2(j-2)(j+1)u_j = F(u_{j-1},\ldots,u_0,\phi_t,\phi_x,\phi_{xx},\ldots)$$

$$\text{for } j = 0,1,2,\ldots\,. \tag{3.5}$$

We note that the recursion relations (3.5) are not defined when $j = -1,2$. These values of j are called the "resonances"[3] of the recursion relation and correspond to points where arbitrary functions of (x,t) are introduced into the expansion. $j = -1$ corresponds to the arbitrary (undefined) singularity manifold ($\phi = 0$). On the other hand, the resonance at $j = 2$ introduces an arbitrary function u_2 and a "compatability condition" on the functions (ϕ,u_0,u_1) that requires the right-hand side of (3.5) vanish identically.

For Burgers' equation, we find from (3.4)

$$j = 0\,, \quad u_0 = -2\sigma\phi_x\,, \tag{3.6}$$

$$j = 1\,, \quad \phi_t + u_1\phi_x = \sigma\phi_{xx}\,, \tag{3.7}$$

$$j = 2\,, \quad \partial_x(\phi_t + u_1\phi_x - \sigma\phi_{xx}) = 0\,. \tag{3.8}$$

By (3.7) the compatability condition (3.8) at $j = 2$ is satisfied identically. Thus, Burgers' equation possesses the Painlevé property. Furthermore, if we set the arbitrary function u_2 equal to zero,

$$u_2 = 0\,, \tag{3.9}$$

and require that

$$u_{1t} + u_1u_{1x} = \sigma u_{1xx}\,, \tag{3.10}$$

then

$$u_j = 0\,, \quad j \geqslant 2\,.$$

In this case, we find the following Bäcklund transform[9] for Burgers' equation,

$$u = -2\sigma\phi_x/\phi + u_1\,, \tag{3.11}$$

where (u,u_1) satisfy Burgers' equation and

$$\phi_t + u_1\phi_x = \sigma\phi_{xx}\,. \tag{3.12}$$

When $u_1 = 0$, the Cole–Hopf transform[6] is obtained, and when $u_1 = \phi$ is found that

$$u = -2\sigma\phi_x/\phi + \phi\,, \tag{3.13}$$

where

$$\phi_t + \phi\phi_x = \sigma\phi_{xx}\,. \tag{3.14}$$

The Bäcklund transform (3.13) and (3.14) was discovered by Fokas,[10] using the method of Lie symmetries. The general form of the Bäcklund transform, (3.11) and (3.12), appears to be a new result.

IV. KORTEWEG–DE VRIES EQUATION

Following the procedure applied in the preceding section, we find that for the KdV equation

$$u_t + uu_x + \sigma u_{xxx} = 0\,, \tag{4.1}$$

$$u = \phi^{-2} \sum_{j=0}^{\infty} u_j \phi^j\,. \tag{4.2}$$

The recursion relations for the u_j are presented elsewhere.[11] It is found that the "resonances" occur at

$$j = -1,4,6\,. \tag{4.3}$$

The compatability conditions at $j = 4,6$ are satisfied identically and the KdV equation possesses the Painlevé property. For instance, when

$$j = 0\,, \quad u_0 = -12\sigma\phi_x^2\,, \tag{4.4}$$

$$j = 1\,, \quad u_1 = 12\sigma\phi_{xx}\,, \tag{4.5}$$

$$j = 2\,, \quad \phi_x\phi_t + \phi_x^2u_2 + 4\sigma\phi_x\phi_{xxx} - 3\sigma\phi_{xx}^2 = 0\,, \tag{4.6}$$

$$j = 3\,, \quad \phi_{xt} + \phi_{xx}u_2 - \phi_x^2u_3 + \sigma\phi_{xxxx} = 0\,, \tag{4.7}$$

$j = 4$, compatability condition:

$$\frac{\partial}{\partial x}(\phi_{xt} + \sigma\phi_{xxxx} + \phi_{xx}u_2 - \phi_x^2u_3) = 0\,. \tag{4.8}$$

By (4.7) the compatability condition (4.8) at $j = 4$ is satisfied identically. The compatability condition at $j = 6$ involves extensive calculation and is, therefore, presented elsewhere.[11]

We now specialize (4.2) by setting the "resonance" functions

$$u_4 = u_6 = 0\,. \tag{4.9}$$

Furthermore, by requiring

$$u_3 = 0\,, \tag{4.10}$$

it is easily demonstrated that

$$u_j = 0\,, \quad j \geqslant 3\,, \tag{4.11}$$

if

$$u_{2t} + u_2u_{2x} + \sigma u_{2xxx} = 0\,. \tag{4.12}$$

Collecting (4.4)–(4.12), we find that

(i) $u_j = 0$, $j \geqslant 3$,

(ii) $u_0 = -12\sigma\phi_x^2$, $u_1 = 12\sigma\phi_{xx}$,

(iii) (a) $\phi_x\phi_t + \phi_x^2u_2 + 4\sigma\phi_x\phi_{xxx} - 3\sigma\phi_{xx}^2 = 0$,

(b) $\phi_{xt} + \phi_{xx}u_2 + \sigma\phi_{xxxx} = 0$,

(iv) $u_{2t} + u_2u_{2x} + \sigma u_{2xxx} = 0$,

(v) $u = -12\sigma\phi_x^2/\phi^2 + 12\sigma\phi_{xx}/\phi + u_2$,

or

$$u = 12\sigma(\partial^2/\partial x^2)\ln\phi + u_2 ,$$

where

$$u_t + uu_x + \sigma u_{xxx} = 0 . \tag{4.13}$$

(4.13) will define a Bäcklund transform for the KdV equation if equations (4.13.iii) and (4.13.iv) are consistent. By solving Eq. (4.13.iiia) for ϕ_t, then differentiating w.r.t. x, using Eq. (4.13.iiib), and making the change of variable

$$\phi_x = V^2 \tag{4.14}$$

it is found that

$$\begin{aligned} &\text{(i)} \quad 6\sigma V_{xx} + u_2 V = \lambda V , \\ &\text{(ii)} \quad 2V_t + u_2 V_x + \lambda V_x + 2\sigma V_{xxx} = 0 . \end{aligned} \tag{4.15}$$

Thus, (4.13.iii) reduces to (4.15), which is just the Lax pair[12] for (4.13.iv), as may be verified by direct calculation.[11] In summary, our result reads

$$u = 12\sigma\frac{\partial^2}{\partial x^2}\ln\phi + u_2 ,$$

where

$$\phi_x = V^2 ,$$
$$6\sigma V_{xx} + u_2 V = \lambda V ,$$
$$2V_t + u_2 V_x + \lambda V_x + 2\sigma V_{xxx} = 0 ,$$

and

$$u_{2t} + u_2 u_{2x} + \sigma u_{2xxx} = 0 ,$$
$$u_t + uu_x + \sigma u_{xxx} = 0 . \tag{4.16}$$

This Bäcklund transform connects two solutions of the KdV equation via the squared eigenfunction associated with one solution by the IST transform. Although a Bäcklund transform for the KdV is hardly novel, the construction of the transform from a finite Painlevé expansion about the singularity manifold is obtained in a remarkably straightforward way.

V. THE MODIFIED KdV EQUATION

For the modified KdV equation,

$$u_t = \frac{\partial}{\partial x}(u^3 - 2\sigma^2 u_{xx}) , \tag{5.1}$$

it is found that

$$u = \phi^{-1}\sum_{j=0}^{\infty} u_j \phi^j , \tag{5.2}$$

where

$$u_0 = 2\sigma\phi_x .$$

The resonances occur at $j = -1,3$, and 4. The compatability conditions at $j = 3,4$ are satisfied identically and the modified KdV equations possesses the Painlevé property. For reference, the recursion relations are

$$u_{j-3,t} + (j-3)\phi_t u_{j-2} = \psi_{j-1,x} + (j-3)\phi_x \psi_j , \tag{5.3}$$

where

$$\psi_j = \sum_{m=0}^{j}\sum_{l=0}^{m} u_{j-m}u_{m-l}u_l - 2\sigma^2(\theta_{j-1,x} + (j-2)\phi_x\theta_j) \tag{5.4}$$

and

$$\theta_j = u_{j-1,x} + (j-1)\phi_x u_j .$$

We find that

$$j = 0 , \quad u_0 = 2\sigma\phi_x , \tag{5.5}$$

$$j = 1 , \quad u_1 = -\sigma\phi_{xx}/\phi_x , \tag{5.6}$$

$$j = 2 , \quad \phi_t = 6\sigma\phi_x^2 u_2 + 3\sigma^2\phi_{xx}^2/\phi_x - 2\sigma^2\phi_{xxx} . \tag{5.7}$$

To find a Bäcklund transform, the "resonance" functions u_3 and u_4 are set equal to zero. To obtain a finite expansion, it is found that, in general, u_2 must vanish as well. Thus, when

$$u_2 = u_3 = u_4 = 0 , \tag{5.8}$$

(5.3) implies that

$$u_j = 0 , \quad j \geqslant 2 ,$$

and

$$u = 2\sigma\phi_x/\phi + u_1 . \tag{5.9}$$

Furthermore, (5.6) and (5.7) with $u_2 = 0$ lead to

$$u_{1t} = \frac{\partial}{\partial x}(u_1^3 - 2\sigma^2 u_{1xx}) , \tag{5.10}$$

where

$$u_1 = -\sigma\phi_{xx}/\phi_x . \tag{5.11}$$

Equations (5.9), (5.10), and (5.11) constitute an *explicit* Bäcklund transform for the modified KdV equation.

APPENDIX: SINE-GORDON, BOUSSINESQ AND KP EQUATIONS

A. The sine-Gordon equation

$$u_{xt} = \sin u . \tag{A1}$$

The substitution

$$V = e^{iu} \tag{A2}$$

leads to the equation

$$2VV_{xt} - 2V_x V_t = V^3 - V . \tag{A3}$$

We find that

$$V = \phi^{-2}\sum_{j=0}^{\infty} V_j \phi_j , \tag{A4}$$

where

$$V_0 = 4\phi_x\phi_t \quad \text{and} \quad V_1 = -4\phi_{xt} .$$

The resonances occur at

$$j = -1,2 . \tag{A5}$$

Again, the compatability condition at $j = 2$ is satisfied identically and the equation possesses the Painlevé property. The recursion relations are found to be

$$2\sum_{m=0}^{j} V_{j-m}A_m - 2\sum_{m=0}^{j} B_{j-m}C_m = \sum_{m=0}^{j}\sum_{l=0}^{m} V_{j-m}V_{m-l}V_l - V_{j-4} , \tag{A6}$$

where

$$A_j = V_{j-2,xt} + (j-3)(\phi_t V_{j-1,x} + \phi_x V_{j-1,t} + \phi_{xt} V_{j-1}) + (j-2)(j-3)\phi_x \phi_t V_j ,$$

$$B_j = V_{j-1,x} + (j-2)\phi_x V_j ,$$

and

$$C_j = V_{j-1,t} + (j-2)\phi_t V_j .$$

From the recursion relations or more directly from (A3), it follows that any truncation of (A4) must satisfy

$$V_j = 0 , \quad j \geqslant 3 . \tag{A7}$$

In the search for a possible Bäcklund transformation, we are thus led to examine the truncation

$$V = \phi^{-2} V_0 + \phi^{-1} V_1 + V_2 = -4 \frac{\partial^2}{\partial x \partial t} \ln \phi + V_2 . \tag{A8}$$

It then follows that the resonance function V_2 must satisfy Eq. (A3). In addition, for (A7) to be verified (V_2, ϕ) must satisfy an overdetermined system of three nonlinear pde's obtained from the conditions

$$V_3 = V_4 = V_5 = 0 . \tag{A9}$$

[From $V_6 = 0$, V_2 must satisfy (A31). Simplification of the conditions defining the BT (A8) is currently in progress.]

B. The Boussinesq equation

$$u_{tt} + 2uu_{xx} + 2u_x^2 + \tfrac{1}{3} u_{xxxx} = 0 . \tag{A10}$$

It is found that

$$u = \phi^{-2} \sum_{j=0}^{\infty} u_j \phi^j . \tag{A11}$$

The resonances occur at

$$j = -1,4,5,6 . \tag{A12}$$

From the recursion relations we find that

$$j = 0 , \quad u_0 = -2\phi_x^2 , \tag{A13}$$

$$j = 1 , \quad u_1 = 2\phi_{xx} , \tag{A14}$$

$$j = 2 , \quad \phi_t^2 - \phi_{xx}^2 + \tfrac{4}{3} \phi_x \phi_{xxx} + 2u_2 \phi_x^2 = 0 , \tag{A15}$$

$$j = 3 , \quad \phi_{tt} + \tfrac{1}{3} \phi_{xxxx} + 2\phi_{xx} u_2 - 2\phi_x^2 u_3 = 0 , \tag{A16}$$

$j = 4$ (resonance), u_4 is arbitrary if the compatability condition

$$\frac{\partial^2}{\partial x^2} (\phi_t^2 - \phi_{xx}^2 + \tfrac{4}{3} \phi_x \phi_{xx} + 2u_2 \phi_x^2) = 0 \tag{A17}$$

is satisfied. By (A15) this is so.

$j = 5$ (resonance), u_5 is arbitrary if

$$\frac{\partial^2}{\partial x^2} (\phi_{tt} + \tfrac{1}{3} \phi_{xxxx} + 2\phi_{xx} u_2 - 2\phi_x^2 u_3) = 0 . \tag{A18}$$

By (A16) this is always satisfied.

$j = 6$ (resonance): We find that u_6 is arbitrary if a rather complicated compatibility condition is satisfied. After algebraic reduction, this condition is found to involve (ϕ, u_2, u_3), or, using (A15) and (A16), to reduce to an identity in the arbitrary function ϕ. The verification of this identity becomes trivial when the "reduced anzatzs"[13]

$$\phi = x - \psi(t) \tag{A19}$$

is employed. Basically, when $\phi_x \neq 0$ [see (A13)], the implicit function theorem indicates that $\phi = \phi(x,t)$ may be "locally" (near $\phi = 0$) represented in the form (A19), where

$$\phi(\psi(t), t) = 0 . \tag{A20}$$

In this manner the compatability condition is verified, and the Boussinesq equation is found to identically Painlevé.

To define a Bäcklund transform it is necessary to truncate (A11) at the "constant" level term. That is,

$$u_j = 0 , \quad j \geqslant 3 . \tag{A21}$$

We find from (A13)–(A16) that

$$u = 2 \frac{\partial^2}{\partial x^2} \ln \phi + u_2 , \tag{A22}$$

where (u, u_2) satisfy (A10), defines a Bäcklund transform, if

$$u_{2t} + 2u_2 u_{2x} + 2u_{2x}^2 + \tfrac{1}{3} u_{2xxx} = 0 , \tag{A23}$$

$$\phi_t^2 - \phi_{xx}^2 + \tfrac{4}{3} \phi_x \phi_{xx} + 2u_2 \phi_x^2 = 0 , \tag{A24}$$

and

$$\phi_{tt} + \tfrac{1}{3} \phi_{xxxx} + 2\phi_{xx} u_2 = 0 . \tag{A25}$$

By a method presented in Ref. 14 the Lax pair can be found from (A24) and (A25).

C. The Kadomstev–Petviashvili, or two-dimensional KdV, equation

The KP equation[15]

$$u_{tx} + u_x^2 + uu_{xx} + \delta u_{xxxx} + u_{yy} = 0 \tag{A26}$$

has an expansion of the form

$$u = \phi^{-2} \sum_{j=0}^{\infty} u_j \phi^j \tag{A27}$$

with resonances at

$$j = -1,4,5,6 . \tag{A28}$$

From the recursion relations we find that

$$j = 0 , \quad u_0 = -12\sigma\phi_x^2 , \tag{A29}$$

$$j = 1 , \quad u_1 = 12\sigma\phi_{xx} , \tag{A30}$$

$$j = 2 , \quad \phi_t \phi_x + 4\sigma\phi_x \phi_{xxx} - 3\sigma\phi_{xx}^2 + \phi_y^2 + u_2 \phi_x^2 = 0 , \tag{A31}$$

$$j = 3 , \quad \phi_{xt} + \sigma\phi_{xxxx} + \phi_{yy} + \phi_{xx} u_2 - \phi_x^2 u_3 = 0 , \tag{A32}$$

$j = 4$ (resonance), u_4 is arbitrary if the compatability condition

$$\frac{\partial^2}{\partial x^2} (\phi_x \phi_t + 4\sigma\phi_x \phi_{xxx} - 3\sigma\phi_{xx}^2 + \phi_y^2 + u_2 \phi_x^2) = 0 . \tag{A33}$$

By (A31) this is so.

$j = 5$ (resonance), u_5 is arbitrary if

$$\frac{\partial^2}{\partial x^2} (\phi_{xt} + \sigma\phi_{xxxx} + \phi_{yy} + \phi_{xx} u_2 - \phi_x^2 u_3) = 0 . \tag{A34}$$

By (A32) this is so.

$j = 6$ (resonance): Again u_6 will be arbitrary if a complicated compatibility condition is satisfied. By the "reduced" anzatz

(see the last section)

$$\phi = x - \psi(t,y) \tag{A35}$$

it can be shown that this condition is satisfied identically.

By the above considerations the KP equation is found to be identically Painlevé.

The associated Bäcklund transform, defined by truncating at the "constant" term, is

$$u = 12\sigma \frac{\partial^2}{\partial x^2} \ln \phi + u_2 . \tag{A36}$$

In Ref. 14 we demonstrate the consistency of this transform by finding the Lax pairs for this equation.

[1](a) E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), (b) Einar Hille, *Ordinary Differential Equations in the Complex Plane* (Wiley, New York, 1976).

[2]S. Kowalevskaya, Acta Math. **14**, 81 (1890).

[3]Y. F. Chang, M. Tabor, and J. Weiss, "Analytic structure of the Henon–Heiles Hamiltonian in integrable and nonintegrable regimes," J. Math. Phys. **23**, 531 (1982).

[4]T. Bountis, H. Segur, and F. Viualdi, "Integrable Hamiltonian systems and the Painlevé property," Phys. Rev. A **25**, 1257 (1982).

[5]M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).

[6]G. B. Whitham, *Linear and Non-Linear Waves* (Wiley, New York, 1974).

[7]D. V. Chudnovsky, "Riemann Monodromy Problem, Isomonodromy Deformation Equations and Completely Integrable Systems," in *Bifurcation Phenomena in Mathematical Physics*, edited by C. Bardos and D. Bessis (D. Reidel, New York, 1979).

[8]W. F. Osgood, *Topics in the Theory of Functions of Several Complex Variables* (Dover, New York, 1966).

[9]R. M. Miura, Ed., *Bäcklund Transformations, the Inverse Scattering Method, Solitons, and Their Applications*, Lecture Notes in Mathematics, Vol. 515 (Springer-Verlag, New York, 1976).

[10]A. S. Fokas, "Invariants, Lie–Bäcklund Operators, and Bäcklund Transformations," Ph.D. thesis, California Institute of Technology, 1979.

[11]J. Weiss, "The Painlevé Property for Partial Differential Equations," La Jolla Institute Report LJI-R-82-186, 1982.

[12]P. D. Lax, "Integrals of Nonlinear Equations of Evolution and Solitary Waves," Commun. Pure Appl. Math. **21**, 674 (1968).

[13]M. Kruskal (private communication).

[14]J. Weiss, "The Painlevé property for partial differential equations. II. Bäcklund transformations, Lax pairs, and the Schwarzian derivative," J. Math. Phys. (to be published).

[15]V. S. Dryuma, "Analytic Solutions of the Two-Dimensional Kortweg–de Vries Equation," Pis'ma Zh. Eksp. Teor. Fiz. **19**, 753 (1974) [JETP Lett. **19**, 387 (1974)].

Real axis integration of Sommerfeld integrals with applications to printed circuit antennas[a)]

P. B. Katehi and N. G. Alexopoulos
Electrical Engineering Department, University of California-Los Angeles, Los Angeles, California 90024

(Received 16 March 1982; accepted for publication 17 September 1982)

Printed circuit antennas are becoming an integral part of imaging arrays in microwave, millimeter, and submillimeter wave frequencies. The electrical characteristics of such antennas can be analyzed by solving integral equations of the Fredholm first kind. The kernel involves Sommerfeld integrals which are particularly difficult to solve when source and field points lie on an electrical discontinuity, as it occurs in the determination of the characteristics of printed circuit antennas. An analytic-numeric real axis integration technique has been developed for such integrals and it is combined with piece-wise sinusoidal expansions to solve the Fredholm integral equation for the unknown current density.

PACS numbers: 02.30.Rz, 84.40.Gf, 85.20.Lg

I. INTRODUCTION

When sources of electromagnetic radiation are in the proximity of or on the interface of electrically dissimilar materials, the resulting electromagnetic fields involve Sommerfeld type integrals.[1-8] If the source is near the interface, these integrals include exponentially decaying terms in the integrand which facilitate convergence.[5-8] On the other hand whenever the source and observation points are on the interface these integrals converge very slowly.

The recent evolution of printed circuit antennas[2-4, 9-14] with applications in microwave, millimeter and submillimeter wave frequencies, has necessitated the development of special techniques to solve the resulting Sommerfeld type integrals. The printed circuit antenna is an extended source typically printed on a grounded dielectric substrate. In order to evaluate the electrical properties of such antennas, the current distribution must be obtained by solving Pocklington's integral equation, a Fredholm integral equation of the first kind.[15] In the process of solving for the antenna current distribution, the field and source points lie on the interface between vacuum and the substrate. This renders the resulting Sommerfeld-type integrals very slowly convergent.

In this paper special techniques are developed to address this case. This is done in the effort to obtain the current distribution of dipoles printed on a substrate, as compared to previous approaches where the antenna current was assumed to be uniform and the antennas were chosen to be off the interface.[5-8] The integral equation for the unknown current distribution is reduced into a system of linear equations by applying the method of moments.[16] The solution of the resulting matrix equation leads to the evaluation of integrals which are of the form

$$I=\int_0^\infty \mathscr{L}\{J_0(\lambda\rho)\}f(\lambda)\,d\lambda\,,$$

where $\mathscr{L}$ is a self-adjoint operator which involves integration with respect to ρ, $J_0(\lambda\rho)$ is the zeroth order Bessel function of the first kind and $f(\lambda)$ is given by $f(\lambda)=Q(\lambda)/P(\lambda)$ where $Q(\lambda)$ and $P(\lambda)$ are expressions involving transcendental functions. The semi-infinite path of integration along the real axis is divided in four subintervals, with the integrand being treated differently in each one. The fourth subinterval extends from a constant $A\gg 1$ to infinity, wherein the integrand is a fast varying function and converges very slowly.

The difficulties which are encountered within the fourth subinterval are resolved by approximating the integrand with controlled error and by applying a combination of analytical and numerical evaluation of the resulting integral. This procedure introduces an error which is of the order 10^{-4} of the correct value of the integral.

II. REDUCTION OF INTEGRAL EQUATION TO A SYSTEM OF LINEAR EQUATIONS

The integral equation resulting from the solution of the boundary value problem (see Fig. 1) gives an expression for the radiated electric field in the form

$$\vec{E}(\vec{r})=\int_L [k^2\bar{\bar{I}}+\vec{\nabla}\vec{\nabla}]\cdot\bar{\bar{G}}(\vec{r},\vec{r}')\cdot\vec{J}(\vec{r}')dr'\,, \tag{2.1}$$

where $\bar{\bar{G}}(\vec{r},\vec{r}')$ is the Green's function of the problem given by

$$\bar{\bar{G}}(\vec{r},\vec{r}')=\int_0^\infty J_0(\lambda\,|\vec{r}-\vec{r}'|)\bar{\bar{F}}(|\vec{r}-\vec{r}'|,\lambda)\,d\lambda \tag{2.2}$$

with

$$\bar{\bar{F}}(|\vec{r}-\vec{r}'|,\lambda)=f_x(|\vec{r}-\vec{r}'|\lambda)\hat{x}\hat{x}+f_y(|\vec{r}-r'|\lambda)\hat{z}\hat{x}\,, \tag{2.3}$$

while $\vec{J}(\vec{r}')$ is the unknown current distribution on the an-

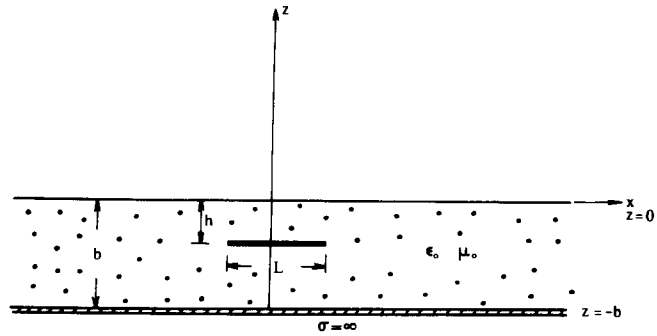


FIG. 1. Horizontal electric dipole embedded in a grounded dielectric slab.

a) Work sponsored by U. S. Army Contract DAAG 29-79-C-0050.

 0022-2488/83/030527-07$02.50

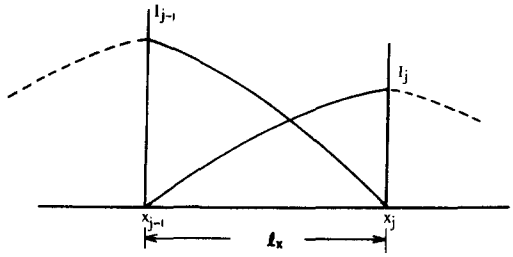


FIG. 2. Piecewise-sinusoidal currents on a wire segment.

tenna. The reduction of the integral equation to a matrix equation requires the expansion of the current distribution in a finite series of basis functions $\phi_i(|\vec{r}'|,x_i)$ as follows

$$\vec{J}(\vec{r}') = \sum_{i=1}^{N} I_i \phi_i(|\vec{r}'|,x_i), \tag{2.4}$$

where I_i are unknown constants to be determined. For the problem of interest the most appropriate set of basis functions has been found to be in the form (see Fig. 2)

$$\phi_i(|\vec{r}'|,x_i) = \left\{ P_{i-1} \frac{\sin[k(x'-x_{i-1})]}{\sin kl_x} + P_i \frac{\sin[k(x_{i+1}-x')]}{\sin kl_x} \right\}, \tag{2.5}$$

where $\vec{r}' = x'\hat{x}$ and

$$P_{i-1} = \begin{Bmatrix} 1, & x_{i-1} \leqslant x' \leqslant x_i \\ 0 & \text{elsewhere} \end{Bmatrix}. \tag{2.6}$$

By combining Eqs. (2.1), (2.2) and (2.4) the radiated field can be written in the form

$$\vec{E}(\vec{r}) = \sum_{i=1}^{N} I_i \int_0^\infty \overline{\mathscr{L}}_i\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\}\cdot\hat{x}\,d\lambda, \tag{2.7}$$

where

$$\overline{\mathscr{L}}_i\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\} = \int_L [k^2\bar{\bar{I}} + \vec{\nabla}\vec{\nabla}]\cdot\bar{\bar{F}} J_0(\lambda\,|\vec{r}-\vec{r}'|)dr'. \tag{2.8}$$

The integral equation given by (2.7) is solved by Galerkin's method with a set of weighting functions $\vec{w}_m$ being identical to the basis function, i.e.,

$$\{\vec{w}_m\} = \phi_m(|\vec{r}|,x_m), \tag{2.9}$$

subsequently, an inner product is formulated as follows

$$\langle \vec{w}_m,\vec{a}\rangle = \int_L dr(\vec{w}_m\cdot\vec{a}) \quad \text{with } \vec{r} = x\hat{x} \tag{2.10}$$

$$\sum_{i=1}^{N} I_i \int_0^\infty \langle \vec{w}_k,\overline{\mathscr{L}}_i\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\}\cdot\hat{x}\rangle d\lambda = \langle \vec{w}_k,\vec{E}(\vec{r})\rangle$$

for the unknown coefficients I_i. By noting that

$$\langle \vec{w}_k,\overline{\mathscr{L}}_i\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\}\cdot\hat{x}\rangle = \langle \phi_k\hat{x},\overline{\mathscr{L}}_i\cdot\hat{x}\rangle = \langle \phi_k,\mathscr{L}_i\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\}f(\lambda)\rangle \tag{2.11}$$

with

$$\mathscr{L}_i\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\} = \int_L \phi_i(\vec{r}',x_i)J_0(\lambda\,|\vec{r}-\vec{r}'|)dr', \tag{2.12}$$

Eq. (2.10) can be written as

$$[I_i]\cdot[Z_{ik}] = [V_k], \tag{2.13}$$

where V_k is the voltage excitation vector given by

$$V_k = \langle \vec{w}_k,\vec{E}(\vec{r})\rangle = \begin{cases} 1 & \text{feedpoint at } x_k, \\ 0 & \text{elsewhere}, \end{cases} \tag{2.14}$$

and

$$Z_{ik} = \int_0^\infty \mathscr{L}_{ik}\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\}f(\lambda)\,d\lambda. \tag{2.15}$$

Here $\mathscr{L}_{ik}\{\ \}$ is the finite double-integral operator

$$\mathscr{L}_{ik}\{\ \} = \int_L dx \int_L dx'\,\Phi_k(x,x_k)\Phi_i(x',x_i)\{\ \}, \tag{2.16}$$

where

$$\Phi_i(x',x_i) = [\phi_i(x',x_i) + S_{i-1}\delta(x'-x_{i-1}) + S_i\delta(x'-x_i) + S_{i+1}\delta(x'-x_{i+1})] \tag{2.17}$$

and S_{i-1}, S_i, S_{i+1} are independent of x'.

In this manner the integral equation is reduced to a system of linear equations which can be solved for the unknown coefficients I_i. The matrix inversion requires a small fraction of the total computation time, while a considerable part of the computational effort is spent on the evaluation of the elements of the matrix Z.

III. EVALUATION OF THE INTEGRAL IN THE GENERALIZED IMPEDANCE MATRIX

A. Singular points and related surface waves

It can be shown analytically that the matrix elements in the matrix equation are given by

$$Z_{ik} = \int_0^\infty \mathscr{L}_{ik}\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\}\left[\frac{A(\lambda,b)}{f_1(\lambda,b)f_2(\lambda,b)}\right]d\lambda, \tag{3.1}$$

where

$$f_1(\lambda,b) = u_0\sinh(ub) + u\cosh(ub), \tag{3.2}$$

$$f_2(\lambda,b) = \epsilon_r u_0\cosh(ub) + u\sinh(ub), \tag{3.3}$$

$$A(\lambda,b) = [C_1\sinh[u(b-h)] + C_2\sinh(ub)] \times [D_1u_0\cosh(ub) + D_2u\sinh(ub)] \times \left[E_1 f_2(\lambda,b)(\cosh(uh) - \frac{u_0}{u}\sinh(uh)) + E_2\right] \tag{3.4}$$

with C_i, D_i, E_i $(i = 1,2)$ being constants. Since the technique for the evaluation of the elements Z_{ik} is the same for any finite values of the above constants, for simplicity E_1 is considered equal to zero throughout this paper.

In Eq. (3.1) the integrand is a function of the parameter λ through the radicals

$$u = [\lambda^2 - k^2]^{1/2}, \tag{3.5}$$

$$u_0 = [\lambda^2 - k_0^2]^{1/2}, \tag{3.6}$$

where k_0,k are the wavenumbers in vacuum and substrate, respectively. The sign of the radical u does not affect the single-value of the integrals, as the terms involving the radical are even functions of u. For this reason, only the branch cut contribution by the radical u_0 is considered and its direction is determined by convergence of the integrals, the re-

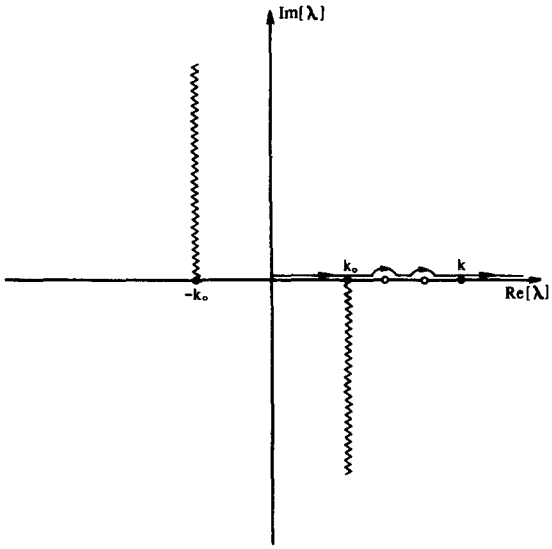


FIG. 3. Path of integration.

quirement that the radiated field is a wave receding from the source and the integrals must be single valued. As a result the restrictions on λ are

$$\mathrm{Re}\,[\lambda]>0\,, \tag{3.7}$$

$$\mathrm{Im}\,[\lambda]>0\,, \tag{3.8}$$

which in turn impose the following behavior for u and u_0:

$$\mathrm{Re}(u_0)>0\,,\quad \mathrm{Im}(u_0)>0\,, \tag{3.9}$$

$$\mathrm{Re}(u)>0\,,\quad \mathrm{Im}(u)>0\,. \tag{3.10}$$

A possible position of the branch cuts governed by these inequalities is shown in Fig. 3.

The integrand in Eq. (3.1), with $E_i=0$, has poles whenever either one of the functions $f_1(\lambda,b)$, $f_2(\lambda,b)$ becomes zero. The zeros of these two functions correspond to surface-wave modes. Particularly the zeros of $f_1(\lambda,b)$ correspond to TE surface waves while the zeros of $f_2(\lambda,b)$ to TM surface waves. In the case of lossless dielectric these TE and TM poles are the roots of the equations $u_0=-u\coth(ub)$ and $\epsilon_r u_0=-u\tanh(ub)$. Furthermore, these poles lie within the range $k_0<\mathrm{Re}(\lambda)<k$.

The infinite integration in Eq. (3.1) is performed along the real axis and it is completed in two steps:

(i) Numerical integration over the interval $[0,A]$ where A satisfies the relationship $\coth((A^2-k^2)^{1/2}b)\doteq 1$;

(ii) Combination of numerical and analytical integration for the evaluation of the tail contribution which is actually the integration over the path $[A,\infty)$.

Subsequently it is concluded that the (i,k)th element of the impedance matrix can be split into two parts as it is shown below

$$Z_{ik}=Z_{ik}^{(1)}+Z_{ik}^{(2)}\,, \tag{3.11}$$

where

$$Z_{ik}^{(1)}=\int_0^A \mathcal{L}_{ik}\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\}\left[\frac{A(\lambda,b)}{f_1(\lambda,b)\cdot f_2(\lambda,b)}\right]d\lambda \tag{3.12}$$

and

$$Z_{ik}^{(2)}=\int_A^\infty \mathcal{L}_{ik}\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\}\left[\frac{A(\lambda,b)}{f_1(\lambda,b)\cdot f_2(\lambda,b)}\right]d\lambda\,. \tag{3.13}$$

B. Numerical integration $\lambda\in[0,A]$

The first part of each element of the generalized impedance matrix given by

$$Z_{ik}^{(1)}=\int_0^A \mathcal{L}_{ik}\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\}\left[\frac{A(\lambda,b)}{f_1(\lambda,b)\cdot f_2(\lambda,b)}\right]d\lambda \tag{3.14}$$

is evaluated numerically. Here $\mathcal{L}_{ik}\{\ \}$ is a double-integral operator with the integrals being finite with respect to space coordinates. In this integral it has been found that the integral operator can make the integrand a slow varying function of λ and therefore it minimizes the error of integration. However, the opposite happens if $\mathcal{L}_{ik}$ operates on the infinite integral as shown below

$$Z_{ik}^{(1)}=\mathcal{L}_{ik}\left\{\int_0^A J_0(\lambda\,|\vec{r}-\vec{r}'|)\left[\frac{A(\lambda,b)}{f_1(\lambda,b)\cdot f_2(\lambda,b)}\right]d\lambda\right\}. \tag{3.15}$$

Therefore at first, for the evaluation of $Z_{ik}^{(1)}$, the computation of the space integrals in the operator is performed numerically using a Gaussian-Quadrature method with fixed points. For the integration with respect to λ, because of the existence of poles in the strip $k_0<\mathrm{Re}(\lambda)<k$, a further division of the integration interval into subintervals takes place as it is shown below:

(i) $0<\lambda<k_0$. The integration with respect to λ is performed numerically using a modified Romberg–Quadrature method.

(ii) $k_0<\lambda<k$. For the integration along this interval a singularity extraction technique is used (see Appendix A) which transforms the integral into a finite series plus an integral of a slowly varying function.[2,4] This finite series gives the contribution of the surface wave modes with the number of its terms dependent on the thickness of the dielectric as well as the dielectric constant.

(iii) $k<\lambda<A$. Numerical integration is invoked here in exactly the same manner as it is performed in the first subinterval.

C. Combination of numerical and analytic integration

In this case the integration with respect to λ is extended along the interval $[A,\infty]$ and the integrals which have to be computed are given by

$$Z_{ik}^{(2)}=\int_A^\infty \mathcal{L}_{ik}\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\}\left[\frac{A(\lambda,b)}{f_1(\lambda,b)\cdot f_2(\lambda,b)}\right]d\lambda\,. \tag{3.16}$$

When the source point and the observation point coincide, then $|r-r'|=0$ and the integral above takes the form

$$Z_{ik}^{(2)}=\int_A^\infty \frac{A(\lambda,b)}{f_1(\lambda,b)\cdot f_2(\lambda,b)}\,d\lambda\,. \tag{3.17}$$

For the integrand in (3.17) it can be proved that when $A \ll \lambda$

$$\frac{A(\lambda,b)}{f_1(\lambda,b)f_2(\lambda,b)} = O(\lambda^{-a}) \quad (a<1), \tag{3.18}$$

and therefore $Z_{ik}^{(2)}$ becomes infinite. In order to avoid this difficulty which arises in the computation of the matrix element, the dipole is assumed to have a very small but finite radius with the source current being along the axis of the dipole and the observation points being considered on its surface. As a result the distance between source and observation points can never be zero but the minimum value it can take is equal to the radius of the dipole. Since A has already been chosen in such a way that

$$\coth[(A^2-k^2)^{1/2}b] \doteq 1. \tag{3.19}$$

Equation (3.16) can be approximated as (see Appendix B)

$$\begin{aligned} Z_{ik}^{(2)} = &\left[\frac{2E_2}{\epsilon_r-1}(D_1-D_2)[1-e_1(A)]\right. \\ &+E_2\frac{\epsilon_r+1}{\epsilon_r-1}(D_2\epsilon_r-D_1)\left[1-\frac{\epsilon_r}{2(\epsilon_r-1)}e_2(A)\right] \\ &\times\mathscr{L}_{ik}\left\{C_1\int_A^\infty J_0(\lambda|\vec{r}-\vec{r}'|)e^{-uh}\frac{\lambda}{u}d\lambda\right. \\ &\left.\left.+C_2[1-e_3(A)]\int_A^\infty J_0(\lambda|\vec{r}-\vec{r}'|)\,d\lambda\right\}\right. \\ &+\text{Error}, \end{aligned} \tag{3.20}$$

where

$$e_1(A) = \frac{k^2-k_0^2}{4A^2-(3k^2+k_0^2)}, \tag{3.21}$$

$$e_2(A) = \frac{k^2-k_0^2}{A^2-k^2+(\epsilon_r/2(\epsilon_r+1))(k^2-k_0^2)}, \tag{3.22}$$

$$e_3(A) = \frac{1}{2}\frac{k^2}{A^2}. \tag{3.23}$$

The error made by the approximation in Eq. (3.20) depends on A. It has been found that if

$$A = O(10^{-2}), \tag{3.24}$$

then

$$\text{Error} = O(10^{-4}Z_{ik}^{(2)}) \tag{3.25}$$

and therefore the approximation considered here is a very good one since the overall error made in the computation of the input impedance is of the order of 0.1%. In Eq. (3.20) the integrals can be written as

$$\begin{aligned} I_1\int_A^\infty &J_0(\lambda|\vec{r}-\vec{r}'|)e^{-uh}\frac{\lambda}{u}d\lambda \\ &=\frac{1}{(h^2+|\vec{r}-\vec{r}'|^2(1+e_3(A))^2)^{1/2}} \\ &\quad-\int_0^{(A^2-k^2)^{1/2}}J_0[t\rho(1+e_3(A))]e^{-th}dt+\mathscr{R}_3, \end{aligned} \tag{3.26}$$

with

$$\mathscr{R}_3 = O(10^{-4}I_1) \tag{3.27}$$

and

$$\begin{aligned} \int_A^\infty &J_0(\lambda|\vec{r}-\vec{r}'|)\,d\lambda \\ &=\frac{1}{|\vec{r}-\vec{r}'|}-\left\{AJ_0(A|\vec{r}-\vec{r}'|)\right. \\ &\quad+\frac{\pi A}{2}\left[J_1(A|\vec{r}-\vec{r}'|)\mathscr{H}_0(A|\vec{r}-\vec{r}'|)\right. \\ &\quad\left.\left.-J_0(A|\vec{r}-\vec{r}'|)\mathscr{H}_1(A|\vec{r}-\vec{r}'|)\right]\right\}. \end{aligned} \tag{3.28}$$

By substituting equations above into (3.20) and by ignoring the error terms, $Z_{ik}^{(2)}$ becomes

$$\begin{aligned} Z_{ik}^{(2)} = &\left[\frac{2E_2}{\epsilon_r-1}(D_1-D_2)[1-e_1(A)]\right. \\ &\left.+E_2\frac{\epsilon_r+1}{\epsilon_r-1}(D_2\epsilon_r-D_1)\left[1-\frac{\epsilon_r}{2(\epsilon_r-1)}e_2(A)\right]\right] \\ &\mathscr{L}_{ik}\{F(|r-r'|)-\phi(|r-r'|)\}, \end{aligned} \tag{3.29}$$

where $F(|\vec{r}-\vec{r}'|)$ is a fast varying function of $|\vec{r}-\vec{r}'|$ given by

$$\begin{aligned} &F(|\vec{r}-\vec{r}'|) \\ &=\frac{C_1}{1+e_3(A)}\frac{1}{(|\vec{r}-\vec{r}'|^2+h^2/(1+e_3(A))^2)^{1/2}} \\ &\quad+C_2[1-e_3(A)]\frac{1}{|\vec{r}-\vec{r}'|}, \end{aligned} \tag{3.30}$$

while $\phi(|\vec{r}-\vec{r}'|)$ is a slow varying function of $|\vec{r}-\vec{r}'|$ and is given by

$$\begin{aligned} \phi(|\vec{r}-\vec{r}'|) = &C_1\int_0^{(A^2-k^2)^{1/2}}J_0[t\rho(1+e_3(A))]e^{-th}dt \\ &+C_2[1-e_3(A)]\left\{AJ_0(A|\vec{r}-\vec{r}'|)\right. \\ &+\frac{\pi A}{2}\left[J_1(A|\vec{r}-\vec{r}'|)\mathscr{H}_0(A|\vec{r}-\vec{r}'|)\right. \\ &\left.\left.-J_0(A|\vec{r}-\vec{r}'|)\mathscr{H}_1(A|\vec{r}-\vec{r}'|)\right]\right\}. \end{aligned} \tag{3.31}$$

Considering the result of the operation of $\mathscr{L}_{ik}$ on the function $F(|\vec{r}-\vec{r}'|)$, $Z_{ik}^{(2)}$ can be written in a final form as

$$\begin{aligned} Z_{ik}^{(2)} = &\left[\frac{2E_2}{\epsilon_{r-1}}(D_1-D_2)[1-e_1(A)]\right. \\ &\left.+E_2\frac{\epsilon_{r+1}}{\epsilon_{r-1}}(D_2\epsilon_r-D_1)\left[1-\frac{\epsilon_r}{2(\epsilon_{r+1})}e_2(A)\right]\right] \\ &\times\left[\sum_{n=-1,0}\sum_{m=-1,0,1}A_{nm}\{\tau\ln((\sigma^2+\tau^2)^{1/2}-\tau)\right. \\ &\left.+(\sigma^2+\tau^2)^{1/2}\}\Big|_{\substack{\tau=\tau_1\\ \tau=\tau_0}}\right. \\ &\left.+\mathscr{L}_{ik}\{\mathscr{R}(|\vec{r}-\vec{r}'|)-\phi(|\vec{r}-\vec{r}'|)\}\right], \end{aligned} \tag{3.32}$$

where

$\tau_1 = x_i - x_k + (n-m)l_x$,

$\tau_0 = x_i - x_k + (n-m-1)l_x$,

l_x = the length of the subinterval considered on the dipole,

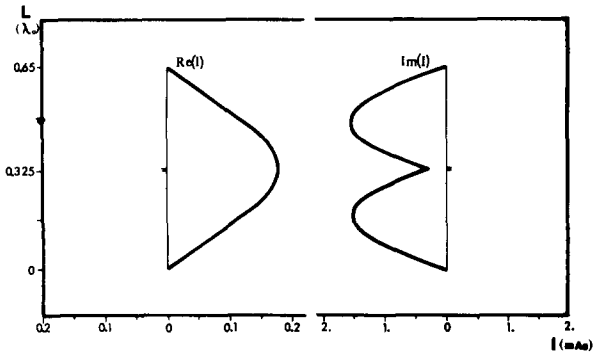


FIG. 4. Current distribution for a printed dipole with $\epsilon_r = 2.35$, $b = 0.1016\lambda_0$, $L = 0.65\lambda_0$.

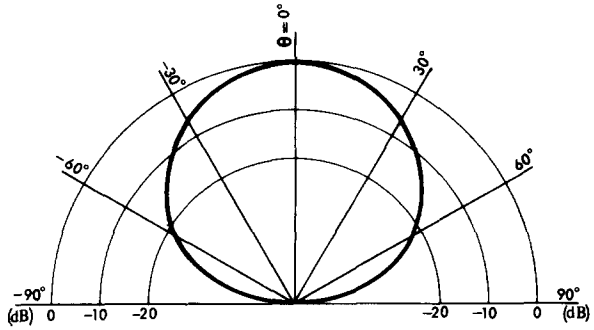


FIG. 6. E-plane normalized power pattern for a printed dipole with $\epsilon_r = 2.35$, $b = 0.1016\lambda_0$ and $L = 0.65\lambda_0$.

and

$\mathscr{R}(|\vec{r} - \vec{r}'|) - \phi(|\vec{r} - \vec{r}'|)$ is a slow-varying function of $|\vec{r} - \vec{r}'|$.

In Eq. (3.32) the term $\mathscr{L}_{ik}\{\ \}$ is evaluated numerically by using the Gaussian-Quadrature method of integration.

IV. AN APPLICATION AND CONCLUSIONS

In the attempt to obtain the current distribution of extended sources of electromagnetic radiation with dipole moments on the interface between two electrically dissimilar materials, the kernel of the resulting Fredholm integral equation of the first kind involves Sommerfeld-type integrals. These integrals include all the physical phenomena of the problem, i.e., radiation into space, evanescent waves as well as surface modes in the substrate. In addition, because of the fact that source and field points must be selected on the substrate surface to solve the integral equation for the current distribution, the Sommerfeld-type integrals converge very slowly and special techniques are necessary for their proper evaluation. This is the crucial contribution of this paper, i.e., a systematic methodology is provided to solve these integrals for sources on dielectric interfaces and subsequently to obtain the current distribution of such extended sources. Previously,[5–8] sources of electromagnetic radiation with a presumed current distribution were selected off the interface, introducing in this manner an exponential decay factor in the integrand of the resulting Sommerfeld-type integrals.

In the process of solving the Fredholm integral equation, a basis set of piece-wise continuous functions has been adopted in the expansion of the unknown current distribution. This device is dictated by two factors: (a) it simplifies the kernel after a series of closed form integrations by parts and (b) it satisfies the boundary conditions of the current distribution on the antenna.[17,18] Finally in order to illustrate the utility of solutions of Eq. (1) for printed circuit antennas, the case of lineal printed dipoles is considered. The current distribution is shown in Fig. 4 for a center-fed, by a unit voltage source, dipole of radius $10^{-4}\lambda_0$ and length $L = 0.65\lambda_0$. The substrate properties are $\epsilon_r = 2.35$ and thickness $b = 0.1016\lambda_0$. The input impedance for this printed circuit dipole is also shown in Fig. 5 and its radiation pattern in Fig. 6.

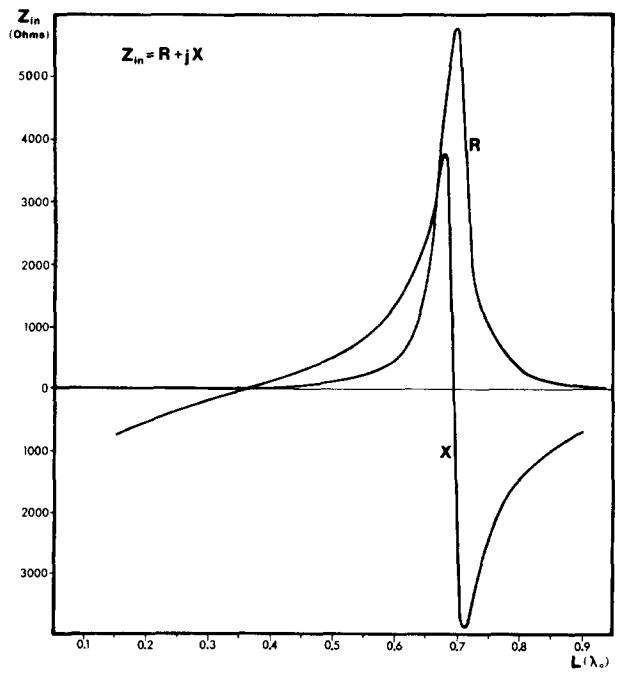


FIG. 5. Input impedance vs dipole length for $\epsilon_r = 2.35$ and $b = 0.1016\lambda_0$.

APPENDIX A: TECHNIQUE FOR THE EXTRACTION OF THE SINGULARITIES

It has been mentioned that the interval $[k_0, k]$ includes a finite set of discrete essential singularities which correspond to surface waves and contribute to the integration along this interval. Their contribution affects seriously the input impedance and radiation characteristics of the printed antenna. The technique which is described below transforms the integral given by

$$I_{ik} = \int_{k_0}^{k} \mathscr{L}_{ik}\{J_0(\lambda|\vec{r} - \vec{r}'|)\}\left[\frac{A(\lambda,b)}{f_1(\lambda,b)\cdot f_2(\lambda,b)}\right] d\lambda \quad \text{(A1)}$$

into a finite series plus an integral of a slowly varying function.

If it is assumed that S is the set of the singularities mentioned above, then S is given by

$$S=\{x_i/[x_i=\text{zero of } f_1(\lambda,b)]\,U\,[x_i=\text{zero of } f_2(\lambda,b)]\;; \quad i=1,2,3,\cdots,N\}\,. \tag{A2}$$

A partition $P[t_{n+1}]$ of the interval $[k_0,k]$ is considered such that

$$P[t_{N+1}]=\{t_0=k_0,\ t_1<x_1,\ldots,t_n<x_N,\ t_{N+1}=k\}\,, \tag{A3}$$

and the integral (A1) can be written as

$$I_{ik}=\sum_{r=0}^{N}\int_{t_r}^{t_{r+1}}d\lambda\,\frac{\phi_r^{ik}(\lambda,b,\epsilon_r)}{\lambda-x_r} \tag{A4}$$

with $\phi_r^{ik}(\lambda,b,\epsilon_r)$ given by

$$\phi_r^{ik}(\lambda,b,\epsilon_r)=\mathscr{L}_{ik}\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\}\left[\frac{A(\lambda,b)(\lambda-x_r)}{f_1(\lambda,b)\cdot f_2(\lambda,b)}\right]. \tag{A5}$$

In (A5) the quantity $(\phi_r^{ik}(x_r,b,\epsilon_r)/(\lambda-x_r))$ is added and subtracted giving

$$I_{ik}=\sum_{r=0}^{N}\int_{t_r}^{t_{r+1}}d\lambda\,\frac{\phi_r^{ik}(\lambda,b,\epsilon_r)-\phi_r^{ik}(x_r,b,\epsilon_r)}{\lambda-x_r}+\sum_{r=0}^{N}\phi_r^{ik}(x_r,b,\epsilon_r)\left[\ln\left|\frac{t_{r+1}-x_r}{x_r-t_r}\right|-j\pi\right].$$

In the equation above, the integrands are slowly varying functions in the intervals $[t_r,t_{r+1}]$ and the integrals can be evaluated using a Gaussian-Quadrature integration method with four fixed points.

APPENDIX B: FORMULATION OF THE INTEGRAL FOR THE TAIL CONTRIBUTION

If we choose A in such a way that

$$\coth[(A^2-k^2)^{1/2}b]\doteq 1\,, \tag{B1}$$

the integral given below

$$Z_{ik}^{(2)}=\int_A^{\infty}\mathscr{L}_{ik}\{J_0(\lambda\,|\vec{r}-\vec{r}'|)\}\left[\frac{A(\lambda,b)}{f_1(\lambda,b)\cdot f(\lambda,b)}\right]d\lambda \tag{B2}$$

can be approximated as follows

$$Z_{ik}^{(2)}=\mathscr{L}_{ik}\left\{\int_A^{\infty}J_0(\lambda\,|\vec{r}-\vec{r}'|)\left[\frac{C(\lambda,b)}{g_1(\lambda,b)\cdot g_2(\lambda,b)}\right]d\lambda\right\}, \tag{B3}$$

where

$$g_1(\lambda,b)=u_0+u\,, \tag{B4}$$

$$g_2(\lambda,b)=\epsilon_r u_0+u\,, \tag{B5}$$

and

$$C(\lambda,b)=\left[C_1\frac{\sinh[u(b-h)]}{\sinh(ub)}+C_2\right](D_1u_0+D_2u)E_2\,. \tag{B6}$$

If one considers that

$$\sinh[u(b-h)]=\sinh(ub)\cosh(uh)-\cosh(ub)\sinh(uh)\,,$$

then

$$C(\lambda,b)=(C_1e^{-uh}+C_2)(D_1u_0+D_2)E_2\lambda\,, \tag{B7}$$

and since the relation

$$\frac{D_1u_0+D_2u}{u+u_0}\cdot\frac{\lambda}{\epsilon_r u_0+u}=\frac{\lambda}{\epsilon_{r-1}}\left\{\frac{D_1-D_2}{u+u_0}+\frac{D_2\epsilon_r-D_1}{\epsilon_r u_0+u}\right\} \tag{B8}$$

is true Eq. (B3) can be written as

$$Z_{ik}^{(2)}=\mathscr{L}_{ik}\left\{\int_A^{\infty}J_0(\lambda\,|\vec{r}-\vec{r}'|)(C_1e^{-uh}+C_2)\frac{E_2}{\epsilon_r-1}\times\left(\frac{D_1-D_2}{1+(u_0/u)}+\frac{D_2\epsilon_r-D_1}{1+\epsilon_r(u_0/u)}\right)\frac{\lambda}{u}\,d\lambda\right\}, \tag{B9}$$

where

$$u=[\lambda^2-k^2]^{1/2}\,, \tag{B10}$$

$$u_0=[\lambda^2-k_0^2]^{1/2}\,. \tag{B11}$$

Because of (B10) and (B11) the functions $1+u_0/u$ and $1+\epsilon_r u_0/u$ in (B9) can be written as

$$1+\frac{u_0}{u}=2\{1-e_1(\lambda)\}^{-1} \tag{B12}$$

and

$$1+\epsilon_r\frac{u_0}{u}=(1+\epsilon_r)\left\{1-\frac{\epsilon_r}{2(\epsilon_r+1)}e_2(\lambda)\right\}^{-1}, \tag{B13}$$

where

$$0\leqslant e_1(\lambda)\leqslant\frac{k^2-k_0^2}{4A^2-3k^2-k_0^2}\,,$$

$$0\leqslant e_2(\lambda)\leqslant\frac{k^2-k_0^2}{A^2-k^2+(\epsilon_r/2(\epsilon_r+1))(k^2-k_0^2)} \tag{B14}$$

with

$$\lambda\epsilon[A,\infty)\,.$$

By substituting (B12) and (B13) into (B9) after some manipulations $Z_{ik}^{(2)}$ can be put in the final form

$$\begin{aligned}Z_{ik}^{(2)}=&\frac{2E_2}{\epsilon_r-1}(D_1-D_2)[1-e_1(A)]\\&\times\mathscr{L}_{ik}\left\{C_1\int_A^{\infty}J_0(\lambda\,|\vec{r}-\vec{r}'|)e^{-uh}\frac{\lambda}{u}\,d\lambda+C_2[1-e_3(A)]\int_A^{\infty}J_0(\lambda\,|\vec{r}-\vec{r}'|)\,d\lambda\right\}+\tilde{\mathscr{R}}_1\\&+E_2\frac{\epsilon_r+1}{\epsilon_r-1}(D_2\epsilon_r-D_1)\left[1-\frac{\epsilon_r}{2(\epsilon_r+1)}e_2(A)\right]\\&\times\mathscr{L}_{ik}\left\{C_1\int_A^{\infty}J_0(\lambda\,|\vec{r}-\vec{r}'|)e^{-uh}\frac{\lambda}{u}\,d\lambda\right\}\\&+\frac{\epsilon_r}{2(\epsilon_r+1)}\tilde{\mathscr{R}}_2\,,\end{aligned} \tag{B15}$$

with

$$\begin{aligned}|\tilde{\mathscr{R}}_1|<&A^{1/2}\frac{k^2-k_0^2}{A^2-k^2}\cdot\left[\frac{2E_2}{\epsilon_r-1}(D_1-D_2)[1-e_1(A)]\right.\\&+\mathscr{L}_{ik}\left\{C_1\int_A^{\infty}J_0(\lambda\,|\vec{r}-\vec{r}'|)e^{-uh}\frac{\lambda}{u}\,d\lambda\right.\\&\left.\left.+C_2[1-e_3(A)]\int_A^{\infty}J_0(\lambda\,|\vec{r}-\vec{r}'|)\,d\lambda\right\}\right]\end{aligned} \tag{B16}$$

and

$$|\tilde{\mathcal{R}}_2| < A^{1/2} \frac{k^2 - k_0^2}{A^2 - k^2} \left[E_2 \frac{\epsilon_r + 1}{\epsilon_r - 1} (D_2 \epsilon_r - D_1) \times \left[1 - \frac{\epsilon_r}{2(\epsilon_r + 1)} e_2(A) \right] \times \mathcal{L}_{ik} \left\{ C_1 \int_A^\infty J_0(\lambda |\vec{r} - \vec{r}'|) e^{-uh} \frac{\lambda}{u} d\lambda + C_2 [1 - e_3(A)] \int_A^\infty J_0(\lambda |\vec{r} - \vec{r}'|) d\lambda \right\} \right]. \quad \text{(B17)}$$

From (B16) and (B17) it can be shown that

$$\left| \tilde{\mathcal{R}}_1 + \frac{\epsilon_r}{2(\epsilon_r + 1)} \tilde{\mathcal{R}}_2 \right| < \frac{k^2 - k_0^2}{A^2 - k^2} Z_{ik}^{(2)} \times \frac{A^{1/2}}{2}.$$

With A being of the order of 100 the error made by ignoring $\left[\tilde{\mathcal{R}}_1 + \frac{\epsilon_r}{2(\epsilon_r + 1)} \tilde{\mathcal{R}}_2 \right]$ is given by

$$\text{Error} = O[10^{-4} \times Z_{ik}].$$

This approximation gives an error 0.1% in the input impedance.

[1] A. Sommerfeld, *Partial Differential Equations in Physics* (Academic, New York, 1949), Vol. VI.

[2] N. K. Uzunoglu, N. G. Alexopoulos, and J. G. Fikioris, "Radiation Properties of Microstrip Dipoles," IEEE Trans. Antennas Propag. **AP-27**, 853–858 (1979).

[3] I. E. Rana and N. G. Alexopoulos, "Current Distribution and Input Impedance of Printed Dipoles," IEEE Trans. Antennas Propag. **AP-29**, 99–105 (1981).

[4] P. B. Katehi, M. S. Thesis, Dept. Elect. Engr., University of California, Los Angeles, CA 90024 (1981).

[5] E. K. Miller, A. J. Poggio, G. J. Burke, and E. S. Selden, "Analysis of Wire Antennas in the Presence of a Conducting Half-space, Part II, Horizontal Antenna in Free Space," Can. J. Phys. **50**, 2614–2627 (1972).

[6] J. N. Brittingham, E. K. Miller, and J. T. Okada, "A Bivariate Interpolation Approach of Efficiently and Accurately Modeling of Antennas Near a Half Space," Electr. Lett. **13**, 690–691 (1977).

[7] Y. Rahmat-Samii, P. Parhami, and R. Mittra, "Transient Response of Loaded Horizontal Antenna over an Imperfect Ground," IEEE Trans. Antennas Propag. **AP-26**, 789–796 (1978).

[8] P. Parhami, Y. Rahmat-Samii, and R. Mittra, "An Efficient Approach for Evaluating Sommerfeld Integrals Encountered in the Problem of a Current Element radiating over a Lossy Ground," IEEE Trans. Antennas Propag. **AP-28**, 100–104 (1980).

[9] K. R. Carver and J. W. Mink, "Microstrip Antenna Technology," IEEE Trans. Antennas Propag. **AP-29**, 2–24 (1981).

[10] R. J. Mailloux, J. G. McIlvenna, and N. P. Kernweis, "Microstrip Array Technology," IEEE Trans. Antennas Propag. **AP-29**, 25–37 (1981).

[11] F. Mizuno, Y. Daiku, and S. Ono, "Design of Printed Resonant Antennas for Monolithic-Diode Detectors," IEEE Trans. Microwave Theory Tech. **MIT-25**, 470–476 (1977).

[12] D. B. Rutledge, S. E. Schwarz, and A. T. Adams, "Infrared and Submillimeter Antennas," Infrared Phys. **18**, 713–729 (1978).

[13] P. Hall *et al.*, "Feasibility of Designing Millimeter Microstrip Antenna Arrays," in *Proc. of Agard Conf. on Millimeter and Sub-Millimeter Wave Propagation and Circuitry*, **245** (1978).

[14] D. B. Rutledge, S. E. Schwartz, T. L. Hwang, D. J. Angelakos, K. K. Mei, and S. Ykota, "Antennas and Waveguides for Far-Infrared Integrated Circuits," IEEE J. Quantum Electron. **QE-16**, 508–516 (1980).

[15] R. S. Elliott, *Antenna Theory and Design* (Prentice-Hall, Englewood Cliffs, NJ, 1981).

[16] R. F. Harrington, *Field Computation by Moment Methods* (MacMillan, New York, 1968).

[17] T. H. Richmond, "A Wire Grid Model for Scattering by Conducting Bodies," IEEE Trans. Antennas Propag. **AP-14**, 782–786 (1966).

[18] W. A. Imbriale and P. G. Ingerson, "On Numerical Convergence of Moment Method Solution of Moderately Thick Wire Antennas Using Sinusoidal Basis Functions," IEEE Trans. Antennas Propag. **AP-21**, 363–366 (1973).

A generalized Weyl correspondence. II. Some general results

M. Gadella
Facultad de Ciencias, Universidad de Santander, Av. los Castros s/n, Santander, Spain

(Received 4 November 1981; accepted for publication 24 September 1982)

After defining a generalized Weyl correspondence, we give some general results. These are presented as comments on a theorem. They mainly refer to the finite-dimensional, unbounded, and non-self-adjoint cases.

PACS numbers: 03.65. — w

I. INTRODUCTION

A generalized Weyl correspondence (GWC) has been explicitly defined in a recent paper.[1] Along with the definitions the authors gave some generalized results which were labeled under Theorem I and Theorem II. The aim of the present paper is to provide a few results which go a little further beyond those contained in both theorems. Here we present our results as comments on Theorem II.

The definition of the GWC is as follows: Let $(X_j)_{j\in J}$ be a set of real random variables and $(\hat{x}_j)_{j\in J}$ a set of linear operators on a separable Hilbert space $\mathscr{H}$, where J is an arbitrary index set. If there exists a trace class operator ρ such that for any finite set of indices $\{j_1,\ldots j_n\}\subset J$ and any real numbers $\alpha_{j_1},\ldots,\alpha_{j_n}$

$$E\{\exp[i(\alpha_{j_1}X_{j_1}+\cdots+\alpha_{j_n}X_{j_n})]\} = \mathrm{tr}\{\rho\exp[i(\alpha_{j_1}\hat{x}_{j_1}+\cdots+\alpha_{j_n}\hat{x}_{j_n})]\} \tag{1}$$

holds, then we say that there exists a GWC between the random variables and the operators. (E here means expectation value). Each of the random variables acts on a sample space Ω which is the cartesian product of the ranges of the variables $(X_j)_{j\in J}$. Ω is endowed with a finite measure μ such that $\mu(\Omega)<\infty$ [usually $\mu(\Omega)=1$], but μ is not necessarily nonnegative [i.e., there may exist a measurable set $A\subset\Omega$ such that $\mu(A)<0$]. X_j is the projection of Ω on its jth coordinate.

From now on we shall call basic variables to the random variables $(X_j)_{j\in J}$ and basic operators to the $(\hat{x}_j)_{j\in J}$.

The GWC may be used to realize a classical stochastic picture of a nonrelativistic quantum system[2] (direct problem). Then each of the operators $\hat{x}_j$ is a basic observable on the description of the system and ρ one of its possible states (ρ is here usually taken positive and with trace one). $\hat{x}_j$ is in correspondence with the variable X_j all $j\in J$ and ρ with a signed measure μ on Ω such that $1=\mathrm{tr}\rho=\mu(\Omega)$.

Also it can be used to represent algebras of random variables, which are commutative algebras of linear operators on a Hilbert space, which are generally noncommutative[3,4] (inverse problem).

Finally we might use the GWC to compare a classical stochastic theory of matter with the quantum one in some particular problems.[5,6]

Other correspondences between random variables and linear operators can be defined.[7] We have chosen the GWC because it generalizes the classical Weyl correspondence,[8,9] which is the most natural one in the context of nonrelativistic problems.[10]

Here we briefly recall Theorem II of Ref. 1. We shall see later some conditions under which we may weaken the statements of Theorem II, obtaining the same results. From them we shall obtain some interesting general results. Theorem II establishes the following[11]:

Theorem II: Let there exist a solution of the GWC fulfilling:

(1) The number of elements in J is finite.

(2) ρ is a positive trace class operator.

(3) The basic operators are self-adjoint.

(4) All the basic variables and the basic operators are bounded.

(5) Each of the basic variables has infinite range.

We call A and B respectively the linear algebras with identity spanned by the basic variables and operators. Then:

(a) There exists a linear mapping g between A and B considered as linear spaces such that

$$g(X_j)=\hat{x}_j,\quad j\in J. \tag{2}$$

(b) (1) is equivalent to

$$\mathrm{tr}[\rho g(y)]=E(y),\quad y\in A. \tag{3}$$

(c) If $y=X_1^{k_1}\cdots X_N^{k_N}$,

$$g(y)=(\hat{x}_1^{k_1}\cdots\hat{x}_N^{k_n})_s, \tag{4}$$

where $(\cdots)_s$ denotes the symmetrized product of the operators within the bracket.

II. VARIABLES WITH FINITE RANGE

The main problem we have when we deal with variables of finite range is that g defined in (2) may not exist. For example, let us assume that we want to make a classical stochastic picture of a spin-$\frac{1}{2}$ system. Then we may think that we can use the GWC where the basic operators $\hat{x}_j$ are to be the Pauli matrices and each of the basic variables X_j the jth coordinate of $\Omega=\{1,-1\}\times\{1,-1\}\times\{1,-1\}$ ($j\in J=\{1,2,3\}$). Observe that Ω must be this set precisely if we want that $g(X_j^k)=\hat{x}_j^k$, all k. The basic variables and operators are bounded here; hence we can conclude that[1]

$$g\{(\alpha X_1+\beta X_2+\gamma X_3)^n\} = (\alpha\sigma_x+\beta\sigma_y+\gamma\sigma_z)^n,\quad \text{all } n, \tag{5}$$

where α, β, and γ are any real numbers. For the sake of simplicity let us make $\gamma=0$. Then, if $n=3$,

$$g\{(\alpha X_1+\beta X_2)^3\} = (\alpha\sigma_x+\beta\sigma_y)^3 = (\alpha+\beta)\alpha\sigma_x+(\alpha+\beta)\beta\sigma_y. \tag{6}$$

 0022-2488/83/030534-05$02.50

On the other hand,

$$(\alpha X_1 + \beta X_2)^3 = (\alpha^3 + 3\alpha\beta^2)X_1 + (\beta^3 + 3\alpha^2\beta)X_2. \quad (7)$$

Whence

$$g\{(\alpha X_1 + \beta X_2)^3\} = (\alpha^3 + 3\alpha\beta^2)\sigma_x + (\beta^3 + 3\alpha^2\beta)\sigma_y, \quad (8)$$

which is clearly different from (6). Then the question that arises is: What are the conditions a finite-dimensional GWC has to fulfill in order that g be well defined?[12] An answer is contained in the next result. We shall label it Theorem III, after Theorem I and Theorem II of Ref. 1.

Theorem III: Assume we have a finite-dimensional GWC and $\rho \geqslant 0$ or $\rho \leqslant 0$.

(1) Suppose $N = 2$ and that for some pair (α_1^0, α_2^0) all the elements in the diagonal of ρ are different from zero in all the basis which diagonalize $\alpha_1^0 \hat{x}_1 + \alpha_2^0 \hat{x}_2$. Then g is a mapping if and only if $[x_1, x_2] = 0$.

(2) If $\rho \geqslant 0$ ($\rho \leqslant 0$) and for all the basic diagonalizing the operators $\hat{x}_2, \hat{x}_3, \ldots, \hat{x}_N$, the elements in the diagonal of ρ are different from zero, then g is a mapping if and only if the algebra B spanned by the basic operators is commutative.

Before we give the proof of the theorem, it will be convenient to write the following lemma:

Lemma: Let $H = \{a_1, \ldots, a_n\}$ be a finite set of real numbers different from zero and each other. Then, if there is a set of numbers $G = \{b_1, \ldots, b_n\}$ such that

$$\sum_{i=1}^{n} b_i a_i^k = 0, \quad k = 1, \ldots, n+1. \quad (9)$$

Then $b_i = 0$, $i = 1, \ldots, n$.

Proof: Consider the algebra with identity $\mathscr{A}$ consisting of all the functions of the elements of H. $\mathscr{A}$ is spanned by the functions of the form:

$$f_k(\alpha_i) = \alpha_i^k, \quad k = 1, \ldots, n+1, \quad (10)$$

by the Stone–Weiesstrass theorem.[13] (Note that since f_{n+1} is a linear combination of $f_n, \ldots, f_1, 1$,[4] then the identity is a linear combination of $f_{n+1}, f_n, \ldots, f_1$). The finite sequence of numbers $(b_1, \ldots, b_n)$ can be looked at as a functional on the f_k ($k = 1, \ldots, n+1$) and can be extended to all of $\mathscr{A}$ by linearity. But the f_k span $\mathscr{A}$. This means that $b_i = 0$, all $i = 1, \ldots, n$.

III. PROOF OF THE THEOREM

Proof 1: Assume that g is a mapping between the algebras A and B. Then

$$\mathrm{tr}[\rho(\alpha_1 x_1 + \alpha_2 x_2)^n] = E(\alpha_1 X_1 + \alpha_2 X_2)^n, \quad \text{all } n, \quad (11)$$

is equivalent to (1).[1] The first term in (11), in a basis which diagonalizes $\alpha_1^0 \hat{x}_1 + \alpha_2^0 \hat{x}_2$, is

$$\mathrm{tr}\begin{vmatrix} \rho_{11} & \cdots & \rho_{k1} \\ \vdots & & \vdots \\ \rho_{1k} & \cdots & \rho_{kk} \end{vmatrix} \begin{vmatrix} a_1^n & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & a_k^n \end{vmatrix}, \quad (12)$$

where $a_1, \ldots, a_k$ are the eigenvalues of $\alpha_1^0 \hat{x}_1 + \alpha_2^0 \hat{x}_2$, which obviously depend on α_1^0 and α_2^0. We may suppose that in this basis $\rho_{ii} \neq 0$, all $i = 1, \ldots, k$. Then

$$\begin{aligned}(12) &= \rho_{11} a_1^n + \cdots + \rho_{kk} a_k^n \\ &= p_1(\alpha_1^0 x_{11} + \alpha_2^0 x_{21})^n + \cdots \\ &\quad + p_m(\alpha_1^0 x_{1m} + \alpha_2^0 x_{2m})^n, \end{aligned} \quad (13)$$

where x_{1j} and x_{2j}, respectively, are the values of the variables X_1 and X_2 at the point $\omega_j \in \Omega$ and p_j its probability.

Now we apply the lemma to conclude that if some a_i were different from all the numbers $(\alpha_1^0 x_{1s} + \alpha_2^0 x_{2s'})$, then $\rho_{ii} = 0$. By hypothesis $\rho_{ii} \neq 0$ ($i = 1, \ldots, k$); hence every a_i must be equal to some element of the form $(\alpha_1^0 x_{1s} + \alpha_2^0 x_{2s'})$. [Note that $p_i = 0$ for at least $n - (k+1)$ points in Ω.]

The lemma also says that $\rho_{ii} = \rho_{ii}(\alpha_1^0, \alpha_2^0)$ must be equal to some p_i. Then the following question arises: Is this true for any other values of α_1 and α_2?

For the sake of simplicity, we restrict ourselves to the case $\alpha_1, \alpha_2 \geqslant 0$ and α_1, α_2, not both zero. Take $\lambda = \alpha_1/(\alpha_1 + \alpha_2)$; then $\lambda \in [0,1]$, whence

$$\alpha_1 \hat{x}_1 + \alpha_2 \hat{x}_2 = (\alpha_1 + \alpha_2)[\lambda \hat{x}_1 + (1 - \lambda)\hat{x}_2]. \quad (14)$$

Since $\hat{x}_1$ and $\hat{x}_2$ are Hermitian, there exists a neighborhood of $[0,1]$ in the complex plane in which the eigenvalues and eigenvectors of $\lambda \hat{x}_1 + (1 - \lambda)\hat{x}_2$ are analytic. If $\psi_j(\lambda)$ is the jth eigenvector of (14),

$$\rho_{jj}(\alpha_1, \alpha_2) = \langle \psi_j(\lambda), \rho \psi_j(\lambda) \rangle \quad (15)$$

is an analytic function of λ. For the initial values α_1^0 and α_2^0 of α_1 and α_2, $\rho_{ii} \neq 0$. Therefore, (15) is different from zero except for possibly finitely many values of λ in $[0,1]$.

Now assume that $\rho \geqslant 0$ (or $\rho \leqslant 0$). Equation (13) can be rewritten as

$$\sum_{j=1}^{k} \langle \psi_j(\lambda), \rho \psi_j(\lambda) \rangle a_j^n(\lambda) = \sum_{i=1}^{m} p_i(\lambda x_{1i} + (1 - \lambda)x_{2i})^n, \quad \lambda \in [0,1], \ n \in N, \quad (16)$$

where $\psi_j(\lambda)$ is the jth eigenvector of λ and $a_j(\lambda)$ its corresponding eigenvalue.

The positivity condition on ρ and the above discussion over analyticity lead to

$$\langle \psi_j(\lambda), \rho \psi_j(\lambda) \rangle > 0, \quad j = 1, \ldots, n, \quad (17)$$

except for possible finitely many values of λ.

Hence either

$$a_j(\lambda) = 0 \quad (18)$$

or these exists an index i (which may depend on j and λ) such that

$$a_j(\lambda) = \lambda x_{1i} + (1 - \lambda)x_{2i}. \quad (19)$$

(18) and (19) along with the analyticity of the eigenvalues on λ yield

$$\sigma(\lambda \hat{x}_1 + (1 - \lambda)\hat{x}_2) \subset \{\lambda \sigma(\hat{x}_1) + (1 - \lambda)\sigma(\hat{x}_2)\} \cup \{0\}. \quad (20)$$

Now we claim that $[\hat{x}_1, \hat{x}_2] = 0$. Figure 1 will help us to understand our reasoning. On the vertical axis crossing $\lambda = 0$, we have placed all the eigenvalues of $\hat{x}_2$ and 0. On that one crossing $\lambda = 1$, we have put the eigenvalues of $\hat{x}_1$ and 0. We have drawn all possible lines connecting the eigenvalues as well as that one joining the zeroes. For any λ_0, the eigenvalues of $\lambda_0 \hat{x}_1 + (1 - \lambda_0)\hat{x}_2$ lie on the vertical line which crosses λ_0 at the intersections with all the other lines. In fact, since $a_j(\lambda)$ are analytic in λ because of the Rellich theorem,[14] the eigenvalues of $\lambda \hat{x}_1 + (1 - \lambda)\hat{x}_2$ cannot jump from one line to the other, even at crossings. Therefore, the subindex i

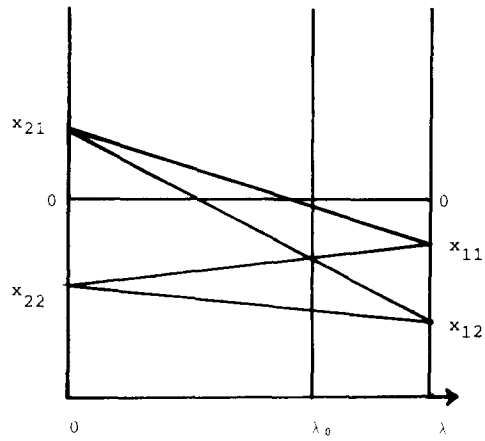


FIG. 1.

in (19) does not depend on λ. This implies that for every j there exist $x_{1j}\in\sigma(\hat{x}_1)$ and $x_{2j}\in\sigma(\hat{x}_2)$ such that

$$\forall\ \lambda:\quad a_j(\lambda)=\lambda x_{1j}+(1-\lambda)x_{2j}. \tag{21}$$

Take now

$$\hat{x}_1+\mu\hat{x}_2=(1+\mu)\left[\frac{1}{1+\mu}\hat{x}_1+\left(1-\frac{1}{1+\mu}\right)\hat{x}_2\right] \tag{22}$$

with $\mu\geqslant 0$. After (21), the eigenvalues of (22) are given by $x_{1j}+\mu x_{2j}$. Let $\psi_j(\mu)$ be the jth eigenvector of (22) and assume that we have ordered the j so that $x_{21}\geqslant x_{22}\geqslant\cdots\geqslant x_{2k}$. For $\psi_1(\mu)$ we can write

$$\langle\psi_1(0),(\hat{x}_1+\mu\hat{x}_2)\psi_1(\mu)\rangle=(x_{11}+\mu x_{21})\langle\psi_1(0),\psi_1(\mu)\rangle. \tag{23}$$

Using analyticity, we obtain

$$\begin{aligned}&\langle\psi_1(0),\hat{x}_1\psi_1(0)\rangle+\mu\langle\psi_1(0),\hat{x}_2\psi_1(0)\rangle\\&\quad+\mu\langle\psi_1(0),\hat{x}_1\psi_1'(0)\rangle+o(\mu^2)\\&=(x_{11}+\mu x_{21})\langle\psi_1(0),\psi_1(0)\rangle+\mu\langle\psi_1(0),\psi_1'(0)\rangle x_{11}\\&\quad+o(\mu^2).\end{aligned} \tag{24}$$

Hence

$$\begin{aligned}&\langle\psi_1(0),\hat{x}_2\psi_1'(0)\rangle+\langle\psi_1(0),\hat{x}_1\psi_1'(0)\rangle\\&\quad=x_{21}+\langle\psi_1(0),\psi_1'(0)\rangle x_{11}.\end{aligned} \tag{25}$$

Realizing that $\hat{x}_1$ is Hermitian and that $\hat{x}_1\psi_1(0)=x_{11}\psi_1(0)$, we finally conclude

$$\max[\sigma(\hat{x}_2)]=x_{21}=\langle\psi_1(0),\hat{x}_2\psi_1(0)\rangle. \tag{26}$$

This implies that

$$\hat{x}_2\psi_1(0)=x_{21}\psi_1(0). \tag{27}$$

In fact, if $(a_1,\ldots,a_k)$ were the components of $\psi_1(0)$ in a basis which diagonalizes $\hat{x}_2$,

$$\begin{aligned}\langle\psi_1(0),\hat{x}_2\psi_1(0)\rangle&=|a_1|^2x_{21}+|a_2|^2x_{22}+\cdots\\&\quad+|a_k|^2x_{2k}\leqslant x_{21},\end{aligned} \tag{28}$$

the last inequality being true because $\psi_1(0)$ has norm 1 and $x_{21}=\max[\sigma(\hat{x}_2)]$. Moreover, the equality is satisfied if and only if $|a_1|=1$ and $|a_2|=\cdots=|a_k|=0$. This concludes (27).

From (27) we can deduce that $\psi_1(0)$ is an eigenvector of $\hat{x}_1+\mu\hat{x}_2$ with eigenvalue $x_{11}+\mu x_{21}$. This is true whatever $\mu\geqslant 0$ we choose. Then

$$\mathbb{C}^k=\{\psi_1(0)\}\oplus\{\psi_1(0)\}^\perp, \tag{29}$$

where $\{\psi_1(0)\}$ is the subspace spanned by $\psi_1(0)$ and $\{\psi_1(0)\}^\perp$ its orthogonal. We may apply the same reasoning to $\{\psi_1(0)\}^\perp$ again so as to find a basis which simultaneously diagonalizes $\hat{x}_1$ and $\hat{x}_2$. This shows that $[\hat{x}_1,\hat{x}_2]=0$.

Proof 2: It is a corollary of part 1.

Remark I: We might have relinquished the condition on the sign of ρ (either $\rho\geqslant 0$ or $\rho\leqslant 0$). However, in this case we have to impose nondegeneracy conditions on the $\hat{x}_i$. Therefore, the above result can work if we substitute the restriction on ρ by the following: the $\hat{x}_i$ do not have degenerate eigenvalues.

Now we prove the converse. If all the basic operators commute all of them can be diagonalized in the same orthogonal basis (we are assuming all of them are Hermitian). In such a basis

$$\rho=\begin{vmatrix}\rho_{11}&\cdots&\rho_{1k}\\\vdots&&\vdots\\\rho_{k1}&\cdots&\rho_{kk}\end{vmatrix}. \tag{30}$$

If we call $x_{ij}, j=1,\ldots,k$, to the jth eigenvalue of the operator $\hat{x}_i$, then we have

$$\begin{aligned}\mathrm{tr}(\rho(\alpha_1\hat{x}_1+\cdots+\alpha_n\hat{x}_n)^l)&=\sum_{i=1}^{k}\rho_{ii}(\alpha_ix_{1i}+\cdots+\alpha_nx_{ni})^l\\&=\sum_{s=1}^{m}p_s(\alpha_1a_{1s}+\cdots+\alpha_na_{ns})^l,\quad\text{all } l,\end{aligned} \tag{31}$$

where $a_{ij}, j=1,\ldots,s$, is the jth value of the range of X_i. By the lemma we conclude that[15]

$$\begin{aligned}&(\alpha_1x_{1s}+\cdots+\alpha_nx_{ns})=(\alpha_1a_{1s}+\cdots+\alpha_na_{ns}),\\&\quad s=1,\ldots,r\leqslant\dim\mathscr{H}\end{aligned} \tag{32}$$

all $\alpha_1,\ldots,\alpha_n\in R$ and $p_{r+1}=\cdots=p_m=0$. On the other hand, it may also happen that $x_{1s}=x_{2s}=\cdots=x_{ns}=0$.

Hence[16]

$$\sigma(\alpha_1\hat{x}_1+\cdots+\alpha_n\hat{x}_n)\subset\mathrm{range}(\alpha_1X_1+\cdots+\alpha_nX_n)\cup\{0\}, \tag{33}$$

and, consequently, the minimal polynomial of $\alpha_1X_1+\cdots+\alpha_nX_n$ has rank bigger or equal of that one of $\alpha_1\hat{x}_1+\cdots+\alpha_n\hat{x}_n$. On the other hand, any $y\in A$ can be spanned as a linear combination of elements of the form $(\alpha_1X_1+\cdots+\alpha_nX_n)^l$.[1] Hence $g: A\to B$ is a mapping.

Remark II: When considering Pauli matrices again, the eigenvalues of $\alpha\sigma_x+\beta\sigma_y+\gamma\sigma_z$ are

$$\lambda=\pm\sqrt{\alpha^2+\beta^2+\gamma^2}. \tag{34}$$

Furthermore, σ_x, σ_y, and σ_z do not commute; therefore, we cannot expect g to be a mapping in this case. However, the result of Theorem III does not exclude the possibility of realizing classical stochastic models for a quantum spin system. In Ref. 2, we implement a stochastic model for the quantum Ising lattice.

IV. POSITIVITY OF ρ

In the proof of Theorem II we have not made use of the positivity of ρ. It is not necessary. Also it is not a necessary condition for the measure μ on Ω to be nonnegative as we can

see in the following example: Consider a k-dimensional GWC in which B is a commutative algebra, and assume that in the basis which diagonalizes the basic operators, the diagonal elements of ρ are all nonnegative. Then the measures of the points in Ω are all nonnegative, but ρ may be neither positive nor self-adjoint.

The fact that a state of nonrelativistic quantum mechanics is described by a positive trace class operator suggests that ρ should be such a kind of operator. In general we cannot guarantee the positivity of ρ even if μ is positive except for a few cases.[17] Nevertheless, a weaker result can be obtained.

Proposition: Assume we have the hypothesis of Theorem II except that we drop the requirement on the positivity of ρ. Then we can replace ρ by a Hermitian operator ρ' and leave the other elements of the correspondence unchanged.

Proof: We have

$$\mathrm{tr}\rho = (\mathrm{tr}\rho^+)^* \tag{35}$$

for any trace class operator. (35) and the properties of the trace yield

$$\begin{aligned}\mathrm{tr}\{\rho(\alpha_1\hat{x}_1 + \cdots + \alpha_n\hat{x}_n)^l\}^+ &= \mathrm{tr}[\rho^+(\alpha_1\hat{x}_1 + \cdots + \alpha_n\hat{x}_n)^l] \\ &= (\mathrm{tr}[\rho(\alpha_1\hat{x}_1 + \cdots + \alpha_n\hat{x}_n)^l])^* \\ &= E\{(\alpha_1X_1 + \cdots + \alpha_nX_n)^l\}, \qquad \text{all } l\in N. \end{aligned} \tag{36}$$

The last equality says that all terms in (36) are real, because of the reality of the basic variables. Hence

$$\begin{aligned}\mathrm{tr}[\rho^+(\alpha_1\hat{x}_1 + \cdots + \alpha_n\hat{x}_n)^l] &= \mathrm{tr}[\rho(\alpha_1\hat{x}_1 + \cdots + \alpha_n\hat{x}_n)^l] \\ &= \mathrm{tr}[\rho'(\alpha_1\hat{x}_1 + \cdots + \alpha_n\hat{x}_n)^l], \quad \text{all } l, \end{aligned} \tag{37}$$

where

$$\rho' = (\rho + \rho^+)/2. \tag{38}$$

Obviously, ρ' is Hermitian and satisfies our requirements.

V. BOUNDEDNESS

The condition of boundedness of the basic operators and variables is useful because it automatically gives that

$$\mathrm{tr}[\rho(\alpha_1\hat{x}_1 + \cdots + \alpha_n\hat{x}_n)^l] = E\{(\alpha_1X_1 + \cdots + \alpha_nX_n)^l\}, \quad \text{all } l\in N, \tag{39}$$

for any subset $\{1,\ldots,n\}\subset J$, is equivalent to (1). This relation allows us to define g. However, when some of the basic operators $\hat{x}_j$ are not bounded, the series

$$\sum_{n=0}^{\infty}\frac{i^n\hat{x}_j^n}{n!} \tag{40}$$

does not make sense, and the equivalence between (39) and (1) may not be true. Moreover, the relation (39) makes no sense when any of the involved variables has an infinite momentum. With this in mind, we can give a sufficient condition to have (39) in the case we have unbounded operators and variables.

Proposition: If all random variables have finite momenta (or equivalently if all linear combinations of them have finite momenta) and there exists an analytic vector[18] ψ of $\alpha_1\hat{x}_1 + \cdots + \alpha_n\hat{x}_n$ for any α_i, then (39) and (1) are equivalent with $\rho = |\psi\rangle\langle\psi|$.

The proof is straightforward.

Example: The vectors

$$\psi_n = (\sqrt{\pi}2^n n)^{-1/2}e^{-x^2/2}H_n(x), \tag{41}$$

where $H_n(x)$ are the Hermite polynomials, are analytic vectors of the real Lie algebra spanned by the position and the momentum operators.[19] We may define a GWC with these operators, the corresponding variables Q and P being defined on R. For all the vector states (41), both (39) and (1) are equivalent; however, only $\rho_0 = |\psi_0\rangle\langle\psi_0|$ will provide a positive density for Q and P.[17]

Remark: When more than one of the basic operators are unbounded, domain problems could appear. We have to be sure that they have a dense common domain and that their sum is self-adjoint in this domain, in order that the exponential in (1) be correctly defined via the Stone theorem.

VI. SELF-ADJOINTNESS OF THE BASIC OPERATORS

If some of the operators $\hat{x}_j$ are not self-adjoint, we can still define the GWC. But B will not be an involutive algebra. In most cases the basic observables are candidates to be the observables of our physical theory. Consequently, we would like them to be self-adjoint.

Nevertheless, we may have from the beginning a solution of the GWC in which some of the basic operators are not self-adjoint. Can we replace them for self-adjoint operators? The next result gives us a partial answer to this problem.

Theorem IV: Let there be a solution of (1) in which n of the basic operators are normal. Then they can be replaced by n Hermitian operators with no other changes, provided that the basic operators and variables are bounded.

Proof: Suppose that $n = 1$ first. Then by making all α_i zero except the one which corresponds to the normal operator $\hat{z}$:

$$E(e^{i\alpha X}) = \mathrm{tr}(\rho e^{i\alpha\hat{z}}). \tag{42}$$

This is equivalent to

$$\begin{aligned}\mathrm{tr}(\rho\hat{z}^n) &= \mathrm{tr}\left[\rho\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+iy)^n\,dE(x)\,dE(y)\right] \\ &= \int(x+iy)^n\,dG(x,y) = E(X^n) = \int_{-\infty}^{\infty}x^n\,dF(x), \end{aligned} \tag{43}$$

where $E(x,y) = E(x)E(y)$ is the spectral family associated with $\hat{z}$.[20] $F(x)$ is the distribution function of X and can be extended to a two-dimensional distribution $F(x,y)$, so that

$$\int(x+iy)^n\,dF(x,y) = \int(x+iy)^n\,dG(x,y), \quad \text{all } n. \tag{44}$$

Hence

$$G(x,y) = F(x,y)^{21} \tag{45}$$

so that

$$\mathrm{tr}[\rho E(x)] = \lim_{y\to\infty}\mathrm{tr}[\rho E(x)E(y)] = \lim_{y\to\infty}F(x,y) = F(x). \tag{46}$$

The operator

$$\hat{x} = \int x \, dE(x) \tag{47}$$

is self-adjoint and fulfills (46). But (46) is equivalent to (42), as we have proven in Ref. 1. Hence the theorem follows for $n = 1$.

Note that $\hat{x}$ is the Hermitian part of $\hat{z}$. This is the key point in the generalization to n. In fact, if we have n normal operators $\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_n$, which give us a solution of (1), we can use the above reasoning with $\hat{z} = \alpha_1 \hat{z}_1 + \cdots + \alpha_n \hat{z}_n$ and $X = \alpha_1 X_1 + \cdots + \alpha_n X_n$ to conclude that $\hat{z}$ can be replaced by its Hermitian part. This is $\alpha_1 \hat{x}_1 + \cdots + \alpha_n \hat{x}_n$, where $\hat{x}_i$ is the Hermitian part of $\hat{z}_i$. This proves the theorem.

ACKNOWLEDGMENT

The author is greatly indebted to the referee, who suggested to him the proof of the first part of Theorem III as here presented. It was wrong in the original draft of the present paper.

[1] M. Gadella and E. Santos, "A generalized Weyl correspondence," J. Math. Phys. **22**, 1651 (1981).
[2] See a particular example in M. Gadella, Nuovo Cimento B **63**, 510 (1981).
[3] E. Santos, J. Math. Phys. **15**, 1954 (1974).
[4] J. F. Cariñena, M. Gadella, and E. Santos, "An Algebra C^* of Observables," in *In Memoriam of A. Estevez* (Servicio de publicaciones del GIFT, Zaragoza, Spain, 1975).
[5] Reference 1, Application II.
[6] E. Santos, Nuovo Cimento B **22**, 201 (1977).
[7] G. S. Agarwal and E. Wolf, Phys. Rev. D **2**, 161 (1970).
[8] H. Weyl, *Theory of Groups and Quantum Mechanics* (Dover, New York, 1931).
[9] S. de Groot, *La Transformation de Weyl et la Fonction de Wigner: Une Forme Alternative de la Mécanique Quantique* (Université de Montreal, 1974).
[10] J. G. Krüger and A. Poffin, Physica A **85**, 84 (1976).
[11] We can drop the last condition of Theorem II in Ref. 1 because of the way in which the basic variables have been constructed. We also could exclude condition 1 because (1) always deals with a finite number of variables and we assume consistency. We maintain it for the sake of simplicity only.
[12] We say that the GWC is finite-dimensional if and only if the Hilbert space $\mathscr{H}$ where the basic operators are defined is finite-dimensional. If the basic variables have finite range, the basic operators must have finite spectrum; otherwise g will not be a mapping (because then $\hat{x}_j$ will not have a minimal polynomial of finite degree). If we keep J finite, $\mathscr{H}$ may be chosen finite-dimensional.
[13] W. Rudin, *Functional Analysis* (McGraw-Hill, New York, 1973), p. 115.
[14] M. Reed and B. Simon, *Analysis of Operators* (Academic, New York, 1978), Chap. XII.
[15] Note that this implies that $x_{is} = a_{is}$, $s = 1,\ldots,k = \dim\mathscr{H}$.
[16] This means that $\sigma(\alpha_1 \hat{x}_1 + \cdots + \alpha_n \hat{x}_n) = \sigma(\alpha_1 X_1 + \cdots + \alpha_n X_n)$. The last number has been defined in Ref. 1 as the elements in the range of $\alpha_1 X_1 + \cdots + \alpha_n X_n$ with nonzero probability. If in the range of every random variable, we neglect the values with zero probability, then g is indeed a bijection.
[17] R. L. Hudson, Rep. Math. Phys. **6**, 1249 (1974); also Ref. 9.
[18] For the definition of analytic vectors, see, for instance, B. Simon, "Topics in Functional Analysis," in *Mathematics of Contemporary Physics*, edited by R. F. Streater (Academic, London, 1972), p. 24.
[19] A. O. Barut and R. Rayczka, *Theory of Group Representations and Applications* (Polish Scientific Publishers, Warsaw, 1977), p. 346.
[20] We here adopt the notation in the book, G. Bachman and L. Narici, *Functional Analysis* (Academic, New York, 1966). The spectral properties of bounded normal operators are described in Chap. 28, pp. 484–93.
[21] Notice that the equalities (42) for all n are equivalent to the fact that the correspondent two-dimensional characteristic functions are equal. Then (42) follows [R. B. Ash, *Real Analysis and Probability* (Academic, New York, 1972), p. 327].

An upper bound on the allowed bands of the Bloch spectrum of one-dimensional Schrödinger operators with periodic potentials

N. N. Meiman
Former member of the Institute for Theoretical and Experimental Physics, Moscow, USSR

(Received 17 September 1981; accepted for publication 16 April 1982)

Using the principle of increasing flux and its complement, an upper bound associated with the allowed bands of the Bloch spectrum of one-dimensional Schrödinger operators with periodic potentials is derived. Namely, let a be the period of the potential, E_0 the lower bound of the Bloch spectrum and $(E_{2(j-1)},E_{2j-1})$ the allowed bands ($j=1,2,...$). Then

$$\pi > a[(E_{2j-1}-E_0)^{1/2}-(E_{2(j-1)}-E_0)^{1/2}].$$

PACS numbers: 03.65. — w

Let H be the one-dimensional Schrödinger operator $H=-d^2/dx^2+U_a$, with U_a a periodic potential of period a. The Bloch spectrum of H is defined as the set of values E for which all the solutions of the equation $(H-E)\phi=0$ are bounded on the whole real axis. The Block spectrum consists of the half-line $E_0<E$ from which certain closed intervals $[E_{2j-1},E_{2j}]$ ($j=1,2,...$) are omitted. These are the forbidden bands (or gaps), so the $(E_{2(j-1)},E_{2j-1})$ ($j=1,2,...$) are the allowed bands. It will be assumed that U_a nowhere equals the constant E_0.

In this note, the inequality

$$\pi > a[(E_{2j-1}-E_0)^{1/2}-(E_{2(j-1)}-E_0)^{1/2}] \quad (j=1,2,...) \tag{1}$$

will be proved. When the degeneracy $E_{2j-1}=E_{2j}$ occurs, the large allowed band (E_{2j-2},E_{2j+1}) is to be regarded as the two allowed bands, (E_{2j-2},E_{2j-1}) and (E_{2j},E_{2j+1}), with the inequality (1) valid for each of them.

This inequality is almost an immediate consequence of *the principle of increasing flux* and its *complement* proved previously.[1,2] We shall recall these principles below, as well as certain analytic properties of the quasimomentum function $p(E)$ previously established.

The principle of increasing flux may be stated as follows. Let V be a harmonic function which is positive in the region G of the complex plane and let α be an arc of the boundary ∂G of the region. Deform the arc α into the region G, keeping its endpoints fixed and call the resulting arc α_1. Define a new harmonic function V_1 in the smaller region $G_1\subset G$ bounded by $\partial G\setminus\alpha\cup\alpha_1$ by taking V_1 to have the same boundary values on $\partial G\setminus\alpha$ as V, and to vanish on α_1. Then if n is the internal normal,

$$\int_{\alpha_1}\frac{\partial V_1}{\partial n}\,ds>\int_{\alpha}\frac{\partial V}{\partial n}\,ds. \tag{2}$$

The complement of this principle of increasing flux is that for any arc $\beta\subset\partial G\setminus\alpha$,

$$\int_{\beta}\frac{\partial V_1}{\partial n}\,ds<\int_{\beta}\frac{\partial V}{\partial n}\,ds, \tag{3}$$

notation as before.

If $f(z)=U(z)+iV(z)$ and $f_1(z)=U_1(z)+V_1(z)$ are functions analytic in G and G_1, respectively, then the inequalities (2) and (3) imply the inequalities

$$U_1(z)|_{\alpha_1}>U(z)|_{\alpha}, \tag{2'}$$

$$U_1(z)|_{\beta}<U(z)|_{\beta}, \tag{3'}$$

where $U(z)|_{\alpha}$ is the increase of U along the arc α and so on. It is easy to prove that both principles are valid not only for two-dimensional domains but for arbitrary dimensions as well.

Let $\phi_j(x;E)$ ($j=1,2$) be a fundamental system of solutions of the equation $(H-E)\phi=0$. The effect of a displacement through one period is given by

$$\begin{pmatrix}\phi_1(x+a;E)\\ \phi_2(x+a;E)\end{pmatrix}=\Upsilon(E)\begin{pmatrix}\phi_1(x;E)\\ \phi_2(x;E)\end{pmatrix}, \tag{4}$$

where the monodromy, or shift, matrix $\Upsilon(E)$ satisfies $\det[\Upsilon(E)]=1$. Upon transforming to another fundamental system of solutions, $\Upsilon(E)$ is transformed into a conjugate matrix. Consequently the eigenvalues $\rho_j(E)$ ($j=1,2$) of $\Upsilon(E)$ are independent of the choice of basis. Moreover, $\rho_1\rho_2=1$ and so we can introduce the quasimomentum function $p(E)$ by

$$\rho_{1,2}(E)=\exp[\mp iap(E)]. \tag{5}$$

The factor a arises from the fact that $\Upsilon(E)$ depends not on E but on the dimensionless quantity $a^2(E-E_0)$.

Let C_0 be the complex E plane with the cut $[E_0,+\infty)$. The function $z=a[E-E_0]^{1/2}$ maps C_0 conformally onto the upper half-plane $\operatorname{Im} z>0$. Now the function $t=ap(E)$ is known[1] to depend only on z, so we write $ap(E)=\hat{p}(z)$. The function $\hat{p}$ is determined uniquely by the conditions $\hat{p}(0)=0$ and $\hat{p}(z)\sim z$ at infinity. It is also known[1] that the function t maps C_0 conformally onto the upper half-plane $\operatorname{Im} t>0$ with vertical cuts. The bases of these cuts are at points which are multiples of π. The cuts are the images of the upper and lower edges of the gaps $[E_{2j-1},E_{2j}]$. The intervals $\pm((j-1)\pi,j\pi)$ of the t plane are the images of the bands $(E_{2(j-1)},E_{2j-1})$ under this mapping. The conformal image of C_0 under t we denote by T; it is symmetric with respect to the imaginary axis and completely defined by the lengths L_j ($j=1,2,...$) of its vertical cuts.

We can apply the complement of the principle of increasing flux to the conformal mapping of the region T into the upper half-plane $\operatorname{Im} z>0$ by regarding the vertical cuts as arcs α_j and the intervals $((j-1)\pi,j\pi)$ as arcs β_j. The image of β_j in the half-plane $\operatorname{Im} z>0$ is the interval $(X_{2(j-1)},X_{2j-1})$, where $X_k=a(E_k-E_0)^{1/2}$. According to, e.g., (3'), if the length L_k of at least one vertical cut decreases

 0022-2488/83/030539-02$02.50

under the mapping of T, the differences $X_{2j-1} - X_{2(j-1)}$ increase ($j = 1,2,...$), reaching a maximum when all the image lengths vanish. In this case $t = z$, $X_{2j-1} = X_{2j}$, and $X_{2j-1} - X_{2(j-1)} = \pi$. Thus, the inequality Eq. (1) has been proved.

In Ref. 1, the inequality

$$a(E_{2N} - E_0)^{1/2} = X_{2N} > N\pi$$

was proved for the case when the Bloch spectrum has N bands. Considered as a special case of the analysis here, the vertical cuts are such that $L_j = 0$ for all $j > N$ and the arc α is just that part of the boundary ∂T between $-X_{2N} - 0$ and $X_{2N} + 0$.

[1]N. N. Meiman, J. Math. Phys. **18**, 834 (1977).

[2]The following points should be noted in Ref. 1. In the formula Eq. (2.8), z_n should be replaced by z_{2n}. In the Introduction and the beginning of Sec. 1, the condition $U_a(x) = U_a(x/a)$ was mentioned by mistake.

Explicit Hilbert space representations of Schrödinger states: Definitions and properties of Stieltjes–Tchebycheff orbitals

M. R. Hermann
Department of Chemistry, Indiana University, Bloomington, Indiana 47405

P. W. Langhoff
Department of Chemistry,[a] Indiana University, Bloomington, Indiana 47405
and Max Planck Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, 8046 Garching bei München, Federal Republic of Germany
and Laboratoire des Collisions Atomique et Moleculaires, Université Paris-Sud, Bâtiment 351, 91405 Orsay Cedex, France

(Received 18 June 1982; accepted for publication 22 October 1982)

Stieltjes–Tchebycheff orbital approximations are described for the discrete and continuum Schrödinger states of self-adjoint (Hamiltonian) operators, and certain of their properties are established and clarified. The nth order orbitals are defined in accordance with corresponding moment-theory approximations to spectral densities as eigenfunctions of the appropriate operator in an n-term Cauchy–Hilbert space. Their eigenvalues are consequently the generalized Gaussian or Radau quadrature points of the associated spectral density function formed by projection with an appropriate test function on the Schrödinger states, and their norms provide the corresponding (reciprocal) quadrature weights. Stable algorithms are described for their construction employing recursive Lanczos and orthogonal-polynomial methods. In finite orders the spatial characteristics of the orbitals correspond to spectral averages in the neighborhoods of the quadrature points over the correct Schrödinger states. The spectral content of an individual orbital is obtained in closed form without reference to the correct underlying Schrödinger states. Convergence ($n \to \infty$) is obtained in the discrete spectral region to Schrödinger eigenstates of finite norm, whereas in the essential portion of the spectrum the orbitals converge to scattering states of improper (infinite) norm. Connections with matrix partitioning and optical potential theory are made, indicating that finite-order orbitals can provide exact Schrödinger states over local predetermined portions of configuration space. Illustrative computational studies of the regular l-wave spectra of the Coulomb Hamiltonian are provided.

PACS numbers: 03.65.Db, 03.65.Ge, 31.15 + q

I. INTRODUCTION

Hilbert space provides an appropriate setting for theoretical and computational studies of Schrödinger states.[1] In spite of this, so-called improper vectors are commonly employed in solutions of scattering problems associated with the essential portions of Hamiltonian spectra.[2] Difficulties can arise in the explicit construction of such scattering states, particularly for many-particle systems in which noncentral and nonlocal forces are present, and in which particle permutation symmetry must be satisfied. Since L^2 (Ritz) variational methods have proved useful for construction of the bound states of such complex systems,[3] there is considerable motivation to devise related approximations for the associated scattering states.[4] Techniques based on the theory of moments,[5,6] which have received widespread applications in solid-state,[7] atomic and molecular,[8] and nuclear physics,[9] seem particularly well suited for construction of explicit Hilbert-space approximations of uniform quality to the discrete and continuum portions of Schrödinger spectra.

In the present article, an account is given of Stieltjes–Tchebycheff orbitals that provide explicit Hilbert-space representations of Schrödinger states, and certain of their properties are established and clarified. The orbitals are defined in accordance with corresponding moment-theory approximations to spectral densities, distributions, and Stieltjes-integral representations of operator resolvents that arise upon introduction of appropriate test-function projections on Schrödinger spectra.[10] Although moment-theory approximations are familiar in these latter contexts[5–10] and the general spectral theory is well established,[11] relatively few applications of the corresponding Stieltjes–Tchebycheff approach to L^2 representations of Schrödinger states have been reported in the literature.[12]

In Sec. II, required definitions are provided, and connection is made with previous work describing convergent moment-theory approximations to spectral densities.[10] The Stieltjes–Tchebycheff orbitals are defined in Secs. III and IV as Hamiltonian eigenstates in n-term Cauchy–Hilbert basis sets,[13] and explicit algorithms for their construction are given employing Lanczos recurrence methods.[14] In Sec. V, the nth-order orbitals are shown to provide optical-potential solutions of the full matrix representation of the Schrödinger problem. Their convergence in the limit of large n to eigenstates of finite norm in discrete spectral regions, and to scat-

[a] Permanent address.

 0022-2488/83/030541-07$02.50

tering states of improper (infinite) norm in the essential spectrum, is indicated. In finite orders, the orbitals are seen to provide spectral averages over the correct Schrödinger states in the neighborhoods of the generalized Gaussian or Radau quadrature points of the corresponding spectral density. Moreover, their norms are seen to give the corresponding (reciprocal) quadrature weights. Consequently, in finite orders the orbitals are expected to sensibly span the important dense portions of the spectrum. Closed form expressions are obtained for the spectral content of the Stieltjes–Tchebycheff orbitals without explicit reference to or construction of the appropriate Schrödinger states. These and other properties of Stieltjes–Tchebycheff orbitals are clarified with illustrative calculations of the regular l-wave spectra of the one-electron Coulomb Hamiltonian, given in Sec. VI. Concluding remarks are made in Sec. VII.

II. SPECTRAL DENSITIES AND DISTRIBUTIONS

It is convenient to consider the expectation value of the resolvent $(z-H)^{-1}$ of the self-adjoint (Hamiltonian) operator H of interest with respect to a unit-norm test function Φ. This can be written for nonsingular spectra as a Stieltjes integral[5]:

$$R_\Phi(z) \equiv \langle \Phi | (z-H)^{-1} | \Phi \rangle = \int_{-\infty}^{+\infty} \frac{g(\epsilon)\, d\epsilon}{z-\epsilon} = \int_{-\infty}^{+\infty} \frac{d\Gamma(\epsilon)}{z-\epsilon}, \tag{1}$$

where

$$g(\epsilon) = \sum_i \langle \Phi | \varphi_i \rangle |^2 \, \delta(\epsilon - \epsilon_i) + | \langle \Phi | \varphi_\epsilon \rangle |^2 \tag{2}$$

is the (nonnegative) spectral density function of interest and

$$\Gamma(\epsilon) = \int_{-\infty}^{\epsilon} g(\epsilon')\, d\epsilon', \quad \Gamma(-\infty) = 0 \tag{3}$$

is the corresponding (nondecreasing) distribution. The denumerable (possibly infinite) number of bound-state eigenfunction φ_i ($\epsilon_i < 0$) and the scattering functions φ_ϵ ($\epsilon \geqslant 0$) of Eq. (2) satisfy the Schrödinger problem

$$(H - \epsilon_i)\, \varphi_i = 0, \quad \epsilon_i < 0, \tag{4a}$$

$$(H - \epsilon)\, \varphi_\epsilon = 0, \quad \epsilon \geqslant 0, \tag{4b}$$

as well as appropriate boundary conditions in each case. By convention, the orthogonal eigenstates φ_i are made to satisfy physical normalization,

$$\langle \varphi_i | \varphi_j \rangle = \delta_{ij}, \tag{5a}$$

whereas the scattering solutions φ_ϵ satisfy Dirac delta-function normalization,[2]

$$\langle \varphi_\epsilon | \varphi_{\epsilon'} \rangle = \delta(\epsilon - \epsilon'), \tag{5b}$$

in the energy variable. Appropriate choices of the test function Φ result in spectral densities [Eq. (2)] and Stieltjes integrals [Eq. (1)] having well-defined and familiar physical significance.[10] Common cases include the absorption and dispersion of electromagnetic radiation in matter,[15,16] aspects of equilibrium statistical mechanics,[17,18] crystal-lattice vibrations,[19–21] and certain formulations of particle scattering,[22] to mention some representative examples.

Spectral power moments of the form (A, p. 141; CL, p. 652)

$$\mu_k = \langle \Phi | H^k | \Phi \rangle = \int_{-\infty}^{+\infty} \epsilon^k\, d\Gamma(\epsilon), \quad k \geqslant 0, \tag{6}$$

provide a basis for construction of moment-theory approximations to the resolvent, density, and distribution of Eqs. (1)–(3) under appropriate conditions.[5,6,10] In the present development, corresponding Stieltjes–Tchebycheff orbital approximations to the spectrum of bound-state eigenfunctions and scattering solutions of Eqs. (4) and (5) are described, and their properties are established and clarified. When the moments of Eq. (6) fail to converge, the development can be generalized by replacing H in Eq. (6) with an appropriate operator $A(H)$ chosen to assure convergence. This flexibility has important consequences in computational applications, as indicated further in Sec. VI.

It is assumed throughout for ease of presentation that a strongly convergent sequence of Jacobi-matrix representatives of the operator H can be constructed in a separable Hilbert space, the operator consequently permits a closure, the moments of Eq. (6) converge, and the corresponding moment problem is a determinate one. Under these conditions there is a close correspondence between the moment problem and construction of a spectral resolution of the corresponding self-adjoint operator H (A, Chap. 4).

III. STIELTJES ORBITALS

Stieltjes approximations to the resolvent, density, and distribution of Eqs. (1)–(3) are obtained from nth-order generalized Gaussian quadrature points $\epsilon_i^{(n)}$ and weights $\Gamma_i^{(n)}$ that satisfy the reduced moment problem (ST, p. 28; CL, Sec. 3)

$$\sum_{i=1}^{n} (\epsilon_i^{(n)})^k\, \Gamma_i^{(n)} = \mu_k, \quad 0 \leqslant k \leqslant 2n-1. \tag{7}$$

Corresponding nth-order Stieltjes orbitals $\{\Phi_i^{(n)};\ i = 1\text{–}n\}$ that have the $\epsilon_i^{(n)}$ as eigenvalues and the $\Gamma_i^{(n)}$ as (reciprocal) norms are constructed as the eigenfunctions of an $n \times n$ matrix representative of the operator H in a so-called Cauchy–Hilbert moment basis[13,14]

$$\theta_k = H^{k-1}\Phi, \quad k = 1,2,\ldots,n. \tag{8}$$

Since the metric of the Cauchy basis is positive definite $\{\langle \theta_i | \theta_j \rangle \equiv \mu_{i+j-2}\}$ for a moment problem with solution (ST, p. viii), the n functions of Eq. (8) form a linearly independent invariant Hilbert subspace that includes the test function $\Phi\ (\equiv \theta_1)$. That the $\epsilon_i^{(n)}$ of Eq. (7) are Hamiltonian eigenvalues in the Cauchy basis follows from explicit construction of the secular equation,[23] or from the observation that the development is a variational one and that the Hamiltonian matrix is determined by the first $2n$ spectral moments $\{\langle \theta_i | H | \theta_j \rangle = \mu_{i+j-1}\}$.[24]

It is well known that instabilities can arise when the Cauchy basis of Eq. (8) is employed directly in forming an eigenvalue problem.[7–9] These difficulties are avoided

through use of orthonormal linear combinations of the Cauchy basis that bring the Hamiltonian to tridiagonal Jacobi form.[6,7] The resulting Lanczos basis satisfies the recurrence relation[14]

$$\beta_j v_{j+1} = (H - \alpha_j)\, v_j - \beta_{j-1} v_{j-1}, \quad j = 1,...,n-1, \tag{9}$$

with $v_0 = 0$ and $v_1 = \Phi$. Equation (9) and the orthonormality requirement on the v_j are sufficient to determine the Lanczos basis, as well as the recurrence coefficients α_j and β_j ($j = 1$–n). In computational applications an appropriate convenient complete L^2 basis is generally employed to represent the v_j and operator H of Eq. (9), thereby avoiding explicit reference to the Cauchy basis of Eq. (8). Necessary boundary conditions on the Schrödinger states are conveniently incorporated into the choice of basis (Sec. VI), and the Hamiltonian H is replaced in Eq. (9) with A (H) in a generalization of the development required when the moments of Eq. (6) fail to converge, as indicated above.

The Schrödinger problem in this Lanczos basis takes the form

$$(\mathbf{J}^{(n)} - \epsilon_i^{(n)}\, \mathbf{I}^{(n)}) \cdot \mathbf{q}_i^{(n)} = 0, \quad i = 1,...,n, \tag{10}$$

where $\mathbf{J}^{(n)}$ is the $(n \times n)$ tridiagonal Jacobi matrix (A, p. 138) [cf. Eq. (9)]

$$(\mathbf{J}^{(n)})_{i,i} = \alpha_i, \tag{11a}$$

$$(\mathbf{J}^{(n)})_{i,i+1} = \beta_i. \tag{11b}$$

The components of the eigenvector of Eq. (10) are (A, p. 140)

$$(\mathbf{q}_i^{(n)})_j = q_j(\epsilon_i^{(n)}), \tag{11c}$$

where $q_j(\epsilon)$ is the jth-order polynomial orthonormal with respect to the distribution $\Gamma(\epsilon)$ (ST, p. 35)

$$\int_{-\infty}^{+\infty} q_i(\epsilon)\, q_j(\epsilon)\, d\Gamma(\epsilon) = \delta_{ij}, \tag{12}$$

and consequently can be computed from the recurrence relation (ST, p. 47)

$$\beta_j\, q_{j+1}(\epsilon) = (\epsilon - \alpha_j)\, q_j(\epsilon) - \beta_{j-1}\, q_{j-1}(\epsilon), \tag{13}$$

where $q_0(\epsilon) = 0$, $q_1(\epsilon) = 1$, similar to Eq. (9).[25] Writing out Eq. (13) for $j = 1$ to n gives a matrix equation similar to Eq. (10), with the zero on the right-hand side there replaced by a vector having the polynomial $q_{n+1}(\epsilon)$ in the nth position and zeros elsewhere. Consequently, the nth-order Stieltjes energies, or generalized Gaussian quadrature points, are obtained from diagonalization of the $\mathbf{J}^{(n)}$ matrix or, equivalently, from the roots of the orthogonal polynomial [CL, Eq. (13a)]

$$q_{n+1}(\epsilon) = (-1)^n \det|\, \mathbf{J}^{(n)} - \epsilon \mathbf{I}^{(n)}|/(\beta_1 \beta_2 \cdots \beta_n). \tag{14}$$

Corresponding quadrature weights are given by the Christoffel expression (ST, pp. 39–41)

$$\Gamma_i^{(n)} = \left[\sum_{j=1}^{n} q_j(\epsilon_i^{(n)})^2 \right]^{-1}. \tag{15}$$

Finally, associated Stieltjes orbitals are written in the form (A, pp. 140, 141)

$$\Phi_i^{(n)} = \sum_{j=1}^{n} q_j(\epsilon_i^{(n)})\, v_j = \sum_{j=1}^{n} q_j(\epsilon_i^{(n)})\, q_j(H)\, \Phi. \tag{16}$$

The metric of the Stieltjes orbitals

$$\langle \Phi_i^{(n)} |\, \Phi_j^{(n)} \rangle = \sum_{k=1}^{n} q_k(\epsilon_i^{(n)})\, q_k(\epsilon_j^{(n)}) = \delta_{ij}/\Gamma_i^{(n)}, \tag{17}$$

where the Christoffel–Darboux relation has been employed (A, pp. 9, 22), indicates the (reciprocal) norms of the orbitals provide nth-order Gaussian quadrature weights. By contrast, the norms of the Cauchy functions [Eq. (8)] are seen to be the spectral moments of Eq. (6).

Properties of the Stieltjes orbitals of Eq. (16) are indicated and clarified in Sec. V below. It is noted here that, since their eigenvalues and norms provide generalized Gaussian quadratures of the appropriate density function, the Stieltjes orbitals are sure to appear in the important portion of the spectrum. Moreover, because of the interlacing theorem satisfied by the roots of the polynomials $q_n(\epsilon)$ and $q_{n+1}(\epsilon)$ (A, p. 10), the Stieltjes energies are dense in the essential spectrum in the limit of large n, whereas monotone convergence to the Schrödinger eigenvalues is obtained in the bound-state region.

IV. TCHEBYCHEFF ORBITALS

Tchebycheff approximations to the resolvent, density, and distribution of Eqs. (1)–(3) are obtained from the nth-order Radau quadrature points $\epsilon_i^{(n)}(\epsilon)$ and weights $\Gamma_i^{(n)}(\epsilon)$ that satisfy the reduced moment problem (CL, Sec. IV)

$$\epsilon^k\, \Gamma^{(n)}(\epsilon) + \sum_{i=1}^{n-1} [\epsilon_i^{(n)}(\epsilon)]^k\, \Gamma_i^{(n)}(\epsilon) = \mu_k, \quad 0 \leqslant k \leqslant 2n-2. \tag{18}$$

Equation (18) differs from Eq. (7) defining the generalized Gaussian quadratures in that one of the points $\epsilon_n^{(n)}(\epsilon)$ ($\equiv \epsilon$) is assigned a preselected value in the interval $-\infty$ to $+\infty$. Consequently, the weight at that point $\Gamma_n^{(n)}(\epsilon)$ [$\equiv \Gamma^{(n)}(\epsilon)$], as well as the $n-1$ other points $\epsilon_i^{(n)}(\epsilon)$ and $n-1$ weights $\Gamma_i^{(n)}(\epsilon)$, are explicit functions of the preselected value ϵ.[10]

The nth-order orthonormal Tchebycheff orbitals corresponding to the quadratures of Eq. (18) are obtained in the Lanczos basis of Eq. (9) from the equation

$$(\mathbf{J}^{(n)}(\epsilon) - \epsilon \mathbf{I}^{(n)}) \cdot \mathbf{q}_\epsilon^{(n)} = 0. \tag{19}$$

Here, the $n \times n$ tridiagonal matrix $\mathbf{J}^{(n)}(\epsilon)$ is identical to the matrix of Eqs. (10) and (11) with the exception of the n,n element (α_n), which is replaced by a so-called modified recurrence coefficient [CL, Eq. (23)]:

$$(\mathbf{J}^{(n)}(\epsilon))_{n,n} = \alpha_n(\epsilon) = \epsilon - \beta_{n-1}\, q_{n-1}(\epsilon)/q_n(\epsilon) = \alpha_n + \beta_n\, q_{n+1}(\epsilon)/q_n(\epsilon). \tag{20a}$$

Moreover, the vector of Eq. (19) is given by (A, p. 140)

$$(q_\epsilon^{(n)})_j = q_j(\epsilon), \tag{20b}$$

as in Eq. (11c). Equation (19) provides n Radau quadrature points or Tchebycheff eigenvalues $\epsilon_i^{(n)}(\epsilon)$ and n corresponding orthonormal eigenstates, with the former obtained from diagonalization of the $\mathbf{J}^{(n)}(\epsilon)$ matrix or, equivalent, from the roots of the quasiorthogonal polynomial [CT, Eq. (25a)]

$$q_{n+1}(\epsilon', \epsilon) = (-1)^n \det|\, \mathbf{J}^{(n)}(\epsilon) - \epsilon' \mathbf{I}^{(n)}|/(\beta_1 \beta_2 \cdots \beta_n)$$
$$= q_{n+1}(\epsilon') - [q_{n+1}(\epsilon)/q_n(\epsilon)]\, q_n(\epsilon'). \quad (21)$$

The corresponding quadrature weights are obtained from the Christoffel expression (ST, pp. 39–41)

$$\Gamma^{(n)}(\epsilon) = \left[\sum_{i=1}^{n} q_i(\epsilon)^2 \right]^{-1}, \quad (22)$$

for ϵ equal to any of the Radau points. Finally, the Tchebycheff orbitals can be written (A, pp. 140, 141)

$$\Phi_\epsilon^{(n)} = \sum_{j=1}^{n} q_j(\epsilon)\, v_j$$
$$= \sum_{j=1}^{n} q_j(\epsilon)\, q_j(H)\, \Phi, \quad (23)$$

and the corresponding metric is

$$\langle \Phi_\epsilon^{(n)} |\, \Phi_{\epsilon'}^{(n)} \rangle = \sum_{j=1}^{n} q_j(\epsilon)\, q_j(\epsilon')$$
$$= 1/\Gamma^{(n)}(\epsilon), \quad \epsilon = \epsilon'. \quad (24)$$

Although the Tchebycheff orbitals of order n are not orthogonal for arbitrary ϵ and ϵ', they are so at the Radau quadrature points or roots of Eq. (21). As for the Stieltjes orbitals [Eq. (17)], Tchebycheff orbital norms are seen to provide the (reciprocal) Radau quadrature weights. Since the preselected point ϵ can be placed anywhere in the spectrum, it is seen the Tchebycheff orbital of Eq. (23) can be constructed at an arbitrary discrete or continuum Schrödinger energy.

Properties of the Tchebycheff orbitals of Eq. (23) are indicated and clarified in Sec. V below. It is convenient to note here that construction of the Lanczos functions [Eq. (9)] and orthogonal polynomials [Eq. (13)] required in solution of the Stieltjes problem [Eq. (10)] is also sufficient for the construction of the Tchebycheff orbitals. Since the roots of orthogonal [Eq. (14)] and quasiorthogonal polynomials [Eq. (21)] of common order satisfy an interlacing theorem (ST, p. 37), one and only one Tchebycheff orbital will appear in each of the $n+1$ distinct intervals defined by the nth-order Stieltjes eigenvalues. This observation proves to have computationally useful consequences (CL, p. 654).

V. PROPERTIES OF STIELTJES–TCHEBYCHEFF ORBITALS

Properties of the Stieltjes–Tchebycheff orbitals defined in the preceding sections are established and clarified here. First, an important connection with matrix partitioning or optical potential theory is provided, and second the appropriate physical renormalization and convergence of the orbitals is indicated.

A. Optical potential property

The modified recurrence coefficient $\alpha_n(\epsilon)$ of Eq. (20a) provides the exact optical potential arising from a matrix partitioning solution of the infinite-order Stieltjes–Schrödinger problem [Eq. (10), $n \to \infty$].

To establish and clarify the nature of this property, the Schrödinger equation in the full Lanczos basis [Eq. (10), $n \to \infty$] is written in the partitioned form

$$\begin{bmatrix} \mathbf{J}^{(n)} & \boldsymbol{\beta} \\ \boldsymbol{\beta}^\dagger & \mathbf{R} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{q}_\epsilon^{(n)} \\ \mathbf{Q}_\epsilon \end{bmatrix} = \epsilon \begin{bmatrix} \mathbf{q}_\epsilon^{(n)} \\ \mathbf{Q}_\epsilon \end{bmatrix}. \quad (25)$$

Here, $\mathbf{J}^{(n)}$ is the $(n \times n)$ tridiagonal matrix of Eqs. (10) and (11), $\boldsymbol{\beta}$ is an $(n \times \infty)$ matrix having only one nonzero element (β_n, in the lower left-hand corner), $\boldsymbol{\beta}^\dagger$ is its $(\infty \times n)$ adjoint, and $\mathbf{R}$ is an $(\infty \times \infty)$ tridiagonal remainder matrix having α_{n+1} as its first (1,1) element. The first n terms of the solution vector of Eq. (25) are written as $\mathbf{q}_\epsilon^{(n)}$ in anticipation of the desired result [cf. Eq. (19)], with $\mathbf{Q}_\epsilon$ representing the infinitely many remaining components. Solving Eq. (25) for $\mathbf{q}_\epsilon^{(n)}$ by eliminating $\mathbf{Q}_\epsilon$ in the standard way gives[26]

$$(\mathbf{J}^{(n)} + \mathbf{V}^{(n)}(\epsilon) - \epsilon \mathbf{I}^{(n)}) \cdot \mathbf{q}_\epsilon^{(n)} = 0, \quad (26)$$

where the second (optical potential) matrix on the left-hand side of Eq. (26) is given by the formal expression

$$\mathbf{V}^{(n)}(\epsilon) = \boldsymbol{\beta} \cdot [\, \mathbf{I}(\epsilon + i0) - \mathbf{R}\,]^{-1} \cdot \boldsymbol{\beta}^\dagger. \quad (27a)$$

It is shown by inspection of Eq. (25) and explicit construction that $\mathbf{V}^{(n)}(\epsilon)$ has only one nonzero element (n,n), given by[27]

$$[\mathbf{V}^{(n)}(\epsilon)]_{n,n} = \cfrac{\beta_n^2}{\epsilon - \alpha_{n+1} - \cfrac{\beta_{n+1}^2}{\epsilon - \alpha_{n+2} - \cdots}}$$
$$= \beta_n\, q_{n+1}(\epsilon)/q_n(\epsilon). \quad (27b)$$

As a consequence of the recursive definition of the orthonormal polynomials $q_n(\epsilon)$ [Eq. (13)], the infinite-order continued fraction of Eq. (27b) can be represented by the indicated rational function for any energy ϵ in the Schrödinger spectrum [Eqs. (4) and (25)]. Comparison of this result with the modified recurrence coefficient $\alpha_n(\epsilon)$ of Eq. (20a) indicates that Eq. (26) is identical with Eq. (19) satisfied by the Tchebycheff orbitals, establishing the optical potential property.

Comments: As a consequence of this property, it is seen that the $q_j(\epsilon)$ of Eq. (20b) provide projections of Lanczos functions v_j on the correct Schrödinger state for energy ϵ to within a common constant (normalization) factor. These projections are not order-dependent in any sense, a consequence of the recursive definition of the Lanczos functions of Eq. (9). The Tchebycheff orbitals can be made, therefore, to provide highly reliable approximations in finite orders to local portions of Schrödinger states by appropriate choice of test function Φ. Since the polynomial $q_{n+1}(\epsilon)$ of Eq. (27b) vanishes when ϵ is set equal to a Gaussian quadrature or Stieltjes point $\epsilon_i^{(n)}$ of Eq. (7), nth-order Stieltjes orbitals [Eq. (16)] are seen to provide special matrix partitioning solutions for which the optical potential of Eqs. (26) and (27) vanishes identically. At the roots of the polynomial $q_n(\epsilon)$, which are distinct from those of $q_{n+1}(\epsilon)$, the optical potential diverges, and the n orthonormal eigensolutions of Eqs. (19) and (26) include one state at infinity having zero weight [Eq. (22)].

B. Convergence and renormalization property

For appropriate choices of test function Φ, the Tchebycheff orbitals of Eq. (23) converge in the pointwise sense in the limit of large n to Schrödinger eigenstates of finite norm ($\epsilon = \epsilon_i < 0$) or to scattering states of improper (infinite) norm ($\epsilon \geqslant 0$) that can be renormalized in the physical or Dirac delta-function sense, respectively.

Choice of test function Φ selects a symmetry subspace in which convergence to Schrödinger states may be

achieved. Convergence in this subspace of the Tchebycheff orbitals for large n to Schrödinger states [Eqs. (4) and (5)] follows from property (a) above, the linear independent of the n Lanczos functions of Eq. (9), and the strong convergence of a sequence of corresponding nth-order Jacobi matrices [Eqs. (10) and (11)] to a representative of the self-adjoint operator H in the limit $n\to\infty$.[11] The norms of the nth-order Tchebycheff orbitals are given by [Eq. (24), $\epsilon=\epsilon'$]

$$\langle \Phi_\epsilon^{(n)} | \Phi_\epsilon^{(n)} \rangle = \sum_{j=1}^{n} q_j(\epsilon)^2, \tag{28}$$

which is known to satisfy the conditions (A, p. 140)

$$\sum_{j=1}^{\infty} q_j(\epsilon)^2 < \infty, \quad \epsilon=\epsilon_i, \tag{29a}$$

$$\to\infty, \quad \epsilon\neq\epsilon_i, \tag{29b}$$

where the ϵ_i are the appropriate Schrödinger eigenvalues [Eqs. (4a) and (10), $n\to\infty$]. Consequently, physically normalized Tchebycheff orbitals appropriate for both discrete and essential portions of the spectrum can be written

$$\psi_\epsilon^{(n)} = [\, g^{(n)}(\epsilon)\,]^{1/2} \sum_{j=1}^{n} q_j(\epsilon)\, v_j, \tag{30}$$

where

$$g^{(n)}(\epsilon) = \Gamma^{(n)}(\epsilon), \quad \epsilon<0, \tag{31a}$$

$$= G^{(n)}(\epsilon), \quad \epsilon\geqslant 0. \tag{31b}$$

Here, $\Gamma^{(n)}(\epsilon)$ is the nth-order quadrature weight of Eq. (22), and $G^{(n)}(\epsilon)$ is the corresponding Stieltjes–Tchebycheff derivative approximation to the density $|\langle \Phi | \varphi_\epsilon \rangle|^2$ in the essential portion of the spectrum.[10] Noting that [CL, Lemma (a)]

$$g^{(n)}(\epsilon) \overset{n\to\infty}{\to} \sum_i |\langle \Phi \mid \varphi_i \rangle|^2\, \delta(\epsilon-\epsilon_i) + |\langle \Phi \mid \varphi_\epsilon \rangle|^2, \tag{32}$$

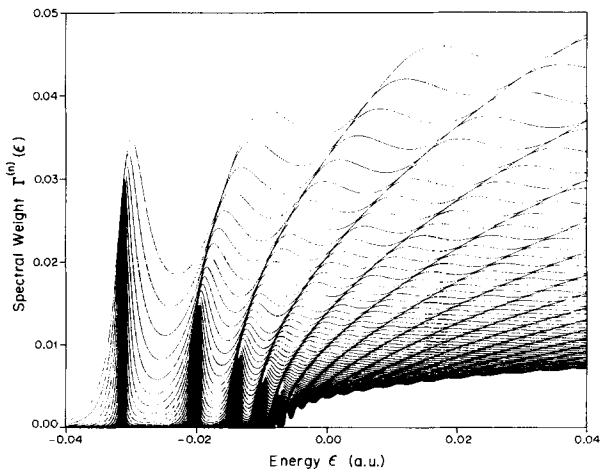


FIG. 1. Tchebycheff weights $\Gamma^{(n)}(\epsilon)$ of Eq. (22) [n = 20–100] for the p-wave or dipole spectrum in atomic hydrogen. Convergence to discrete transition strengths at the appropriate Coulomb energies is evident, whereas there is seen to be monotone decrease in the essential portion of the spectrum, in accordance with Eqs. (29). The waves evident in the figure are a consequence of the interlacing theorem that causes Tchebycheff results of adjacent orders to coincide at the Stieltjes eigenvalues. Hartree atomic units are employed.

the orthonormality of the states of Eq. (30) in the limit of large n,

$$\langle \psi_\epsilon^{(n)} | \psi_{\epsilon'}^{(n)} \rangle = [\, g^{(n)}(\epsilon)\, g^{(n)}(\epsilon')\,]^{1/2} \sum_{j=1}^{n} q_j(\epsilon)\, q_j(\epsilon')$$

$$\overset{n\to\infty}{\to} \delta_{ij}, \quad \epsilon=\epsilon_i, \quad \epsilon'=\epsilon_j,$$

$$\overset{n\to\infty}{\to} \delta(\epsilon-\epsilon'), \quad \epsilon,\epsilon'\geqslant 0,$$

$$\overset{n\to\infty}{\to} 0, \quad \epsilon<0, \quad \epsilon'>0, \tag{33}$$

follows from well-known properties of the kernel polynomial appearing on the right-hand side (A, p. 9).

Comments: In finite order, the Tchebycheff orbital $\psi_\epsilon^{(n)}$ of Eq. (30) corresponds to a spectral average over Schrödinger states largely in the neighborhood of the energy ϵ. With increasing order the spectral width of $\psi_\epsilon^{(n)}$ diminishes in accordance with the corresponding convergence of Tchebycheff orbitals to Schrödinger states indicated above. Since the correct Schrödinger states can be written as a limit ($n\to\infty$) of Eq. (30), the spectral contribution of an arbitrary discrete or scattering state $\varphi_{\epsilon'}$ [Eqs. (4)] to $\psi_\epsilon^{(n)}$ is [cf. Eq. (28)]

$$\langle \varphi_{\epsilon'} | \psi_\epsilon^{(n)} \rangle = [\, g(\epsilon')\, g^{(n)}(\epsilon)\,]^{1/2} \sum_{j=1}^{n} q_j(\epsilon')\, q_j(\epsilon)$$

$$\overset{n\to\infty}{\to} \delta_{ij}, \quad \epsilon=\epsilon_i, \quad \epsilon'=\epsilon_j,$$

$$\overset{n\to\infty}{\to} \delta(\epsilon-\epsilon'), \quad \epsilon,\epsilon'\geqslant 0. \tag{34}$$

Similarly, the spectral content of a Lanczos function is

$$\langle \varphi_{\epsilon'} \mid v_j \rangle = [\, g(\epsilon')\,]^{1/2}\, q_j(\epsilon'), \tag{35}$$

emphasizing the spatial-spectral correspondence between Lanczos functions and orthogonal polynomials.

VI. ILLUSTRATIVE APPLICATION

Stieltjes–Tchebycheff orbitals are constructed for the one-electron Coulomb Hamiltonian as an illustration of the foregoing development. Test functions of the form $\Phi\sim r^l P_l(\cos\theta)\,\varphi_0(r)$, where $\varphi_0(r)$ is the ground-state eigenfunction, are chosen to pick out the various regular l-wave spectra. In order to obtain convergent spectral moments in these cases, it is necessary to replace the Hamiltonian H in the development with the operator $A(H) = (H-\epsilon_0)^{-1}$, where ϵ_0 is the ground state eigenvalue. Since this operator has the same eigenfunctions as H itself, the Stieltjes–Tchebycheff orbitals will converge to the appropriate states, in accordance with remarks made in Sec. II. The choice $l=1$ picks out the dipole spectrum for ground-state hydrogen, and $g(\epsilon)$ is proportional to the photoexcitation/ionization cross section. Attention is focused in the following on this familiar case.

The Cauchy [Eq. (8)] and Lanczos [Eq. (9)] functions for the dipole or regular p-wave spectrum in atomic hydrogen are obtained in terms of associated Laguerre polynomials of appropriate order multiplied by the ground-state eigenfunction. These functions are denumerable and complete in the

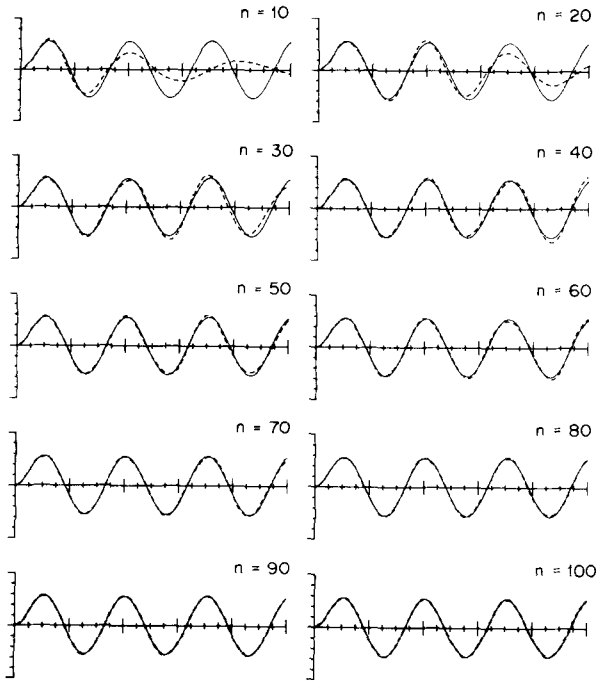


FIG. 2. Tchebycheff orbitals $r\psi_\epsilon^{(n)}(r)$ of Eq. (23) [n = 10–100] for regular p-waves of energy $\epsilon = 2$ a.u. in atomic hydrogen. Convergence to the correct Coulomb functions (solid lines) is seen to be rapid in the neighborhood of the origin, emphasizing that the Tchebycheff orbitals provide local approximations to scattering states appropriate for the dipole spectrum in this case. Hartree atomic units are employed, with the abscissa covering $10a_0$, and the ordinate 1.0 atomic units.

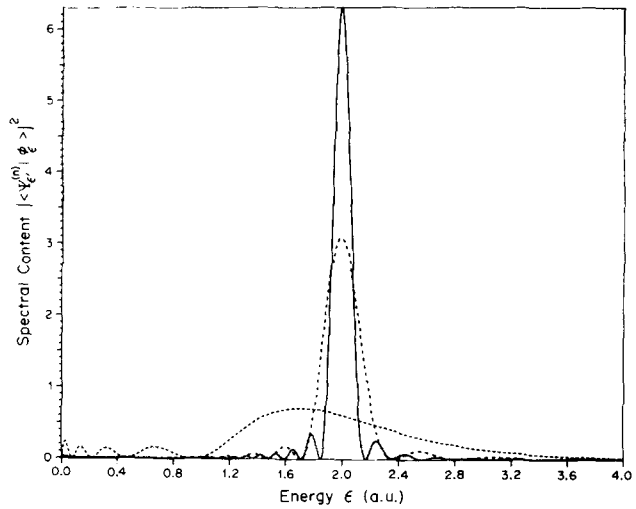


FIG. 3. Spectral content $|\langle\psi_\epsilon^{(n)}|\varphi_\epsilon\rangle|^2$ of Eq. (34) [n = 10, 50, and 100, $\epsilon' = 2$ a.u.] for regular p-waves in atomic hydrogen. Convergence to a delta-function-like structure in accordance with Eq. (32) is evident. The points of zero spectral content occur at the Tchebycheff–Radau energies [Eq. (18)] of appropriate order. Hartree atomic units are employed.

limit of large numbers.[28] The corresponding polynomial recurrence coefficients [Eqs. (9)–(13)] are also obtained in explicit analytical form.[29] Consequently, Stieltjes–Tchebycheff orbitals of arbitrary order can be easily constructed, and the results compared with the correct Coulomb p-states and regular p-waves.[30]

Irregular l-waves can be constructed using a basis that incorporates a $1/r$ factor and an appropriate cutoff to insure integral convergence. Alternatively, a second set of solutions linearly independent of the regular l-waves can be constructed in the Laguerre basis by introduction of so-called orthonormal polynomials of the second kind in Eq. (23) (A, p. 8). The resulting Tchebycheff orbitals of the second kind are solution of an inhomogeneous Schrödinger equation with the test function providing the appropriate inhomogeneity. These orbitals are regular at the origin but are asymptotically linear combinations of regular and irregular Coulomb waves.[31] It should also be noted that solutions of the Weyl type for complex energy ($\epsilon \to \epsilon \pm i\lambda$, $\epsilon, \lambda > 0$),[32] which remain L^2 in the limit $n \to \infty$, can also be constructed employing appropriate combinations of orthonormal polynomials of the first and second kind (A, p. 15).

Tchebycheff–Radau quadrature weights [Eq. (22)] $\Gamma^{(n)}(\epsilon)$ of various orders up to $n = 200$ are constructed in an energy interval spanning the discrete states and extending well into the essential spectrum. In Fig. 1 are shown results for n = 20–100 in the neighborhood of the ionization threshold $\epsilon = 0$. Evidently, the weights converge to finite values, proportional to so-called f numbers in this case,[30] at the correct discrete Coulomb eigenvalues, in accordance with Eq. (29a). By contrast, the weights decrease rapidly with increasing order at energies between the correct eigenvalues, whereas in the essential spectrum the decrease is less rapid but in accordance with Eq. (29b). Note that $\Gamma^{(n+1)}(\epsilon) < \Gamma^{(n)}(\epsilon)$ for all ϵ, in accordance with Eq. (22), with the equality holding at the Stieltjes points $\epsilon = \epsilon_i^{(n)}$, $i = 1$–n, in which case the $(n+1)$-order Techbycheff spectrum has a zero weight at infinity. The Stieltjes energies of increasing orders evidently form apparent rays that converge to the Coulomb eigenvalues monotonically from above, in accordance with the interlacing theorem indicated above.

Convergence of the Stieltjes–Tchebycheff orbitals corresponding to the weights or reciprocal norms of Fig. 1 is verified by detailed comparisons with the correct Coulomb p-states and regular p-waves.[30] It is found that the orbitals in the discrete spectrum converge to Coulomb eigenstates at the correct energies, whereas for energies $\epsilon < 0$ between the ϵ_i, convergence to highly oscillatory unphysical states is obtained. In the essential spectrum $\epsilon \geqslant 0$, smooth convergence to local portions of the regular p-waves is obtained with increasing order. In Fig. 2 are shown representative orbitals of order 10–100 in steps of 10 for $\epsilon = 2$ atomic units energy, in comparison with the correct corresponding regular p-waves.[30] It is seen that the orbitals exhibit smooth convergence to a correctly normalized scattering state. The corresponding spectral contributions of Schrödinger states to the orbitals of Fig. 2 are shown in Fig. 3. These are seen to become delta-function-like in high order, in accordance with Eq. (34).

VII. CONCLUDING REMARKS

Explicit Hilbert-space approximations constructed in accordance with corresponding moment-theory results are described for the discrete and continuum Schrödinger states

of self-adjoint (Hamiltonian) operators. Moment properties of the Lanczos basis set employed in the development insures the resulting Stieltjes–Tchebycheff orbitals provide optical potential solutions of the matrix Schrödinger equation, and insures that they furnish correctly normalized L^2 approximations to the local portions of scattering states, as well as to the bound states of the system. An application in the case of the one-electron Coulomb Hamiltonian illustrates the nature of the convergence achieved. More detailed computational applications are reported elsewhere.[31]

ACKNOWLEDGMENTS

Acknowledgment is made to the Donors of the Petroleum Research Fund, administered by the American Chemical Society, and to the Chemistry and International Programs Divisions of the National Science Foundation, for partial support of this work. It is a pleasure to thank G. H. F. Diercksen, MPI, Garching, and P.-M. Guyon, Orsay, for their kind hospitality to PWL.

[1]J. Von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U.P., Princeton, N. J., 1955).
[2]R. J. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[3]S. T. Epstein, *The Variational Method in Quantum Chemistry* (Academic, New York, 1974).
[4]J. D. Broad, Phys. Rev. A **18**, 1012 (1978).
[5]J. A. Shohat and J. D. Tamarkin, *The Problem of Moments, Mathematical Surveys I* (American Mathematical Society, Providence, R.I., 1943, revised 1970). Referred to as ST in what follows.
[6]N. I. Akhiezer, *The Classical Moment Problem* (Oliver and Boyd, London, 1955). Referred to as A in what follows.
[7]R. Haydock, in *Computational Methods in Classical and Quantum Physics*, edited by M. B. Hooper (Hemisphere, Washington, D.C., 1975), pp. 268–83.
[8]P. W. Langhoff, in *Electron–Molecule and Photon–Molecule Collisions*, edited by T. Rescigno, V. McKoy, and B. Schneider (Plenum, New York, 1979), pp. 183–224.
[9]R. R. Whitehead, in *Theory and Applications of Moment Methods in Many-Fermion Systems*, edited by B. J. Dalton, S. M. Grimes, J. P. Vary, and S. A. Williams (Plenum, New York, 1980), pp. 235–55.
[10]C. T. Corcoran and P. W. Langhoff, J. Math. Phys. **18**, 651 (1977). Referred to as CL in what follows.
[11]D. Mason, in *The Padé Approximant in Theoretical Physics*, edited by G. A. Baker, Jr., and J. L. Gammel (Academic, New York, 1970), pp. 197–218.
[12]M. R. Hermann and P. W. Langhoff, Chem. Phys. Lett. **82**, 242 (1981).
[13]P. W. Langhoff and M. Karplus, J. Chem. Phys. **52**, 1435 (1970).
[14]C. Lanczos, J. Res. Nat. Bur. Stds. **45**, 255 (1950).
[15]P. C. Martin, Phys. Rev. **161**, 143 (1967).
[16]R. Kubo and M. Ichimura, J. Math. Phys. **13**, 1454 (1972).
[17]R. G. Gordon, J. Math. Phys. **9**, 655 (1968).
[18]A. Lonke, J. Math. Phys. **12**, 2422 (1971).
[19]J. Peretti, J. Phys. Chem. Solids **12**, 216 (1960).
[20]J. C. Wheeler and R. G. Gordon, J. Chem. Phys. **51**, 5566 (1969).
[21]P. W. Langhoff and C. T. Corcoran, Mol. Phys. **26**, 203 (1972).
[22]M. Baker, Ann. Phys. **4**, 271 (1958).
[23]Yu. V. Vorobyev, *Method of Moments in Applied Mathematics* (Gordon and Breach, New York, 1965).
[24]P. W. Langhoff, J. S. Sims, and C. T. Corcoran, Phys. Rev. A **10**, 829 (1974).
[25]The polynomials of Eqs. (12) and (13) differ from those of Ref. 10 by explicit incorporation of the appropriate normalization factor in the present development and a shift in the ordering index, whereas the polynomial recurrence coefficients $(\alpha_j\ \beta_j)$ employed here are identical with those $(\alpha_j\ \beta_{j-1}^{1/2})$ of the previous development.
[26]P.-O. Lowdin, Int. J. Quantum Chem. **21**, 69 (1982).
[27]H. S. Wall, *Analytic Theory of Continued Fractions* (Van Nostrand, New York, 1948).
[28]H. Shull and P.-O. Lowdin, J. Chem. Phys. **23**, 1362 (1955).
[29]P. W. Langhoff, C. T. Corcoran, J. S. Sims, F. Weinhold, and R. M. Glover, Phys. Rev. A **14**, 1042 (1976).
[30]H. Bethe and E. Saltpeter, *Quantum Mechanics of One-and Two-Electron Atoms* (Academic, New York, 1957).
[31]M. R. Hermann and P. W. Langhoff, Phys. Rev. A (to be published).
[32]E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGaw-Hill, New York, 1955).

On the correspondence between classical and quantum mechanics.[a] I

Reese T. Prosser
Department of Mathematics, Dartmouth College, Hanover, New Hampshire 03755

(Received 16 June 1982; accepted for publication 15 October 1982)

We consider here a general procedure for implementing the correspondence between classical and quantum mechanical systems of finitely many degrees of freedom. We show that corresponding systems may be formulated within a common framework in such a way that the kinematic, statistical, dynamical, and covariance features may be easily compared.

PACS numbers: 03.65.Fd, 03.65.Ca, 03.65.Bz

INTRODUCTION

We consider here a general procedure for implementing the celebrated correspondence between classical and quantum mechanical systems of finitely many degrees of freedom. This correspondence is usually described in something like the following terms. If a classical system is given in Hamiltonian form, then the corresponding quantum system is obtained by "quantizing" the observable functions of the canonical coordinates, i.e., by replacing the classical functions f,g---by their quantum counterparts F,G---. These are taken as self-adjoint operators on a suitable Hilbert space, in such a way that the classical Poisson bracket $\{f,g\}$ is replaced by (a multiple of) the quantum commutator bracket $(i\hbar)^{-1}[F,G]$. Here, as usual, $\hbar = h/2\pi$, where h is Planck's quantum of action, a fundamental parameter in the quantum theory. If, on the other hand, a quantum system is given, then the corresponding classical system is obtained by taking the limit $\hbar \to 0$; in this limit the quantum observables, $F,G,$---are replaced by their classical counterparts f,g--- and $(i\hbar)^{-1}[F,G]$ by $\{f,g\}$.

This rather vague description serves well in simple cases, well enough, in fact, to form a basis for the historical development of quantum mechanics[1]; but it leaves open a number of important questions. What classical observables, for example, have quantum counterparts, and how are they defined? How should the (noncommutative) quantum coordinates be ordered, and what difference does the ordering make? What classical canonical transformations have quantum counterparts, and how are they obtained? More generally, how much of the Hamiltonian structure of a classical system has a quantum counterpart? How does one quantize a classical system given in curvilinear coordinates, or on a general manifold?

In order to study these questions a more precise description of the correspondence between classical and quantum systems is needed. Such a description should be precise enough to yield detailed information on the nature of the correspondence, general enough to include all known special cases, and flexible enough to provide an access to the questions raised above.

We should observe at the outset that it is somewhat misleading to speak of this correspondence as the process of "quantizing" a classical system, a locution which suggests that the classical system is basic and the quantum system is derived. But in fact the quantum system is usually considered the more accurate model of nature, and the corresponding classical system appears only in the limit $\hbar \to 0$; in this sense the quantum system is basic and the classical system is derived. From our point of view the process of quantizing a classical system becomes the process of recovering the corresponding quantum system from its asymptotic form.

With this much understood, we turn now to the problem of defining the correspondence between classical and quantum systems.

1. QUANTIZATION

We begin by considering those classical systems in Hamiltonian form which have as their phase space the Euclidean space R^{2d} with its usual symplectic structure. We denote by $\mathbf{C}$ the set of classical observable functions $f(x,\xi)$ of the canonical coordinates (x,ξ); we take these here to be smooth and real valued. The corresponding quantum systems in Hilbert space form will have as their state space the Hilbert space $H = L^2(R^d)$. We denote by $\mathbf{Q}$ the set of quantum observable functions $F(X,\Xi)$ of the quantum coordinates (X,Ξ); we take these to be self-adjoint operators on $L^2(R^d)$.

By a *quantization* of $\mathbf{C}$ we understand (provisionally) a mapping $\mathbf{d\theta}: \mathbf{C} \to \mathbf{Q}$ such that $\mathbf{d\theta}$ is linear, and

$$\mathbf{d\theta}\, 1 = I, \tag{1}$$

$$\mathbf{d\theta}\{f,g\} = -i\lambda\,[\mathbf{d\theta}\, f, \mathbf{d\theta}\, g], \tag{2}$$

$$\mathbf{d\theta}(f^2) = (\mathbf{d\theta}\, f)^2 \quad \text{whenever } f^2 \in \mathbf{C}. \tag{3}$$

Here 1 denotes the constant function, I the identity operator, $\{\ ,\ \}$, and $[\ ,\ ]$ the usual Poisson and commutator brackets, and λ a positive parameter.

The conditions (1)–(3) together form a minimal set of reasonable requirements for any quantization process. Condition (1) gives a nontrivial normalization. The importance of condition (2) was emphasized by Dirac,[1] who observed that the corresponding brackets determine the development in time of the classical and quantum systems. Condition (3), taken together with (1), implies that $\mathbf{d\theta}\,(p(f)) = p(F)$ for any real polynomial p for which $p(f) \in \mathbf{C}$; this condition is essential for the standard interpretation of classical and quantum probabilities in terms of moments, as shown by von Neumann.[2]

When we take $\mathbf{C} = P_2$, the set of quadratic polynomials in $C^\infty(R^{2d})$, then we obtain a quantization for P_2 by setting $\mathbf{d\theta}\, 1 = I$, and

[a] Part of this work was supported by NSF Grant No. MCS81-03409.

 0022-2488/83/030548-05$02.50

$$\mathbf{d\theta}(x_i) = X_i = \text{multiplication by } x_i,$$

$$\mathbf{d\theta}(\xi_j) = \Xi_j = (i\lambda)^{-1}\frac{\partial}{\partial x_j}. \tag{4}$$

This prescription, with $\lambda^{-1} = \hbar$, is the usual point of departure for the study of quantum mechanics,[1] and gives the Schrödinger quantization of the harmonic oscillator.

Groenewold[3] has shown, however, that there exists no such quantization satisfying conditions (1)–(3) for any larger class of polynomials on r^{2d}. His result suggests that the quantization of more general classes C of classical observables will require some modification of these conditions. (A nice review of this problem was recently given by Chernoff[4].)

Van Hove[5] has given a more complicated prescription for quantizing C, which forms the basis for the "geometric quantization" of Kostant[6] and Souriau,[7] as well as some of the work of Segal.[8] In our terms, he replaces $L^2(R^d)$ by $L^2(R^{2d})$, and then defines $\mathbf{d\theta}$ for a large class of observables f by

$$\mathbf{d\theta}(f) = (i\lambda)^{-1}\left(\frac{\partial f}{\partial \xi}\frac{\partial}{\partial x} - \frac{\partial f}{\partial x}\frac{\partial}{\partial \xi}\right) + \left(f - \xi\frac{\partial}{\partial \xi}\right). \tag{5}$$

He shows that his prescription satisfies conditions (1) and (2), but violates condition (3), and of course it does not reduce to the Schrödinger quantization of P_2 as given by (4). There is some evidence to suggest that it yields anomolous results for the anharmonic oscillator.[9]

Weyl[10] has given a different prescription for the quantization of a large (unspecified) class C which does reduce to the Schrödinger quantization of P_2 but satisfies neither (2) nor (3). He defines X_i and Ξ_j by (4), and then sets

$$\hat{f}(\eta,y) = (2\pi)^{-d}\int e^{-i(x\eta - y\xi)}f(x,\xi)\,dx\,d\xi, \tag{6}$$

$$\mathbf{d\theta}(f) = (2\pi)^{-d}\int e^{i\lambda(X\eta - y\Xi)}\hat{f}(\eta,y)\,d\eta\,dy. \tag{7}$$

This prescription resolves $f(x,\xi)$ into Fourier components, and then uses these components to weight the Weyl representation $\exp i\lambda(X\eta + y\Xi)$ of the Heisenberg group generated by X_i and Ξ_j.

Hörmander[11] has recently shown that Weyl's prescription can be conveniently and usefully described in terms of the modern theory of pseudodifferential operators acting on $L^2(R^d)$: if f is smooth and suitably restricted at infinity, then $\mathbf{d\theta}f = F$, where for $u \in L^2(R^d)$,

$$Fu(x) = (2\pi)^{-d}\int e^{i\lambda(x-y)\xi}f\left(\frac{x+y}{2},\xi\right)u(y)\,dy\,d\xi. \tag{8}$$

He then shows that this prescription makes sense and yields a symmetric pseudodifferential operator F on $L^2(R^d)$ with symbol f, if f is real-valued and lies in one of his classes $S(m,g)$.

Hörmander then develops an operational calculus for such operators. He shows that if $f \in S(m,g)$, $g \in S(n,g)$, then there exists a function $h = f\#g \in S(mn,g)$ such that, as operators on $L^2(R^d)$,

$$\mathbf{d\theta}(f\#g) = (\mathbf{d\theta}(f))(\mathbf{d\theta}(g)). \tag{9}$$

Moreover, one can express $f\#g$ in terms of f and g as an asymptotic series, valid for large λ:

$$\begin{aligned}(f\#g)(x,\xi) &= (\pi)^{-2d}\int f(x+r,\xi+\rho)\\ &\quad\times g(x+s,\xi+\sigma)e^{2i\lambda(\rho s - r\sigma)}dr\,d\rho\,ds\,d\sigma\\ &= \exp - i\left(\frac{\partial_\xi\partial_y - \partial_x\partial_\eta}{2\lambda}\right)\\ &\quad\times f(x,\xi)g(y,\eta)|_{(y,\eta)=(x,\xi)}\\ &= \sum_{j=0}^{\infty}\frac{1}{j!}\left(\frac{-i}{2\lambda}\right)^j\{f,g\}_j(x,\xi),\end{aligned} \tag{10}$$

where we have put

$$\{f,g\}_j(x,\xi) = (\partial_\xi\partial_y - \partial_x\partial_\eta)^j f(x,\xi)g(y,\eta)|_{(y,\eta)=(x,\xi)}. \tag{11}$$

It follows that

$$\begin{aligned}&\frac{1}{2}(f\#g - g\#f)(x,\xi)\\ &\quad= \cos\left(\frac{\partial_\xi\partial_y - \partial_x\partial_\eta}{2\lambda}\right)f(x,\xi)g(y,\eta)|_{(y,\eta)=(x,\xi)}\\ &\quad= \sum_{j=0}^{\infty}\frac{(-1)^j}{(2j)!(2\lambda)^{2j}}\{f,g\}_{2j}(x,\xi)\\ &\quad= (fg)(x,\xi) - \frac{1}{8\lambda^2}\{f,g\}_2(x,\xi) + O\left(\frac{1}{2\lambda}\right)^4,\end{aligned} \tag{12}$$

$$\begin{aligned}&-\frac{1}{2i}(f\#g - g\#f)(x,\xi)\\ &\quad= \sin\left(\frac{\partial_\xi\partial_y - \partial_x\partial_\eta}{2\lambda}\right)f(x,\xi)g(y,\eta)|_{(y,\eta)=(x,\xi)}\\ &\quad= \sum_{j=0}^{\infty}\frac{(-1)^{j+1}}{(2j+1)!(2\lambda)^{2j+1}}\{f,g\}_{2j+1}(x,\xi)\\ &\quad= \frac{1}{2\lambda}\{f,g\}_1(x,\xi) - \frac{1}{48\lambda^3}\{f,g\}_3(x,\xi) + O\left(\frac{1}{2\lambda}\right)^5.\end{aligned} \tag{13}$$

In particular,

$$f\#f = f^2 - \frac{1}{8\lambda^2}\{f,f\}_2 + O\left(\frac{1}{2\lambda}\right)^4, \tag{14}$$

$$\begin{aligned}\{f,g\}_Q &= +i\lambda(f\#g - g\#f)\\ &= \{f,g\}_1 - \frac{1}{24\lambda^2}\{f,g\}_3 + O\left(\frac{1}{2\lambda}\right)^5.\end{aligned} \tag{15}$$

Note that $\{f,f\}_2 = \partial_\xi^2 f\partial_x^2 f - 2(\partial_\xi\partial_x f)^2$ is the *Hessian* of f, while $\{f,g\} = \partial_\xi f\partial_x g - \partial_x f\partial_\xi g$ is the classical Poisson bracket $\{f,g\}_C$. The form (15) was first introduced by Moyal[12] in his investigation of the Weyl quantization, and is sometimes called the Moyal bracket.

The relations (9), (14), and (15) show that the Weyl quantization procedure "almost" satisfies conditions (2) and (3), in the sense that the differences $f\#f - f^2$ and $\{f,g\}_Q - \{f,g\}_C$ are both of order $(1/2\lambda)^2$ for large λ. Moreover, if f and g are smooth functions with bounded derivatives of order two or more, then these estimates are uniform over R^{2d}. These relations also show that the Weyl procedure satisfies (1), (2), and (3) *exactly* if $\{f,g\}_C$ is replaced by $\{f,g\}_Q$ in (2) and f^2 by $f\#f$ in (3).

The Weyl procedure has other advantages which we shall explore below.

2. STATISTICS

The statistical framework of classical Hamiltonian mechanics may be described as follows. Let $\rho(x,\xi)$ be a smooth

bounded probability density defined over the classical phase space R^{2d}. Then

$$\rho(x,\xi)\geqslant 0, \tag{16}$$

$$(2\pi)^{-d}\int \rho(x,\xi)dx\,d\xi = 1. \tag{17}$$

By the *expected value* of the classical observable $f\in\mathbf{C}$ in the state described by ρ we mean

$$E_\rho(f) = (2\pi)^{-d}\int f(x,\xi)\rho(x,\xi)dx\,d\xi. \tag{18}$$

Then $E_\rho\colon \mathbf{C}\to R$ has all the usual properties of an expectation, and yields the usual physical interpretation.

The statistical framework of quantum mechanics, as developed by von Neumann,[2] may be described in similar terms as follows. Let R be a von Neumann probability density matrix on the quantum state space $L^2(R^d)$. Then

$$R\geqslant 0, \tag{19}$$

$$\mathrm{tr}(R) = 1. \tag{20}$$

Thus R is positive definite and of trace class. By the *expected value* of the quantum observable $F\in\mathbf{Q}$ in the state described by R we mean

$$E_R(F) = \mathrm{tr}\,(FR). \tag{21}$$

Then $E_R\colon \mathbf{Q}\to R$ has all the usual properties and the usual interpretation of an expectation.

Now suppose $R\in\mathbf{Q}$, with symbol $r\in\mathbf{C}$. (It can be shown that the density matrices of this form are dense in the trace norm among all such matrices.) Then (21) becomes

$$\begin{aligned} E_R(F) &= (2\pi)^{-2d}\int e^{i\lambda(x-y)\xi} f\left(\frac{x+y}{2},\xi\right) \\ &\quad\times e^{i\lambda(y-x)\eta} r\left(\frac{y+x}{2},\eta\right) dy\,d\eta\,dx\,d\xi \\ &= (2\pi)^{-d}\int f(z,\zeta)r(z,\zeta)dz\,d\zeta. \end{aligned} \tag{22}$$

In particular, (20) becomes

$$\mathrm{tr}(R) = E_R(I) = (2\pi)^{-d}\int r(z,\zeta)dz\,d\zeta. \tag{23}$$

Comparison of (23) and (22) with (17) and (18) shows that in this setting the classical and quantum expectations have exactly the same form. They are not identical, however, because the class P_C of classical densities satisfying (16), (17) and the class P_Q of quantum densities satisfying (19) and (20) are not the same. To compare them, we rewrite (16), (17) as

$$\begin{aligned} (2\pi)^{-d}\int f^2(x,\xi)\rho(x,\xi)dx\,d\xi &\geqslant 0 \quad \text{for all } f\in\mathbf{C}, \\ &= 1 \quad \text{for } f = 1, \end{aligned} \tag{24}$$

and rewrite (19), (20) as

$$\begin{aligned} (2\pi)^{-d}\int (f\#f)(x,\xi)r(x,\xi)dx\,d\xi &\geqslant 0 \quad \text{for all } f\in\mathbf{C}, \\ &= 1 \quad \text{for } f = 1. \end{aligned} \tag{25}$$

Thus P_C and P_Q differ only in their definitions of positivity.

It is now clear that the Weyl quantization procedure can be carried out entirely within $\mathbf{C}$, and amounts to replacing the Poisson bracket $\{\ ,\ \}_C$ by the Moyal bracket $\{\ ,\ \}_Q$, while replacing the classical probability densities P_C by their quantum counterparts P_Q. All of the statistical features of the quantum mechanical system may now be formulated within $\mathbf{C}$, and may be compared with their classical counterparts within a common framework. This point of view was first espoused by Moyal.[12]

The same is true of the dynamical features, as we see in the next section.

3. DYNAMICS

From now on we suppose that $\mathbf{C}$ consists of all smooth real-valued functions $f(x,\xi)$ satisfying the condition

$$|\partial_x^\alpha\partial_\xi^\beta f(x,\xi)|\leqslant \mathbf{C}_{\alpha\beta} \quad \text{if } |\alpha| + |\beta|\geqslant 2. \tag{26}$$

Note that under this condition $\mathbf{C}$ includes (but is not restricted to) the space P_2 of all quadratic polynomials.

Now if $h(x,\xi)\in\mathbf{C}$, then $h(x,\xi)$ generates a one-parameter group $\gamma(t)$ of global canonical transformations (i.e., a Hamiltonian flow) of the phase space R^{2d}. This flow preserves the phase space volume by Liouville's theorem, and so induces a flow of the space P_C of probability densities on R^{2d}, and a unitary flow of the space $L^2(R^{2d})$.

If we put $(x(t),\xi(t)) = \gamma(t)(x,\xi)$ and, for $g\in L^2(R^{2d})$, $g(t) = g\circ\gamma(-t)$, then by Hamiltonian equations of motion,

$$\begin{aligned} \frac{dx(t)}{dt} &= \{h,x(t)\}_C = \frac{\partial h}{\partial\xi}(x(t),\xi(t)), \\ \frac{d\xi(t)}{dt} &= \{h,\xi(t)\}_C = -\frac{\partial h}{\partial x}(x(t),\xi(t)), \end{aligned} \tag{27}$$

$$\frac{dg(t)}{dt} = -\{h,g(t)\}_C. \tag{28}$$

If $g\in L^2(R^{2d})$ then we may integrate (28) by

$$\begin{aligned} g(t) &= U(t)g \\ &= \exp(+iLt)g, \end{aligned} \tag{29}$$

where L is the Liouville operator associated with h:

$$Lg = i\{h,g\}_C. \tag{30}$$

It is not hard to see that L is essentially self-adjoint on $L^2(R^{2d})$ and generates the unitary flow $U(t)$ induced on $L^2(R^{2d})$ by $\gamma(t)$.

The quantum counterpart H of h, as given by (8), is an essentially self-adjoint operator on $L^2(R^d)$ if $h\in\mathbf{C}$, and so generates a unitary flow $\Gamma(t,\lambda) = \exp i\lambda Ht$ on the quantum state space $L^2(R^d)$. In some sense $\Gamma(t,\lambda)$ is the quantum counterpart of $\gamma(t)$.

To make this precise, we first note that if R is a probability density matrix on $L^2(R^d)$ then, since $\Gamma(t,\lambda)$ is unitary, $\Gamma(-t,\lambda)R\Gamma(+t,\lambda)$ isanother. Thus $\Gamma(t,\lambda)$ induces a flow in the space P_Q of quantum probability density matrices. More generally if G is any Hilbert–Schmidt operator, then $\Gamma(-t,\lambda)\,G\,\Gamma(+t,\lambda)$ is another with the same Hilbert–Schmidt norm, so $\Gamma(t,\lambda)$ induces a unitary flow on the space HS of Hilbert–Schmidt operators on $L^2(R^d)$.

If we put, for $f\in L^2(R^d)$, $f(t) = U(t)f$, and for $G\in HS$, $G(t,\lambda) = \Gamma(-t,\lambda)\,G\,\Gamma(+t,\lambda)$ then by Schrödinger's equations of motion,

$$\frac{df(t)}{dt} = i\lambda Hf(t), \tag{31}$$

$$\frac{dG(t,\lambda)}{dt} = -i\lambda\,[H,G(t,\lambda)]. \tag{32}$$

Now suppose $G \in HS$ is a pseudodifferential operator given by (8), with symbol $g \in \mathbf{C}$. (It is not hard to show that finite linear combinations of such operators are dense in HS in the Hilbert–Schmidt norm). Assume, subject to later verification, that $G(t,\lambda)$ is also a pseudodifferential operator, with symbol $g(t,\lambda)$. In that case (32) becomes

$$\begin{aligned}\frac{dg(t,\lambda)}{dt} &= -i\lambda\,(h\,\#g(t,\lambda) - g(t,\lambda)\#h) \\ &= -\{h,g(t,\lambda)\}_Q. \end{aligned} \tag{33}$$

Moreover, if $G \in HS$ with $g \in \mathbf{C}$, then G^2 is of *trace class*, and we have

$$\begin{aligned}\|G\|_{HS}^2 &= \mathrm{tr}(G^2) \\ &= (2\pi)^{-d}\int (g\#g)(x,\xi)\,dx\,d\xi \\ &= (2\pi)^{-d}\int g^2(z,\zeta)\,dz\,d\zeta. \end{aligned} \tag{34}$$

Thus we may identify the space of Hilbert–Schmidt operators with $L^2(R^{2d})$ and the Hilbert–Schmidt norm with the L^2 norm. Moreover, $\Gamma(t,\lambda)$ induces a unitary flow on $L^2(R^{2d})$ satisfying (33). We may integrate (33) by

$$\begin{aligned} g(t,\lambda) &= W(t,\lambda)g \\ &= \exp(iM(\lambda)t)g, \end{aligned} \tag{35}$$

where $M(\lambda)$ is the Moyal operator associated with h given by [cf. (30)]

$$M(\lambda)g = i\{h,g\}_Q. \tag{36}$$

We may verify directly that $M(\lambda)$, like L, is essentially self-adjoint on $L^2(R^{2d})$ if $h \in \mathbf{C}$.

In this way we see that the dynamics of both classical and quantum mechanics can be formulated within the same framework, in terms of unitary flows on the *same* Hilbert space $L^2(R^{2d})$. These flows are generated by the observables in $\mathbf{C}$ acting via the Poisson bracket as in (30), or via the Moyal bracket, as in (36). The space $L^2(R^{2d})$ contains all the (smooth bounded) probability densities of both classical and quantum type and hence contains all the statistical features of both systems, as well as the dynamics.

We should expect to find that as $\hbar \to 0$, the quantum flows reduce to the classical flows. To see how this comes about, let us assume, again subject to later verification, that the function $g(t,\lambda)$ of (35) admits an expansion in powers of λ of the form

$$g(t,\lambda) = \sum_{j=0}^{\infty}\left(\frac{-1}{4\lambda^2}\right)^j g_{2j}(t). \tag{37}$$

Now insert (37) into (36), and equate powers of λ: [cf. (15)]

$$\begin{aligned}\frac{dg_0(t)}{dt} &= \{h,g_0(t)\}_1, \\ \frac{dg_2(t)}{dt} &= \{h,g_2(t)\}_1 - \frac{1}{3!}\{h,g_0(t)\}_3. \end{aligned} \tag{38}$$

The first of these equations is the same as (28) and shows that we may take

$$g_0(t) = g(t,\infty) = g\circ\gamma(-t) \tag{39}$$

which is just the motion of g under the classical flow. The next equation in (38) may be written

$$\begin{aligned}\frac{\partial g_2(t)}{\partial t} &= \frac{dg_2(t)}{dt} - \{h,g_2(t)\}_1 \\ &= -\frac{1}{3!}\{h,g_0(t)\}_3 \end{aligned} \tag{40}$$

from which

$$g_2(t) = -\frac{1}{3!}\int_0^t \{h,g_0(\tau)\}_3 d\tau, \tag{41}$$

where the integral is taken along the classical path in R^{2d}. In the same way we find that

$$g_{2k}(t) = \sum_{i+j=k}\frac{(-1)^{j+1}}{(2j+1)!}\int_0^t \{h,g_{2i}(\tau)\}_{2j+1}\,d\tau. \tag{42}$$

In this way we establish the expansion (37) for $g(t,\lambda)$, with

$$g(t,\lambda) - g(t,\infty) = -\frac{1}{24}\lambda^2 g_2(t) + O\left(\frac{1}{2\lambda}\right)^4. \tag{43}$$

It remains to establish the validity of our assumptions on $g(t,\lambda)$. To do this we must formulate a more precise description of the classical and quantum flows on $L^2(R^{2d})$, starting from (30) and (31). In part II we show that these flows may be expressed in terms of the modern theory of Fourier integral operators, developed by Hörmander[13] for a related purpose.

In the final section we consider the question of general canonical transformations.

4. SYMMETRIES

Now let $\mathbf{D}$ denote the subset of $\mathbf{C}$ consisting of smooth real-valued functions $f(x,\xi)$ satisfying

$$|\partial_x^\alpha \partial_\xi^\beta f(x,\xi)| \leqslant C_{\alpha\beta}(1 + |x| + |\xi|)^{2-|\alpha|-|\beta|}. \tag{44}$$

Then $\mathbf{D}$ forms a Lie algebra (infinite dimensional) under the Poisson bracket, an algebra of classical infinitesimal generators, which we denote by $\mathfrak{g}_C = (\mathbf{D},\{\ ,\ \}_C)$. Moreover, we know from the Hamilton–Jacobi Theory that this Lie algebra generates a Lie group (again infinite dimensional), a group of classical canonical transformations of the phase space R^{2d}, and hence a group of classical unitary transformations of the Hilbert space $L^2(R^{2d})$ over R^{2d}. We denote this group by G_C.

But $\mathbf{D}$ also forms a Lie algebra under the Moyal bracket, an algebra of quantum infinitesimal generators, which we denote by $\mathfrak{g}_Q = (\mathbf{D},\{\ ,\ \}_Q)$. We know from the previous sections that this Lie algebra also generates a Lie group, a group of quantum unitary transformations of the Hilbert space $L^2(R^{2d})$. Moreover, we know that as $\hbar \to 0$, $\{\ ,\ \}_Q \to \{\ ,\ \}_C$, and so $\mathfrak{g}_Q \to \mathfrak{g}_C$ and $G_Q \to G_C$. In this sense G_Q may be considered as a deformation of G_C, admitting G_C as an asymptotic limit.

Finally, we note that if $h \in \mathbf{C}$ and $g \in \mathbf{D}$, then both $\{g,h\}_C$ and $\{g,h\}_Q$ lie in $\mathbf{C}$. It follows that $\mathbf{C}$ is invariant under the action of both $\mathfrak{g}_C$ and $\mathfrak{g}_Q$ and hence under the action of both G_C and G_Q. In this way the group G_C (G_Q) is a group of symmetries for the classical (quantum) observables.

Since $\mathbf{D}$ contains P_2, the quadratic polynomials, it follows that G_C and G_Q contain the linear symplectic group

$\mathrm{Sp}(2d,R)$ generated by P_2. Moreover, since **D** contains all smooth functions $f(x,\xi)$ with compact support, it follows that G_C contains *all* smooth canonical transformations of R^{2d} which are locally arbitrary but reduced to the identity outside their compact support, and G_Q contains their quantum counterparts.

Note added in proof. The formulation of Weyl's quantization prescription in terms of pseudodifferential operators has also been studied by Grossman, Loupias, and Stein,[14] Voros,[15] and Antonetz.[16] The importance of the Moyal representation has been emphasized by Hermann.[17]

[1]P. A. M. Dirac, *The Principles of Quantum Mechanics*, 4th ed. (Oxford, 1958).

[2]J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton, New Jersey, 1955).

[3]H. J. Groenewold, Physica **12**, 405–460 (1946).

[4]P. R. Chernoff, Hadrionic Journal (to appear).

[5]L. Van Hove, Proc. R. Acad. Belgium **26**, 1–102 (1951).

[6]B. Kostant, *Quantization and Unitary Representations, Lecture Notes in Mathematics*, Vol. 170 (Springer, Berlin, 1970).

[7]J.-M. Souriau, *Structures des Systems Dynamiques* (Dunod, Paris, 1970).

[8]I. E. Segal, J. Math. Phys. **2**, 468 (1960); **5**, 264 (1964).

[9]R. T. Prosser, J. Math. Phys. **5**, 701–707 (1964).

[10]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950).

[11]L. Hörmander, Comm. Pure Appl. Math. **33**, 359–443 (1979).

[12]J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99–124 (1949).

[13]L. Hörmander, Acta Math. **127**, 79–183 (1971).

[14]A. Grossman, G. Loupias, and E. Stein, Ann. Inst. Fourier (Grenoble) **18**, 343–368 (1968).

[15]A. Voros, J. Func. Anal. **29**, 104–132 (1978).

[16]M. A. Antonetz, Mater. Sb. **107** (149), 20–36 (1978).

[17]R. Hermann, *Quantum and Fermion Differential Geometry, Part A* (Math Science, Brookline, MA, 1977).

A class of solvable Pauli–Schrödinger Hamiltonians

Miguel Calvo[a)]
Centro de Física, Instituto Venezolano de Investigaciones Cientificas, Apdo. 1827, Caracas 1010A, Venezuela

Simón Codriansky
Departamento de Matemáticas y Física, Instituto Universitario Pedagógico de Caracas, Avenue Páez, El Paraíso, Caracas 1021, Venezuela

(Received 11 May 1982; accepted for publication 6 August 1982)

The Pauli–Schrödinger equation for a spin 1/2, neutral particle with a nonvanishing magnetic moment μ_0, interacting with an external scalar potential V and a static magnetic field **B**, both functions of only one of the coordinates, is solved exactly for four different choices of the potential and the field. By choosing in the examples the coordinate y, we present these solutions in the following cases:

(i) $V(y)=0$, $\mathbf{B}(y)=(B_0\sin\kappa y, 0, B_0\cos\kappa y)$ where B_0 and κ are two arbitrary constants.

(ii) $V(y)=\lambda' y^2$, $\mathbf{B}(y)=|\alpha y|(\cos 2\kappa y, 0, \sin 2\kappa y)$ where $\lambda=\hbar^2/2m$ and α,κ arbitrary constants and

$$\lambda'=\lambda\left(\frac{\mu_0}{2\kappa\lambda}\right)^2.$$

(iii) $V(y)=\lambda'(\bar{c}_1\tan\alpha y+\bar{c}_2\cot\alpha y)^2$, $\mathbf{B}(y)=B(y)(\sin(2\kappa y+2\delta(y)),0,\cos(2\kappa y+2\delta(y))$ where

$$B^2(y)=(\bar{c}_1\tan\alpha y+\bar{c}_2\cot\alpha y)^2+\frac{\alpha^2}{4\kappa^2}(\bar{c}_1\sec^2\alpha y-\bar{c}_2\csc^2\alpha y)^2$$

and

$$\tan 2\delta(y)=\frac{2\kappa}{\lambda}\frac{\bar{c}_1\tan\alpha y+\bar{c}_2\cot\alpha y}{\bar{c}_1\sec^2\alpha y-\bar{c}_2\csc^2\alpha y};$$

$\bar{c}_1$, $\bar{c}_2$, α and κ are arbitrary constants.

(iv) $V(y)=\lambda'(\bar{a}\tanh z+\bar{b})^2$ where $z=[(y-cd)/d]$,

$$B^2(y)=(\bar{a}\tanh z+\bar{b})^2+(\bar{a}^2/4\kappa^2d^2)\operatorname{sech}^2 z,$$

$$\tan 2\delta(y)=\frac{2\kappa d}{\bar{a}}\frac{\bar{a}\tanh z+\bar{b}}{\operatorname{sech}^2 z};$$

$\bar{a},\bar{b},c,d$, and κ are arbitrary constants restricted only by $(\mu_0/2\kappa\lambda)\bar{a}^2+\bar{a}/d>0$. The functions $B(y)$ and $\delta(y)$ define the vector $\mathbf{B}(y)$ as in (iii). Our method of solution is based on the familiar factorization technique for solving the Schrödinger eigenvalue equation. Several interesting physical results of these solutions are discussed.

PACS numbers: 03.65.Ge

I. INTRODUCTION

Let us consider the motion of a spin 1/2 neutral particle of mass M, magnetic moment μ_0, in the presence of a static magnetic field **B** and an external static potential V. The corresponding eigenvalue equation for the stationary states of the system is given by

$$\left\{-\frac{\hbar^2\nabla^2}{2M}+V(\mathbf{r})+\mu_0\mathbf{B}(\mathbf{r})\cdot\boldsymbol{\sigma}\right\}\Psi(\mathbf{r})=E\Psi(\mathbf{r}), \tag{1}$$

where $\boldsymbol{\sigma}$ are the Pauli matrices, E the eigenvalues, and $\Psi(\mathbf{r})$ is a two component spinor wave function. We shall confine our discussion to the special case in which both the magnetic field and the external potential only depend on one coordinate, say y. Under this circumstance the magnetic field can be written as follows:

$$\mathbf{B}(y)=B(y)\begin{bmatrix}\sin\theta(y)\\0\\\cos\theta(y)\end{bmatrix}, \tag{2}$$

where $B(y)$ and $\theta(y)$ are arbitrary functions. It is clear that, up to a constant component along the y direction, which we shall ignore in this article, Eq. (2) yields the most general field configuration compatible with the divergentless condition

$$\nabla\cdot\mathbf{B}=0 \tag{3}$$

On the other hand it should be noted that, except for the case of constant **B**, this field is not sourceless since its curl does not vanish. In the following pages we shall exhibit some classes of functions $V(y)$, $B(y)$, and $\theta(y)$ for which the eigen-

[a)] Present address: Dept. of Physics and Astronomy, University of Maryland, College Park, MD 20742.

 0022-2488/83/030553-07$02.50

value equation can be solved in closed form. Clearly the motion of the particle in the x or z directions is trivial and can be easily taken into account.

With no loss of generality we may write

$$\theta(y) = 2(\kappa y + \delta(y)), \tag{4}$$

where κ is a real constant parameter and $\delta(y)$ is an arbitrary function. Let us then perform a unitary coordinate dependent transformation in spin space corresponding to a rotation around the y axis, which eliminates the $2\kappa y$ term of the magnetic field in Eq. (1). This transformation amounts to introduce a new curvilinear spin system which is rotated about the y axis by an angle $2\kappa y$. In what follows we shall always refer to this new system. The new spinor wave function, χ, is related to the original one by

$$\chi = \exp\{i\sigma_y \kappa y\}\Psi, \tag{5}$$

where σ_y is the Pauli matrix. The transformed eigenvalue equation becomes[1]

$$\left\{\frac{\hbar^2}{2M}\left(-i\frac{d}{dy} - \kappa\sigma_y\right)^2 + V(y) + \mu_0 B(y)(\sigma_x \sin 2\delta(y) + \sigma_z \cos 2\delta(y))\right\}\chi(y) = E\chi(y). \tag{6}$$

The explicit matrix form of the transformed Hamiltonian, $\mathscr{H}'$, is

$$\mathscr{H}' = \begin{bmatrix} -\lambda\frac{d^2}{dy^2} + V(y) + \lambda\kappa^2 + \mu_0 B(y)\cos 2\delta(y) & 2\kappa\lambda\frac{d}{dy} + \mu_0 B(y)\sin 2\delta(y) \\ -2\kappa\lambda\frac{d}{dy} + \mu_0 B(y)\sin 2\delta(y) & -\lambda\frac{d^2}{dy^2} + V(y) + \lambda\kappa^2 - \mu_0 B(y)\cos 2\delta(y) \end{bmatrix}, \tag{7}$$

where $\lambda = \hbar^2/2M$. Let us introduce the following operators

$$\hat{L}_\pm = \pm\frac{d}{dy} + \frac{\mu_0}{2\kappa\lambda} B(y)\sin 2\delta(y), \tag{8}$$

and set

$$\delta(y) = \tan^{-1}\rho(y). \tag{9}$$

We will now restrict the functions $V(y)$, $B(y)$ and $\rho(y)$ [or $\delta(y)$] so that they satisfy the following relationships:

$$V(y) = \frac{1}{\lambda}\left(\frac{\mu_0}{2\kappa}\right)^2 B^2(y)\sin^2 2\delta(y) + C_1 = \frac{1}{\lambda}\left(\frac{\mu_0}{\kappa}\right)^2\left(\frac{B(y)\rho(y)}{1+\rho^2(y)}\right)^2 + C_1, \tag{10}$$

and

$$B(y)\cos 2\delta(y) = \frac{B(y)(1-\rho^2(y))}{1+\rho^2(y)} = \frac{1}{\kappa}\frac{d}{dy}\left(\frac{B(y)\rho(y)}{1+\rho^2(y)}\right) + \frac{C_2}{\mu_0} = \frac{1}{2\kappa}\frac{d}{dy}(B(y)\sin 2\delta(y)) + \frac{C_2}{\mu_0}, \tag{11}$$

where C_1 and C_2 are arbitrary constants. Clearly these relations will leave us with only one arbitrary function. Once any of the functions $V(y)$, $B(y)$, or $\rho(y)$ is chosen the other two will be determined up to arbitrary constants.

The basic reason for imposing these restrictions on the functions V, B and ρ comes from the fact that in this case the Hamiltonian $\mathscr{H}'$ can be re-expressed in terms of the operators $\hat{L}_+$ and $\hat{L}_-$ of Eq. (8) as follows:

$$\mathscr{H}' = \begin{bmatrix} \lambda\hat{L}_+\hat{L}_- + \lambda\kappa^2 + C_1 + C_2 & 2\lambda\kappa\hat{L}_+ \\ 2\lambda\kappa\hat{L}_- & \lambda\hat{L}_-\hat{L}_+ + \lambda\kappa^2 + C_1 - C_2 \end{bmatrix}. \tag{12}$$

Under these conditions the solution to the corresponding eigenvalue equation can be enormously simplified. In fact, the original problem of solving the system of coupled differential equations can be circumvented by first solving an ordinary second order differential eigenvalue equation.[2] This statement is easily verified by observing that the Hermitian operator $\hat{0}^2$, where

$$\hat{0} = \begin{pmatrix} 0 & \hat{L}_+ \\ \hat{L}_- & 0 \end{pmatrix} \tag{13}$$

commutes with $\mathscr{H}'$ and consequently both can be simultaneously diagonalized. Before computing the energy eigenvalues and their eigenfunctions we will establish some important properties of the operators $\hat{L}_+$ and $\hat{L}_-$.

Let us write the operators $\hat{L}_\pm$ in a somewhat more compact form

$$\hat{L}_\pm = \pm\frac{d}{dy} + f(y), \tag{14}$$

then we easily derive that [assuming $f(y)$ is real]

i) $\hat{L}_+ = \hat{L}^\dagger_-$,

ii) $[\hat{L}_+, \hat{L}_-] = 2f'(y)$,

iii) $[\hat{L}_+\hat{L}_-, \hat{L}_-] = 2f'(y)\hat{L}_-$,

iv) $[\hat{L}_-\hat{L}_+, \hat{L}_+] = -2f'(y)\hat{L}_+$. (15)

Moreover we can also prove the following theorems:

Theorem I: the spectrum of the operators $\hat{L}_+\hat{L}_-$ and $\hat{L}_-\hat{L}_+$ is positive.

Proof: Let $\phi(y)$ be a square integrable function with continuous second derivative in some appropriate range of y. Then

$$\int \phi^*(y)\hat{L}_+\hat{L}_-\phi(y)dy = \int (\hat{L}_-\phi(y))^\dagger(\hat{L}_-\phi(y))dy \geq 0. \tag{16}$$

Thus if $\phi(y)$ is an eigenfunction of $\hat{L}_+\hat{L}_-$ its eigenvalue must be positive or zero. A similar proof holds for $\hat{L}_-\hat{L}_+$.

Theorem II: if $\phi^{(+)}(y)$ is an eigenfunction of $\hat{L}_+\hat{L}_-$ then

$\hat{L}_-\phi^{(+)}(y)$ is an eigenfunction of $\hat{L}_-\hat{L}_+$. Similarly if $\phi^{(-)}(y)$ is an eigenfunction $\hat{L}_-\hat{L}_+$ then $\hat{L}_+\phi^{(-)}(y)$ is an eigenfunction of $\hat{L}_+\hat{L}_-$.

Proof: Let $\phi^{(+)}(y)$ be an eigenstate of $\hat{L}_+\hat{L}_-$ with eigenvalue ζ^2, then the function $\hat{L}_-\phi^{(+)}(y)$ is eigenfunction of $\hat{L}_-\hat{L}_+$ with the same eigenvalue. From Eq. (15 iii) we derive that

$$[\hat{L}_+\hat{L}_-,\hat{L}_-]\phi^{(+)} = \hat{L}_+\hat{L}_-\hat{L}_-\phi^{(+)} - \zeta^2\hat{L}_-\phi^{(+)} = 2f'\hat{L}_-\phi^{(+)}, \tag{17}$$

therefore

$$(\hat{L}_+\hat{L}_- - 2f')\hat{L}_-\phi^{(+)} = \hat{L}_-\hat{L}_+\hat{L}_-\phi^{(+)} = \zeta^2\hat{L}_-\phi^{(+)}, \tag{18}$$

as stated. In a similar fashion we can prove that if $\phi^{(-)}$ is an eigenstate of $\hat{L}_-\hat{L}_+$ with eigenvalue $\bar{\zeta}^2$ then $\hat{L}_+\phi^{(-)}$ is eigenstate of $\hat{L}_+\hat{L}_-$ with the same eigenvalue. An immediate corolary of this theorem is that the spectrum of $\hat{L}_+\hat{L}_-$ and $\hat{L}_-\hat{L}_+$ coincide.

Let us finally prove that except for the subspace which corresponds to a possible zero eigenvalue the correspondence between the eigenstates of $\hat{L}_+\hat{L}_-$ and $\hat{L}_-\hat{L}_+$ is one-to-one.

Theorem III: let $\phi_1^{(+)}(y)...\phi_j^{(+)}(y)$ be a set of degenerate but linearly independent eigenfunctions of $\hat{L}_+\hat{L}_-$ with eigenvalue ζ^2. Then the functions $\hat{L}_-\phi_1^{(+)}(y),...,\hat{L}_-\phi_j^{(+)}(y)$ are also a set of linearly independent eigenfunctions of $\hat{L}_-\hat{L}_+$, provided $\zeta^2 > 0$.

Proof: Let us assume that the transformed eigenfunctions are not linearly independent. Then there must exist a combination

$$a_1\hat{L}_-\phi_1^{(+)} + a_2\hat{L}_-\phi_2^{(+)} + ... + a_j\hat{L}_-\phi_j^{(+)} = 0, \tag{19}$$

for some proper choice of constants $a_1,...,a_j$. Multiplying Eq. (19) by $\hat{L}_+$ we obtain

$$\zeta^2(a_1\phi_1^{(+)} + ... + a_j\phi_j^{(+)}) = 0, \tag{20}$$

but since $\zeta^2 > 0$, we conclude that the set of transformed functions must also be linearly independent. A similar proof also holds for any set of transformed degenerate but linearly independent eigenstates of $\hat{L}_-\hat{L}_+$, corresponding to some nonvanishing eigenvalue. It will be shown in Sec. II that for the zero eigenvalue this correspondence between eigenfunctions is not in general one-to-one.

The computation of the eigenfunctions and energy eigenvalues of $\mathscr{H}'$ is then straightforward. Let $\phi_\zeta^{(+)}$ and $\phi_\zeta^{(-)}$ be eigenstates of $\hat{L}_+\hat{L}_-$ and $\hat{L}_-\hat{L}_+$, respectively, with the same eigenvalue ζ, we can then write $\hat{L}_+\phi_\zeta^{(-)} = C_+\zeta\phi^{(+)}$ and $\hat{L}_-\phi_\zeta^{(-)} = C_-\zeta\phi_\zeta^{(-)}$ with $C_+C_- = 1$.

For simplicity let us assume that both functions are normalized. Then the normalized energy eigenfunction of $\mathscr{H}'$ corresponding to ζ can be written as

$$\chi_{+\zeta}(y) = \begin{bmatrix} \phi_\zeta^{(+)}(y)\cos\alpha \\ \phi_\zeta^{(-)}(y)\sin\alpha \end{bmatrix} \text{ and}$$

$$\chi_{-\zeta}(y) = \begin{bmatrix} -\phi_\zeta^{(+)}(y)\sin\alpha \\ \phi_\zeta^{(-)}(y)\cos\alpha \end{bmatrix}, \tag{21}$$

with a proper choice of the parameter $\alpha = \alpha(\zeta)$. Substituting these expressions into Eq. (6) one obtains

$$E_\pm(\zeta) = \lambda\zeta^2 + \lambda\kappa^2 + C_1 \pm [(2\lambda\kappa\zeta)^2 + C_2^2]^{1/2} \tag{22}$$

and $(\alpha(\zeta) = \alpha^\pm(\zeta))$,

$$\tan\alpha^+(\zeta) = \frac{2\kappa\lambda\zeta}{C_2 + [(2\lambda\kappa\zeta)^2 + C_2^2]^{1/2}} = -\cot\alpha^-(\zeta). \tag{23}$$

Finally we note that if the eigenfunctions of $\phi_\zeta^{(+)}(y)$ are chosen real, the eigenfunctions corresponding to $\mathscr{H}'$ will also be real and consequently they can be rewritten as

$$\chi_{\pm\zeta}(y) = [(\cos\alpha^\pm\phi_\zeta^{(+)}(y))^2 + (\sin\alpha^\pm\phi_\zeta^{(-)}(y))^2]^{1/2} \times \exp\{i\sigma_y\Phi_\zeta^\pm(y)\}\eta_\pm, \tag{24}$$

where

$$\eta_+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \eta_- = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{25}$$

with

$$\tan\Phi_\zeta^\pm(y) = \tan\alpha^\pm(\zeta)\frac{\phi_\zeta^{(-)}(y)}{\phi_\zeta^{(+)}(y)}. \tag{26}$$

It is clear from our previous discussion that any system for which the operator $\hat{L}_+\hat{L}_-$ (or either $\hat{L}_-\hat{L}_+$) has known solution, the corresponding Pauli–Schrodinger equation can also be solved. However in order to obtain the potential energy term associated to the operator $\hat{L}_-\hat{L}_+$ we must integrate a Riccati differential equation whose solution is not generally known (see the Appendix). In the next section we shall present examples where the solution of Eq. (6) can be obtained in closed form.

II. EXACTLY SOLVABLE SYSTEMS

In this section we shall make use of Eq. (10) and Eq. (11) to derive some explicit cases of solvable systems.

i) The simplest nontrivial case corresponds to choosing ρ = constant and $B = B_0$ = constant. This example represents a neutral spinning particle moving in a helical magnetic field with period $2\pi/\kappa$. A detailed study of this case has been given by one of the authors.[3] Here we will only outline the main properties of the solution. The external potential is in this case constant and can be ignored. The corresponding energy spectrum is

$$E_\pm(k) = \lambda\{k^2 + \kappa^2 \pm [(2k\kappa)^2 + (\mu_0 B_{0/\lambda})^2]^{1/2}\}, \tag{27}$$

and the eigenfunctions are

$$\psi_{\pm,k}(y) = \frac{1}{\sqrt{2\pi}} e^{iky}e^{-i\sigma y^{\kappa y}}e^{-i\sigma x^{\theta(k)/2}}\eta_\pm, \tag{28}$$

where

$$\tan\frac{\theta(k)}{2} = \frac{2k\kappa\lambda}{\mu_0 B_0 + [(2k\kappa\lambda)^2 + (\mu_0 B_0)^2]^{1/2}}. \tag{29}$$

We note that the continuous parameter k takes the role here of the parameter ζ defined in Sec. I.

It is clear from Eq. (28) that the spin of the particle rotates following the magnetic field as a function of y, but forms an angle $\theta(k)$ with it for the $\psi_{+,k}$ state, and $\theta(k) + \pi$ for the $\phi_{-,k}$ state. Another remarkable property of this solution is the absence of energy gaps in the spectrum. It is a general feature of periodic potentials to give raise to energy gaps as the pseudomomementum (k in our case) approaches

the boundaries of the Brillouin zones. In this example we also have a periodic interaction, but the spectrum is a continuous function of k. The reader is refered to Ref. 3 for a detailed explanation of this effect.

ii) Another interesting example of solvable hamiltonian corresponds to the choice

$$B(y)\sin 2\delta(y) = \alpha y, \tag{30}$$

in Eq. (10), where α is an arbitrary real parameter. In this case one derives that (ignoring C_1)

$$V(y) = \frac{1}{\lambda}\left(\frac{\mu_0\alpha}{2\kappa}\right)^2 y^2. \tag{31}$$

Furthermore, Eq. (11) implies that

$$B(y)\cos 2\delta(y) = \frac{\alpha}{2\kappa} + \frac{C_2}{\mu_0},$$

from which we conclude that

$$B(y) = \left[(\alpha y)^2 + \left(\frac{\alpha}{2\kappa} + \frac{C_2}{\mu_0}\right)^2\right]^{1/2} \tag{32}$$

and

$$\cos 2\delta(y) = \frac{\alpha/\kappa + C_2/\mu_0}{[(\alpha y)^2 + (\alpha/2\kappa + C_2/\mu_0)^2]^{1/2}}. \tag{33}$$

For simplicity let us choose $C_2 = -\mu_0/2\kappa$ so that the above equations become

$$\mathbf{B}(y) = \alpha y\hat{i}, \tag{34}$$

where $\hat{i}$ is the unit vector along the rotating x axis.

This case represents a particle moving in a harmonic potential well whose spin interacts with a magnetic field which rotates uniformly around the y axis with period $2\pi/\kappa$, but whose magnitude grows linearly with y as $|\alpha y|$.

The operators $\hat{L}_+\hat{L}_-$ and $\hat{L}_-\hat{L}_+$ are then given by

$$\hat{L}_+\hat{L}_- = -\frac{d^2}{dy^2} + w^2y^2 + w$$

and

$$\hat{L}_-\hat{L}_+ = -\frac{d^2}{dy^2} + w^2y^2 - w, \tag{35}$$

where we denoted $\mu_0\alpha/2\kappa\lambda$ by w. These are the familiar harmonic oscillator Hamiltonians whose spectrum is given by $2w(n+1)$ and $2wn$ with $n = 0,1,...$, respectively.[4]
The corresponding normalized eigenfunctions are

$$\varphi_n(y) = \left(\frac{w}{\pi}\right)^{1/4}\frac{1}{\sqrt{2^n n!}}H_n((w)^{1/2}y)\exp\left\{-\frac{wy^2}{2}\right\}, \tag{36}$$

where the H_n are the standard Hermite polynomials.[5] The operators $\hat{L}_+$, $\hat{L}_-$ are related to the usual raising and lowering operators $a^\dagger$, a of the harmonic oscillator:

$$L_+ = (2w)^{1/2}a, \quad \text{and} \quad L_- = (2w)^{1/2}a^\dagger. \tag{37}$$

The eigenfunctions of the operator $\hat{0}^2$ are given by

$$\chi_{+,n}(y) = \begin{bmatrix}\cos\alpha_n\varphi_n(y)\\ \sin\alpha_n\varphi_{n+1}(y)\end{bmatrix} \quad \text{and}$$

$$\chi_{-,n}(y) = \begin{bmatrix}-\sin\alpha_n\varphi_n(y)\\ \cos\alpha_n\varphi_{n+1}(y)\end{bmatrix} \tag{38}$$

with α_n arbitrary and $n = 0,1,...$. There is an additional nondegenerate eigenfunction of zero eigenvalue

$$\chi_0(y) = \begin{bmatrix}0\\ \varphi_0(y)\end{bmatrix}. \tag{39}$$

Moreover it is easily checked that

$$\begin{aligned}\hat{L}_+\varphi_{n+1}(y) &= (2w(n+1))^{1/2}\varphi_n(y)\\ \hat{L}_-\varphi_n(y) &= (2w(n+1))^{1/2}\varphi_{n+1}(y).\end{aligned} \tag{40}$$

The corresponding energy eigenvalues and eigenfunctions are then the following:

$$\begin{aligned}E_{\pm,n} = {}& 2\lambda w(n+1) + \lambda\kappa^2 \pm [(\mu_0\alpha/2\kappa)^2\\ &+ 8\lambda^2\kappa^2 w(n+1)]^{1/2}\end{aligned} \tag{41}$$

and the eigenfunctions are those of Eq. (38) with $\alpha_n = \alpha_n^+$ given by

$$\begin{aligned}\tan\alpha_n^+ &= \frac{[8\lambda^2\kappa^2 w(n+1)]^{1/2}}{[(\mu_0\alpha/2\kappa)^2 + 8\lambda^2\kappa^2 w(n+1)]^{1/2} - \alpha\mu_0/2\kappa}\\ &= -\cot\alpha_n^-.\end{aligned} \tag{42}$$

In addition, we have the eigenstate given by Eq. (39) whose energy is

$$E_0 = \lambda\kappa^2 + \mu_0\alpha/2\kappa. \tag{43}$$

We observe in Eq. (41) that depending on the values of the parameters, the ground state may or may not correspond to $E_{-,0}$. In fact one easily checks that taking κ sufficiently large we can have

$$E_{-,1} < E_{-,0}. \tag{44}$$

In an extreme situation we could also get an accidental ground state degeneracy such that for the ground state energy we have

$$E_{-,i} = E_{-,j} \quad \text{for } i \neq j. \tag{45}$$

Finally we note that the eigenfunction $\chi_{\pm,n}$ can be rewritten in the following form:

$$\begin{aligned}\chi_{\pm,n}(y) = {}& [(\cos\alpha_n^\pm\varphi_n)^2 + \sin\alpha_n^\pm\varphi_{n+1})^2]^{1/2}\\ &\times\exp\{-i\sigma_y\Phi_n^\pm(y)\}\eta_\pm\end{aligned} \tag{46}$$

with

$$\tan\Phi_n^\pm(y) = \frac{\tan\alpha_n^\pm}{(2(n+1))^{1/2}}\frac{H_{n+1}((w)^{1/2}y)}{H_n((w)^{1/2}y)}. \tag{47}$$

Equations (46) and (47) have interesting physical consequences. In fact, they tell us how the spin of the particle behaves as a function of y for each eigenstate $\chi_{\pm,n}$. Because the zeros of the Hermite polynomials H_n, H_{n+1} are simple and alternate on the real axis, it is clear that the function H_{n+1}/H_n will have $n+2$ simple poles and $n+1$ simple zeros. One can easily conclude that the spin of the particle does not "follow" the magnetic field except for the eigenfunction χ_0. In the remaining cases it will perform $n+1$ more turns around the y axis than the magnetic field does. This is clearly a case in which the adiabatic hypothesis does not work.

iii) As the third example we will choose the scalar potential and the magnetic field in Eq. (10) and Eq. (11) as follows (we set $C_1 = C_2 = 0$ and $c_i = \bar{c}_i\mu_0/2\kappa\lambda$, $i = 1,2$)

$$V(y)=\lambda(\bar{c}_1\tan\alpha y+\bar{c}_2\cot\alpha y)^2=\frac{1}{\lambda}\left(\frac{\mu_0}{2\kappa}\right)^2 B^2(y)\sin^2 2\delta(y), \tag{48}$$

$$B^2(y)=(c_1\tan\alpha y+c_2\cot\alpha y)^2 + \frac{\alpha^2}{4\kappa^2}(c_1\sec^2\alpha y-c_2\csc^2\alpha y)^2, \tag{49}$$

and

$$\tan 2\delta(y)=\left(\frac{2\kappa}{\lambda}\right)\frac{c_1\tan\alpha y+c_2\cot\alpha y}{c_1\sec^2\alpha y-c_2\csc^2\alpha y}, \tag{50}$$

where $\alpha>0$, c_1 and c_2 are arbitrary constants. In this case the operators $\hat{L}_\pm$ are given by

$$\hat{L}_\pm=\pm\frac{d}{dy}+c_1\tan\alpha y+c_2\cot\alpha y. \tag{51}$$

Because of the poles of $\tan\alpha y$ and $\cot\alpha y$, the variable y is restricted to the interval $[0,\pi/2\alpha]$; between these singular points the functions $V(y)$ and $B(y)$ are well-behaved. On the other hand, $\tan 2\delta(y)$ tends to zero when $y\to 0$ or $y\to\pi/2\alpha$, and it will have a zero (or a pole) at $y=y_0$ given by

$$\cot^2\alpha y_0=c_1/c_2, \tag{52}$$

provided $c_1/c_2<0$ (or $c_1/c_2>0$). When $c_1/c_2<0$, the magnetic field points along the positive z direction at the origin ($y=0$). For $y>0$, it first rotates counterclockwise until its phase $2\delta(y)$ attains its maximum value $2\delta_1$ at $y=y_1$. For $y>y_1$, it starts rotating clockwise until its phase attains its minimum value $2\delta_2$ at $y=y_2$. For $y>y_2$ it again rotates counterclockwise until it reaches $\delta=0$ at $y=\pi/2\alpha$. When $c_1/c_2>0$, the magnetic field points along the positive z direction at $y=0$. At the pole position, $y=y_0$, given by (52), it becomes parallel to the positive x direction and finally when $y\to\pi/2\alpha$ it points along the negative z direction.

From now on and without loss of generality, we will restrict the parameters such that $c_1>\alpha$, $c_2>\alpha$. The operators $\hat{L}_+\hat{L}_-$ and $\hat{L}_-\hat{L}_+$ are given by

$$\hat{L}_+\hat{L}_-=-\frac{d^2}{dy^2}+\alpha^2c_1'(c_1'-1)\sec^2\alpha y + \alpha^2c_2'(c_2'-1)\csc^2\alpha y-(c_1-c_2)^2 \tag{53}$$

and

$$\hat{L}_-\hat{L}_+=-\frac{d^2}{dy^2}+\alpha^2c_1''(c_1''-1)\sec^2\alpha y + \alpha^2c_2''(c_2''-1)\csc^2\alpha y-(c_1-c_2)^2, \tag{54}$$

where the new constants are defined as follows

$$c_1'=\frac{c_1}{\alpha}+1,\quad c_2'=\frac{c_2}{\alpha},$$
$$c_1''=\frac{c_1}{\alpha},\quad c_2''=\frac{c_2}{\alpha}+1. \tag{55}$$

We have introduced them so as to cast $\hat{L}_+\hat{L}_-$ and $\hat{L}_-\hat{L}_+$ in the form of a Schrödinger operator with a Pöschl–Teller potential.[6] The eigenfunctions of the operator $\hat{O}^2$ are

$$\chi_{+,n}(y)=\begin{bmatrix}\cos\theta_n^+\psi_{+,n}(y)\\ \sin\theta_n^+\psi_{-,n}(y)\end{bmatrix},$$
$$\chi_{-,n}(y)=\begin{bmatrix}-\sin\theta_n^-\psi_{+,n}(y)\\ \cos\theta_n^-\psi_{-,n}(y)\end{bmatrix}, \tag{56}$$

where (A_n,B_n are normalization constants)

$$\Psi_{+,n}(y)=A_n(\sin\alpha y)^{c_2'}(\cos\alpha y)^{c_1'} .F(-n,c_1'+c_2'+n,c_2'+\tfrac{1}{2};\sin^2\alpha y), \tag{57}$$

and

$$\Psi_{-,n}(y)=B_n(\sin\alpha y)^{c_2''}(\cos\alpha y)^{c_1''} .F(-n,c_1''+c_2''+n,c_2''+\tfrac{1}{2};\sin^2\alpha y), \tag{58}$$

are normalized eigenfunctions of $\hat{L}_+\hat{L}_-$ and $\hat{L}_-\hat{L}_+$ respectively, with eigenvalue

$$\xi_n^2=\alpha^2(c_1'+c_2'+2n)^2-(c_1-c_2)^2, \tag{59}$$

$F(a,b,c;z)$ is the hypergeometric function and θ_n is an arbitrary constant. It is readily seen from Eq. (59) that there is no zero eigenvalue and that the spectra of $\hat{L}_+\hat{L}_-$ and $\hat{L}_-\hat{L}_+$ coincide. Invoking the recurrence relations for the hypergeometric functions one easily verifies the following relations

$$\hat{L}_+\Psi_{-,n}(y)=\frac{R_n\xi_n^2}{2\alpha A_n}\Psi_{+,n}(y), \tag{60}$$

and

$$\hat{L}_-\Psi_{+,n}(y)=\frac{2\alpha A_n}{R_n}\Psi_{-,n}(y), \tag{61}$$

with

$$R_n=\frac{B_n(c_2'+1/2)}{(c_1'+n-1/2)(c_2'+n+1/2)}. \tag{62}$$

The normalized energy eigenfunctions are given by Eq. (56) with ($\theta_n=\theta_n^+$)

$$\tan\theta_n^+=\frac{2\alpha A_n}{R_n\xi_n}=-\cot\theta_n^-, \tag{63}$$

and the corresponding energy eigenvalues being

$$E_{\pm,n}=\lambda\xi_n^2+\lambda\kappa^2\pm 2\kappa\lambda\xi_n. \tag{64}$$

The solution of Eq. (56) can be rewritten in the form

$$\chi_{\pm,n}(y)=[(\cos\theta_n^\pm\Psi_{+,n}(y))^2+(\sin\theta_n^\pm\Psi_{-,n}(y))^2]^{1/2} \times\exp\{i\sigma_y\Phi_n^\pm(y)\}\eta_\pm, \tag{65}$$

where

$$\tan\Phi_n^\pm(y)=\tan\theta_n^\pm\frac{\Psi_{-,n}(y)}{\Psi_{+,n}(y)}. \tag{66}$$

For the special case $n=0$, the spin at $y=0$ points parallel to (χ_+) or antiparallel (χ_-) to the z axis. When $y\to\pi/2\alpha$ the spin points antiparallel (χ_+) or parallel (χ_-) to the z axis after a rotation in π. For higher values of n, the exact behavior of the spin depends critically on the values of the parameters c_1, c_2 and α.

For instance, if $n=1$,

$$\frac{\Psi_{-,1}(y)}{\Psi_{+,1}(y)}=\frac{B_1}{A_1}\tan\alpha y\frac{1-\tau_1\sin^2\alpha y}{1-\tau_2\sin^2\alpha y}, \tag{67}$$

where

$$\tau_1=\frac{2(c_1+c_2+\alpha)}{2c_2+3\alpha},\quad \tau_2=\frac{2(c_1+c_2+\alpha)}{2c_2+\alpha}. \tag{68}$$

Because of the factor $\tan \alpha y$ in Eq. (67) there is a zero at $y = 0$ and a pole at $y = \pi/2\alpha$. If $c_1 = c_2 = c$ the second factor will produce a pole at $\pi/4\alpha$ and a zero. Both the pole and the zero cancel in the limit $c \to \infty$. Under the conditions stated in Eq. (53), both τ_1 and τ_2 are larger than unity and the second factor in Eq. (67) will produce a zero and a pole which, for special values of c_1, c_2 and α, may cancel against the pole of $\tan \alpha y$.

iv) As our last example we shall take[7]

$$\hat{L}_{\pm} = \pm \frac{d}{dy} + \frac{\mu_0}{2\kappa\lambda}(\bar{a} \tanh z + \bar{b}), \tag{69}$$

where $z = (y - cd)/d$ and $\bar{a}$, $\bar{b}$, c, d are arbitrary parameters restricted so that $(\mu_0/2\kappa\lambda)a^2 \pm a/d > 0$. The scalar potential and the magnetic field are given by (we ignore once more the constants C_1 and C_2 in Eq.(10) and Eq. (11) and set $a = \bar{a}(\mu_0/2\kappa\lambda)$, $b = \bar{b}(\mu_0/2\kappa\lambda)$

$$V(y) = \lambda(a \tanh z + b)^2, \tag{70}$$

$$B^2(y) = (a \tanh z + b)^2 + \frac{a^2}{4\kappa^2 d^2} \operatorname{sech}^4 z, \tag{71}$$

$$\tan 2\delta(y) = \frac{2\kappa d}{a} \frac{a \tanh z + b}{\operatorname{sech}^2 z}. \tag{72}$$

The bound state eigenfunctions and eigenvalues of $\hat{L}_+\hat{L}_-$ and $\hat{L}_-\hat{L}_+$ are

$$\Psi_{\pm,n}(z) = \frac{N_{\pm,n}\exp(-a_{n\pm}z)}{(\exp z + \exp(-z))^{b_{n\pm}}} \times F\left(-n, A_{\pm}, B_{\pm}; \frac{\exp(-z)}{\exp z + \exp(-z)}\right), \tag{73a}$$

where

$$A_+ = A_- - 2 = 2ad - n - 1,$$
$$B_{\pm} = a_{n\pm} + b_{n\pm} + 1, \tag{73b}$$

and

$$\xi^2_{+,n} = \xi^2_{-,(n+1)} = \frac{2a(n+1)}{d} - \frac{(n+1)^2}{d^2} + b^2 - \frac{a^2b^2d^2}{(ad - n - 1)^2}, \tag{74}$$

with $n = 0, 1\ldots$ up to that integer smaller than $ad(1-b) - 1$, and where $b_{n+} = b_{n-} - 1 = ad - n - 1$, $a_{n\pm} = abd^2/b_{n\pm}$. The eigenfunctions of $\hat{O}^2$ are

$$\chi_{n,+}(z) = \begin{bmatrix} \cos\theta_n \Psi_{+,n}(z) \\ \sin\theta_n \Psi_{-,n+1}(z) \end{bmatrix}, \tag{75a}$$

$$\chi_{n,-}(z) = \begin{bmatrix} -\sin\theta_n \Psi_{+,n}(z) \\ \cos\theta_n \Psi_{-,n+1}(z) \end{bmatrix}, \tag{75b}$$

with eigenvalue $\xi^2_n = \xi^2_{+,n}$ and where θ_n is an arbitrary parameter. The eigenfunction for $n = -1$ is

$$\chi_{-1}(z) = \frac{N_{-1}\exp(-bdz)}{(\exp z + \exp(-z))^{ad}} \eta_-, \tag{76}$$

with eigenvalue $\xi^2_{-1} = 0$.

Since the analysis is similar to the preceeding case, we will only present the results for the $n = 0$ states. The energy eigenfunctions are obtained from Eq. (75a) and Eq. (75b) with

$$\tan\theta_0^+ = \frac{\Delta}{\xi_0} = -\cot\theta_0^-, \tag{77}$$

and the corresponding eigenvalues are

$$E_{0\pm} = \lambda\xi_0^2 + \lambda\kappa^2 \pm 2\kappa\lambda\xi_0, \tag{78}$$

where

$$\Delta = \frac{N_{+,0}}{N_{-,1}} \frac{(2ad-1)(a+b-1/d)}{(ad-1)}. \tag{79}$$

The spinors in Eqs. (75a,b) can be rewritten as

$$\chi_{0,\pm} = [(\cos\theta_0^{\pm}\Psi_{+,0})^2 + (\sin\theta_0^{\pm}\Psi_{-,1})^2]^{1/2} \times \exp\{-i\Phi^{\pm}(z)\sigma_y\}\eta_{\pm}, \tag{80}$$

with

$$\tan\Phi^{\pm}(z) = \tan\theta_0^{\pm} \frac{N_{-,1}}{N_{+,0}} \frac{\{2bd + (ad-1)\tanh z\}}{\{ad + 2bd - 1\}}. \tag{81}$$

There is also a continuum spectrum that starts at $(a-b)^2$. We will not discuss these eigenfunctions since they do not add any new feature to the ones we have already described.

As a final remark we note that in the limit $\kappa \to 0$ the operators $\hat{O}^2$ and thus $\mathscr{H}$ are singular, except for the case when the magnetic field is helical with a constant magnitude (Example 1). This means that once the transformation to the rotating system has been performed one cannot undo it by letting κ go to zero.

APPENDIX

Let us suppose that the solution to the eigenvalue problem for the operator $\hat{L}_+\hat{L}_-$ is known

$$\hat{L}_+\hat{L}_-\Psi_\xi(y) = \left(\frac{-d^2}{dy^2} + f^2(y) + f'(y)\right)\Psi_\xi(y) = \left(\frac{-d^2}{dy^2} + U_+(y)\right)\Psi_\xi(y) = \xi\Psi_\xi(y), \tag{A1}$$

with

$$U_+(y) = f^2(y) + f'(y), \tag{A2}$$

and where $U_+(y)$ is a known function. From Theorem III we can immediately construct the eigenvalues and eigenfunctions corresponding to $\hat{L}_-\hat{L}_+$

$$\hat{L}_-\hat{L}_+ = \frac{-d^2}{dy^2} + f^2(y) - f'(y) = \frac{-d^2}{dy^2} + U_-(y). \tag{A3}$$

To obtain the function $U_-(y)$ we must first integrate Eq. (A2) to evaluate $f(y)$ and then compute

$$U_-(y) = U_+(y) - 2f(y). \tag{A4}$$

Equation (A2) is a Riccati equation whose solution is not generally known.

[1]We have not included the x and z dependence of the wave function, which is just a multiplicative exponential. Correspondingly the energy will be shifted by an appropriate kinetic term.

[2]L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951).

[3]M. Calvo, Phys. Rev. B **18**, 5073 (1978).

[4]Note that $\hat{L}_-\hat{L}_+$ has a nontrivial eigenfunction with eigenvalue zero namely $\varphi_0(y)$, whereas $\hat{L}_+\hat{L}_-$ has no nontrivial eigenfunction with eigenvalue zero. This example illustrates the exemption made in Theorem III of Sec. I.

[5]A. Messiah, *Mécanique Quantique* (Dunod, Paris, 1960), Vol. I, Ch. XII.

[6]S. Flugge, *Practical Quantum Mechanics* (Springer–Verlag, New York, 1974).

[7]D. E. McCumber and B. I. Halperin, Phys. Rev. B **1**, 1054 (1970). They solved a particular case of our example (iii) in the study of thin superconducting wires.

Irregular solutions of the s-wave radial Schrödinger equation for van der Waals potentials

W. J. Romo[a)]
Department of Physics, Carleton University, Ottawa, Canada K1S 5B6

(Received 24 September 1982; accepted for publication 5 November 1982)

Simple ideas based on quantum chromodynamics suggest that the strong nuclear force might possess a long-range van der Waals component. Such a component, if it exists, would be most apparent in low-energy interactions between pairs of hadrons, at least one of which is neutral. In this article, as the initial part of a program to set limits on the magnitude of the strength of such a van der Waals interaction, we construct low-energy series expansions of the two irregular solutions $f_0(\pm k,r)$ to the s-wave Schrödinger equation that describes the interaction of a pair of particles interacting through a potential with a van der Waals tail of the form $V(r)\sim -\lambda\hbar c r_0^{N-1}/r^N$, for arbitrary values of the potential parameters, λ, r_0, and N. Accurate approximation schemes based on the series expansions are developed, and the results are used to obtain an expression for the low-energy elastic cross section. The effects of the van der Waals component of this interaction are clearly evident in the energy dependence of the cross section.

PACS numbers: 03.65.Ge, 03.65.Nk

I. INTRODUCTION

A one-particle exchange potential has a range equal to the Compton wavelength of the particle being exchanged. Thus, the range of an interaction involving a number of exchange particles will be determined by the mass of the lightest exchange particle. For many years it was believed that the lightest mass particle involved in the strong interaction was the pion, but recent theories allow for multiple gluon exchange between hadrons. Since gluons are massless particles, the strong interaction may possess a long-range van der Waals component. Feinberg and Sucher[1] carried out a phenomenological analysis that placed limits on the coupling constant λ_N of a hypothetical strong van der Waals potential of the form $V_N(r)\sim(\lambda_N/r)\,(r_0/r)^{N-1}$. Their analysis was based on a number of sources including Etövös and Cavendish type experiments, hyperfine structure of the hydrogen molecule, and the level structure of exotic atoms. In an attempt to improve or provide independent supporting evidence for their limits, we shall employ low-energy neutron–nucleus scattering at low bombardment energies as an additional source. Since most of the information about low-energy scattering is contained in the s-wave radial wavefunction, it is essential to have accurate solutions to the s-wave radial wave equation for a potential with a van der Waals tail. In this paper we develop methods for calculating these wave functions, and in a subsequent paper[2] we apply these results to the problem of obtaining limits on $|\lambda_N|$ by means of an analysis of low energy neutron–nucleus scattering data.

The wavefunction we seek satisfies the s-wave radial wave equation

$$\frac{-d^2u_0(k,r)}{dr^2}+\frac{2m}{\hbar^2}V(r)u_0(k,r)=k^2u_0(k,r), \tag{1.1}$$

where $E=\hbar^2k^2/2m$ is the kinetic energy of relative motion of the two particles at infinite separation and m is their reduced mass. There are three solutions to Eq. (1.1) that we shall be particularly interested in, a regular solution $\phi_0(k,r)$ and two irregular solutions $f_0(\pm k,r)$. We shall follow Newton's[3] convention in defining these solutions. In addition to satisfying (1.1), $\phi_0(k,r)$ also satisfies the boundary condition

$$\lim_{r\to 0} r^{-1}\phi_0(k,r)=1, \tag{1.2}$$

while the two linearly independent irregular solutions $f_0(\pm k,r)$ satisfy the boundary conditions

$$\lim_{r\to\infty} e^{\pm ikr}f_0(\pm k,r)=1. \tag{1.3}$$

Since Eq. (1.1) is a linear second-order differential equation, it only has two linearly independent solutions for each value of k; hence $\phi_0(k,r)$ can be written as a linear combination of $f_0(k,r)$ and $f_0(-k,r)$, namely

$$\phi_0(k,r)=\{f_0(k)f_0(-kr)-f_0(-k)f_0(k,r)\}/2ik, \tag{1.4}$$

where $f_0(k)$ and $f_0(-k)$ are the s-wave Jost functions.

The physical solution to (1.1) must be regular at the origin; hence it must be proportional to $\phi_0(k,r)$. The coefficient of proportionality depends upon the boundary conditions imposed on the physical solution. One usually requires that the asymptotic form of the physical solution consists of an incident plane wave and outgoing spherical scattered waves. The s-wave radial component will then be given by[3]

$$u_0(k,r)=k\phi_0(k,r)/f_0(-k). \tag{1.5}$$

We shall consider the class of potentials $V(r)$ for which

$$V(r)=-(\lambda\hbar c/r)(r_0/r)^{N-1} \tag{1.6}$$

for $r>R$, with $R\gg r_0$, but for now we shall not specify the form of $V(r)$ for $r\leqslant R$. The form chosen for the van der Waals potential, (1.6), agrees with the convection used in Ref. 1. For their description of the strong van der Waals potential Feinberg and Sucher[1] set $r_0=1$ fm. Because of the boundary condition (1.2), one cannot construct the solution $\phi_0(k,r)$ without first specifiying the behavior of $V(r)$ in the region $0\leqslant r\leqslant R$, but one can construct the irregular solutions $f_0(\pm k,r)$ for all values of $r>R$ with the information given. If

a) Work supported by the Natural Sciences and Engineering Research Council of Canada.

 0022-2488/83/030560-13$02.50

one assumes that $V(r)$ is continuous or has a finite discontinuity at $r = R$, then $\phi_0(k,r)$ and its first derivative with respect to r are continuous at $r = R$. Consequently, if one specifies $V(r)$ in the integral region ($r \leqslant R$) and solves for $\phi_0(k,r)$ in this region, the Jost functions can be determined by equating the logarithmic derivative of $\phi_0(k,r)$ at $r = R$ to the right-hand side of (1.4) evaluated at $r = R$. Then, (1.4) and the irregular solutions will give us $\phi_0(k,r)$ in the external region $r > R$.

In Sec. II we shall contruct a series solution for $f_0(k,r)$ that is valid in the region of space $r > R$ for which $r < |k|^{-1}$. We will only consider the $N > 2$ cases, since strong van der Waals potentials with $N = 1$ and 2 were ruled out.[1] Furthermore, well-known solutions exist for the $N = 1$ and 2 cases. In Sec. III we examine the properties of the expansion coefficients of that series and determine the behavior of $f_0(k,r)$ in the strong coupling (large $|\lambda|$) limit. In Sec. IV, an approximation scheme is developed that gives reasonable accuracy not only in the region of validity of the series solution and at large r, where a first-order Born calculation of $f_0(k,r)$ gives accurate results, but also in the transition region that bridges the gap between the two regions. In Sec. V an expression for the cross section is developed that clearly indicates the effects of the van der Waals interaction on low energy scattering. Comments and conclusions are given in Sec. VI.

II. LOW-ENERGY EXPANSION OF $f_0(k,r)$

The irregular solution $f_0(k,r)$ can be expressed as an infinite sum[3]

$$f_0(k,r) = \sum_{j=0}^{\infty} f_0^{(j)}(k,r), \tag{2.1}$$

where

$$f_0^{(0)}(k,r) = e^{-ikr} \tag{2.2}$$

and

$$f_0^{(j)}(k,r) = -\frac{2m}{\hbar^2 k}\int_r^{\infty} \sin(k(r-r'))V(r')f_0^{(j-1)}(k,r')\,dr'. \tag{2.3}$$

Since the potential $V(r)$ is given by (1.6) for $r > R$, if one defines two new constants M and Λ, a new variable z, and a new function $F(\Lambda,z)$ by

$$M = N - 2, \tag{2.4a}$$

$$\Lambda = 2m\lambda c(2ikr_0)^M r_0 \hbar^{-1}, \tag{2.4b}$$

$$z = 2ikr, \tag{2.4c}$$

$$F(\Lambda,z) = e^{ikr} f_0(k,r), \tag{2.4d}$$

then Eqs. (1.1)–(1.3) may be replaced by

$$F(\Lambda,z) = \sum_{j=0}^{\infty} \Lambda^j F^{(j)}(z), \tag{2.5}$$

$$F^{(0)}(z) = 1, \tag{2.6}$$

$$F^{(j)}(z) = \int_z^{\infty} (\exp(z-\zeta) - 1)\zeta^{-N} F^{(j-1)}(\zeta)\,d\zeta. \tag{2.7}$$

In carrying out the substitutions given by Eqs. (2.4a)–(2.4d), one finds that the integral on the right-hand side of (2.7) runs from $\zeta = z$ to $\zeta = \infty$ along the contour C_z given by $\arg\zeta' = \arg z$. In Appendix A it is shown that the contour of integration can be taken to run from $\zeta = z$ to $\zeta = +\infty$ along any simple path that does not encircle the origin.

Taking $j = 1$ in (2.7) and using (2.6) gives

$$\begin{aligned} F^{(1)}(z) &= e^z \int_z^{\infty} e^{-\zeta}\zeta^{-N}\,d\zeta - \int_z^{\infty} \zeta^{-N}\,d\zeta \\ &= z^{1-N}\{e^z E_N(z) - (N-1)^{-1}\}, \end{aligned} \tag{2.8}$$

where $E_N(z)$ is an exponential integral.[4] If one uses series expansions[4] for $E_N(z)$ and e^z, this becomes

$$F^{(1)}(z) = z^{-M}\left\{ -\sum_{m=0}^{M-1}(-z)^m \sum_{j=0}^{m+1}(-1)^j(j!(m+1-j)!(j+M-m))^{-1} + \sum_{m=M}^{\infty} z^m\left(\frac{(-1)^{M+1}\psi(M+2)}{(M+1)!(m-M)!} - (-1)^m \times \sum_{j=0}^{m+1}\frac{(-1)^j}{j!(m+1-j)!(j+M-m)}\right) - \frac{(-1)^{M+1}}{(M+k)!}\ln z \sum_{m=M}^{\infty}\frac{z^m}{(m-M)!}\right\}, \tag{2.9}$$

where M was defined by Eq. (2.4a),

$$\psi(n) = -\gamma + \sum_{m=1}^{n-1} m^{-1},$$

and γ is Euler's constant, $\gamma = 0.577\,21\ldots$.

Having $F^{(1)}(z)$, one can compute $F^{(2)}(z)$ by Eq. (2.7), and by repeated application of (2.7) one can construct $F^{(n)}(z)$ for arbitrary n. A few iterations reveals that $F^{(n)}(z)$ has the general form

$$F^{(n)}(z) = (z^{-M})^n \sum_{L=0}^{n}\sum_{m=LM}^{\infty} (\ln z)^L z^m \beta_m^{(n)}(L). \tag{2.10}$$

Comparison of (2.10) with (2.6) and (2.9) reveals that

$$\begin{aligned} \beta_m^{(1)}(0) &= (-1)^{m+1}\sum_{j=0}^{m+1}(-1)^j\{j!(m+1-j)!(k+M-m)\}^{-1}, \quad m < M, \\ &= \frac{(-1)^{M+1}\psi(M+2)}{(M+1)!(m-M)!} - (-1)^m \sum_{\substack{j=0 \\ j\neq m-M}}^{m+1}\frac{(-1)^j}{(m+1-j)!j!(j+M-m)}, \quad m \geqslant M, \end{aligned} \tag{2.11a}$$

$$\beta_m^{(1)}(1)\begin{cases} = 0, & m < M, \\ = (-1)^M((M+1)!(m-M)!)^{-1}, & m \geqslant M, \end{cases} \tag{2.11b}$$

$$\beta_m^{(0)}(L) = \delta_{L0}\delta_{m0}. \tag{2.11c}$$

To verify that (2.10) is the correct form for $n > 1$, one evaluates (2.7) with $F^{(j-1)}(\zeta')$ on the right-hand side of the equation given by (2.10) with n replaced by $j-1$. This procedure not only confirms the validity of the induction hypothesis (2.10), but it also provides recurrence relations for the coefficients $\beta_m^{(j)}(L)$. One gets[5]

$$\beta_m^{(j)}(L) = \frac{1}{L!}\sum_{l=L}^{[m/M]} l! \sum_{k=0}^{m+1-lM} \frac{(jM+k-m)}{k!}^{L-l-1} \sum_{\substack{n=0 \\ (n\neq -k)}}^{m+1-lM-k} \frac{(-1)^n}{n!}\beta_{m+1-n-k}^{(j-1)}(l) \tag{2.12a}$$

for m such that $LM \leqslant m \leqslant jM$ $(L < j)$, and

$$\begin{aligned}\beta_m^{(j)}(L) = {} & \frac{1}{L!}\sum_{l=L}^{j-1} l! \sum_{\substack{k=0 \\ (k\neq m-jM)}}^{m+1-lM} \frac{(jM+k-m)}{k!}^{L-l-1} \sum_{\substack{n=0 \\ (n\neq -k)}}^{m+1-lM-k} \frac{(-1)^n}{n!}\beta_{m+1-n-k}^{(j-1)}(l)(1-\delta_{L,j}) \\ & + (C_1^{(j)} - C_2^{(j)}\delta_{m,jM}[(m-jM)!]^{-1}\delta_{L,0} \\ & - \frac{1}{L(m-jM)!}\sum_{\substack{n=0 \\ (n\neq jM-m)}}^{(j-L+1)M+1} \frac{(-1)^n}{n!}\beta_{jM+1-n}^{(j-1)}(L-1)(1-\delta_{L,0})\end{aligned} \tag{2.12b}$$

for $m \geqslant jM$, where $C_n^{(j)}$ with $n = 1$ and 2 are constants of integration given by

$$C_1^{(j)} = \int_z^\infty e^{-\zeta} F^{(j-1)}(\zeta)\zeta^{-M-2}\,d\zeta + \mathscr{F}_1^{(j)}(z) \tag{2.13a}$$

and

$$C_2^{(j)} = \int_z^\infty F^{(j-1)}(\zeta)\zeta^{-M-2}\,d\zeta + \mathscr{F}_2^{(j)}(z). \tag{2.13b}$$

$\mathscr{F}_n^{(j)}(z)$ is an indefinite integral of the integrand appearing on the right-hand side of (2.13a) for $n = 1$, or of (2.13b) for $n = 2$. Explicitly,

$$\begin{aligned}\mathscr{F}_1^{(j)}(z) = {} & -\sum_{L=0}^{j-1}\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\sum_{\substack{m=LM \\ (m\neq jM+1-n)}}^{\infty}\beta_m^{(j-1)}(L)\frac{z^{-jM+m+n-1}L!}{(jM+1-m-n)^{L+1}}\sum_{l=0}^{L}\frac{\{(jM+1-m-n)\ln z\}^l}{l!} \\ & + \sum_{L=0}^{j-1}\sum_{n=0}^{(j-L)M+1}\frac{(-1)^n}{n!}\beta_{jM+1-n}^{(j-1)}(L)\frac{\{(\ln z)\}^{L+1}}{L+1},\end{aligned} \tag{2.14a}$$

$$\begin{aligned}\mathscr{F}_2^{(j)}(Z) = {} & -\sum_{L=0}^{j-1}\sum_{\substack{m=LM \\ m\neq jM+1}}^{\infty}\beta_m^{(j-1)}(L)\frac{z^{-jM+m-1}}{(jM+1-m)^{L+1}}L!\sum_{l=0}^{L}\frac{\{(jM+1-m)\ln(z)\}^l}{l!} \\ & + \sum_{L=0}^{j-1}\beta_{jM+1}^{(j-1)}(L)\frac{(\ln z)^{L+1}}{L+1}.\end{aligned} \tag{2.14b}$$

Since the integrals on the right-hand side of (2.13a) and (2.13b) vanish as $z \to \infty$, one has

$$C_n^{(j)} = \lim_{z\to\infty}\mathscr{F}_n^{(j)}(z). \tag{2.15}$$

The values of the $C_n^{(j)}$ for a few of the small values of j are computed in Appendix B. In (2.12a) and(2.12b) and throughout the rest of this paper, we adopt the convention that any sum whose upper index is smaller than its lower index must be set equal to zero.

An examination of (2.12)–(2.15) reveals that the coefficients $\beta_m^{(j)}(L)$ and $C_n^{(j)}$ are independent of Λ and z, which implies that they are also independent of the energy, the radius, and the strength λ of the potential (however, they do depend upon the form of the potential through the constant $M = N-2$). Thus, a single set of coefficients $\beta_m^{(j)}(L)$ will enable one to calculate $f_0(k,r)$ for arbitrary λ, k, and r [cf. (2.10)].

The low-energy expansion of $f_0(k,r)$ can now be obtained by combining (2.4d), (2.5), (2.10), and the Taylor expansion of the exponential function e^{-ikr}. Combining terms of equal order in k, one gets

$$f_0(k,r) = \sum_{L=0}^{\infty}(\Lambda\ln z)^L\sum_{m=0}^{\infty}z^m\mathscr{A}_{L,m}(\xi), \tag{2.16}$$

where

$$\mathscr{A}_{L,m}(\xi) = \sum_{j=0}^{\infty}\xi^j b_{LM+m}^{(L+j)}(L), \tag{2.17}$$

$$\xi = 2m\lambda c r_0\hbar^{-1}(r_0/r)^M, \tag{2.18}$$

$$b_{LM+m}^{(j)}(L) = \sum_{n=0}^{m}\left(-\frac{1}{2}\right)^n\beta_{LM+m-n}^{(j)}(L)/n!, \tag{2.19}$$

and where Λ and z are again given by (2.4b) and(2.4c), respectively. The functions $\mathscr{A}_{L,m}(\xi)$ are independent of k, and by Eqs. (2.17) and (2.18) it can be seen that they fall off rapidly with r. The expansion coefficients $b_{LK+M}^{(j)}(L)$ can easily be computed once the set of coefficients $\beta_\mu^{(j)}(L)$ are known for all μ in the range $LM \leqslant \mu \leqslant LM+m$. Furthermore, the coefficients $b_m^{(j)}(L)$ like the coefficients $\beta_m^{(j)}(L)$ are independent of both k and λ.

In the next section we shall examine these coefficients in more detail.

III. THE EXPANSION COEFFICIENTS

An examination of the recurrence relations (2.12a) and (2.12b) reveals that the coefficients $\beta^{(j)}_m(L)$ vanish for $j<L$ or $m<LM$ for each fixed value of L. Then from (2.19) it follows that $b^{(j)}_m(L)$ vanishes for the same range of j and M. Since we are interested in a low-energy expansion of $f_0(k,r)$, we shall be most interested in coefficients $b^{(j)}_m(L)$, for which m and L are small. In general, we shall need these coefficients for all $j\geqslant L$, but, as was mentioned in the previous section, if r is sufficiently large or λ sufficiently small, the series defining $\mathscr{A}_{L,m}(\zeta)$ converges rapidly and so only a few of the terms should suffice. Since $b^{(j)}_m(L)$ is the coefficient of a term of order $k^m(\ln k)^L$ in the expansion of $f_0(k,r)$, if one wishes to determine the small k expansion of $f_0(k,r)$ up to terms of order k^μ, for some integer μ, he must calculate all coefficients $b^{(j)}_m(L)$ with L in the range $0\leqslant L\leqslant[m/M]$ and m in the range $0\leqslant m\leqslant\mu$. From (2.19) it can be seen that these coefficients can be constructed if one can calculate the set of coefficients $\beta^{(j)}_m(L')$ with $0\leqslant L'\leqslant[m'/M]$ and $0\leqslant m'\leqslant\mu$. Inspection of the recurrence relations then shows that the set of coefficients $\{\beta^{(j)}_m(L);\ 0\leqslant L\leqslant[m/M],\ 0\leqslant m\leqslant\mu\}$ will suffice in determining the set $\{\beta^{(j+1)}_m(L);\ 0\leqslant L\leqslant[m/M],\ 0\leqslant m\leqslant\mu\}$, which in turn will determine the coefficients $b^{(j+1)}_m(L)$ over the desired range of L and m. Hence, in constructing the needed coefficients one need not construct any coefficients associated with terms smaller than order k^μ. The necessary coefficients can easily be obtained by computer calculations, but it is instructive to examine a few general algebraic solutions.

We shall obtain closed form expressions for $\beta^{(j)}_{LM+n}(L)$ with $n=0$ and 1. In the $n=0$ case with $j>L$, (2.12a) applies and one obtains

$$\beta^{(j)}_{LM}(L)=-\beta^{(j-1)}_{LM}(L)M^{-2}(j-L)^{-1}(j-L+M^{-1})^{-1}, \tag{3.1}$$

while in the $n=1$ case with $j>L+M^{-1}$ and $M>1$, (2.12a) again applies and one gets[6]

$$\beta^{(j)}_{LM+1}(L)=\frac{\beta^{(j-1)}_{LM}(L)}{M^3(j-L-M^{-1})(j-L)(j-L+M^{-1})}-\frac{\beta^{(j-1)}_{LM+1}(L)}{M^2(j-L)(j-L-M^{-1})}. \tag{3.2}$$

From Eq. (2.19) one finds

$$b^{(j)}_{LM}(L)=\beta^{(j)}_{LM}(L), \tag{3.3}$$

$$b^{(j)}_{LM+1}(L)=\beta^{(j)}_{LM+1}(L)-\tfrac{1}{2}\beta^{(j)}_{LM}(L). \tag{3.4}$$

Combining (3.2) with (3.4) yields

$$b^{(j)}_{LM+1}(L)=-[M^2(j-L)(j-L-M^{-1})]^{-1}b^{(j-1)}_{LM+1}(L), \tag{3.5}$$

for $j>L+M^{-1}$. Iterating Eqs. (3.1) and (3.5) and using Eq. (3.3) then gives

$$b^{(L+j)}_{LM}(L)=\frac{(-1)^j\Gamma(1+M^{-1})}{M^{2j}j!\Gamma(j+1+M^{-1})}\beta^{(L)}_{LM}(L), \tag{3.6}$$

for all $M\geqslant1$ and $j\geqslant0$, and

$$b^{(L+j)}_{LM+1}(L)=\frac{(-1)^j\Gamma(1-M^{-1})}{M^{2j}j\Gamma(j+1-M^{-1})}b^{(L)}_{LM+1}(L) \tag{3.7}$$

for all $M\geqslant2$ and $j\geqslant0$. To complete the analysis, we must compute $\beta^{(L)}_{LM}(L)$ and $b^{(L)}_{LM+1}(L)$. If j is taken to be L, and m is taken to be $LM+m$ in Eq. (2.12b), one gets

$$\beta^{(L)}_{LM+M}(L)=-\frac{1}{L!M!}\sum_{\substack{n=0\\(n\neq-m)}}^{M+1}\frac{(-1)^n}{n!}\beta^{(L-1)}_{(L-1)M+M+1-n}(L-1). \tag{3.8}$$

The set of equations given by Eqs. (2.11c) and (3.8) with L any integer greater than zero has the solution

$$\beta^{(L)}_{LM+m}(L)=\frac{1}{L!\{(M+1)!\}^L m!}\times\begin{cases}(-1)^{LM/2}\delta_{m,0} & \text{for } L \text{ even},\\ (-1)^{(L+1)M/2} & \text{for } L \text{ odd}.\end{cases} \tag{3.9}$$

Equation (3.9) is also valid for $L=0$ [cf. (2.11c)]. Finally, if we combine Eqs. (3.3), (3.4), (3.6), (3.7), and (3.9), we find that

$$b^{(L+j)}_{LM}(L)=\frac{(-1)^{M[(L+1)/2]}\Gamma(1+M^{-1})(-1)^j}{M^{2j}j!L!\Gamma(j+1+M^{-1})\{(M+1)!\}^L}, \tag{3.10}$$

for all $j\geqslant L$, $L\geqslant0$, and $M\geqslant1$, and also

$$b^{(L+j)}_{LM+1}(L)=\frac{(-1)^{L-1+M[(L+1)/2]}\Gamma(1-M^{-1})(-1)^j}{2M^{2j}j!L!\Gamma(j+1-M^{-1})\{(M+1)!\}^L} \tag{3.11}$$

for all $j\geqslant L$, $L\geqslant0$, and $M\geqslant2$, The square brackets in the exponent of -1 in (3.10) and (3.11) again signifies the greatest integer less than or equal to the argument, in this case $(L+1)/2$. Combining (2.17), (3.10), or (3.11) with (9.1.10) of Ref. 4 gives

$$\mathscr{A}_{L,0}(\xi)=\frac{(-1)^{M[(L+1)/2]}\Gamma(1+M^{-1})}{L!\{(M+1)!\}^L(\sqrt{\xi}/M)^{1/M}}J_{1/M}(2\sqrt{\xi}/M), \tag{3.12}$$

$$\mathscr{A}_{L,1}(\xi)=-\frac{(-1)^{M[(L+1)/2]}\Gamma(1+M^{-1})}{2L!\{(M+1)!\}^L(\sqrt{\xi}/M)^{-1/M}}\times J_{-1/M}(2\sqrt{\xi}/M), \tag{3.13}$$

where $J_\nu(z)$ is a Bessel Function.[4] Equation (3.12) holds for $M\geqslant1$ and (3.13) for $M\geqslant2$. From (2.4c), (2.16), (3.12), and (3.13) one finds that

$$f_0(k,r)=\frac{(1+M^{-1})J_{1/M}(2\sqrt{\xi}/M)}{(\sqrt{\xi}/M)^{1/M}}-\frac{ikr\Gamma(1-M^{-1})J_{-1/M}(2\sqrt{\xi}/M)}{(\sqrt{\xi}/M)^{-1/M}}+o(kr) \tag{3.14a}$$

for $N+M+2>4$, and

$$f_0(k,r)=\{J_1(2\sqrt{\xi})/\sqrt{\xi}\}\{1-ikr\xi\ln(2ikr)\}+o(kr\ln(kr)) \tag{3.14b}$$

for[7] $N=3$. If one substitutes the first two terms of the expansion for $f_0(k,r)$ given by (3.14a) or (3.14b) into the s-wave Schrödinger equation (1.1) and employs (1.6) and (3.18), it is found that they provide a solution to (1.1) up to terms that

vanish faster than k for the $N>3$ cases, or faster than $k\ln k$ for the $N=3$ case; in fact, the leading term is an exact solution to (1.1) and (1.3) when $k=0$. Equation (3.14a) is especially simple when $N=4$, or equivalently $M=2$. Using definitions for the Bessel Functions $J_{\pm\frac{1}{2}}$ given in Abramowitz and Stegun,[4] one finds that

$$f_0(k,r)=\sin\sqrt{\xi}/\sqrt{\xi}-ikr\cos\sqrt{\xi}+o(k,r),$$

if $M=2$. Equation (3.14a) may be used to find the asymptotic form of $f_0(k,r)$ when $M\geqslant 2$ and both k and r tend to zero, if one assumes that (1.6) holds for all r. From (9.1.10) of Ref. 4, one finds that, as k and r tend to zero,

$$f_0(k,r)\sim\frac{\Gamma(1+M^{-1})}{\sqrt{\pi}}\left(\frac{M}{\sqrt{\xi}}\right)^{1/2+1/M}\left\{\cos\left(\frac{2\sqrt{\xi}}{M}-\frac{\pi}{2M}-\frac{\pi}{4}\right)+O(|\xi|^{-1/2})\right\}$$
$$-\frac{ikr\Gamma(1+M^{-1})}{\sqrt{\pi}}\left(\frac{M}{\sqrt{\xi}}\right)^{1/2-1/M}\left\{\cos\left(\frac{2\sqrt{\xi}}{M}+\frac{\pi}{2M}-\frac{\pi}{4}\right)+O(|\xi|^{-1/2})\right\} \tag{3.15}$$

for an attractive van der Waals potential, $\lambda>0$, or

$$f_0(k,r)=\frac{1}{2\sqrt{\pi}}\left(\frac{M}{\sqrt{|\xi|}}\right)^{1/2}e^{2\sqrt{|\xi|}/M}\left\{\Gamma(1+M^{-1})\left(\frac{M}{\sqrt{|\xi|}}\right)^{1/M}\right.$$
$$\left.-ikr\Gamma(1-M^{-1})(M/\sqrt{|\xi|})^{-1/M}\right\}(1+O(|\xi|^{-1/2})) \tag{3.16}$$

for a repulsive van der Waals potential, $\lambda>0$. Since $|\xi|\to\infty$ as $r\to 0$, the cosine term dominates the asymptotic behavior in (3.15) and the exponential term dominates in (3.16), which means that the irregular solution oscillates more and more rapidly as $r\to 0$ for the attractive case, while its magnitude diverges exponentially for the repulsive case.

IV. APPROXIMATION METHODS

Let us now return to the topic of approximation schemes for $f_0(k,r)$ based on (2.16)–(2.19). These methods will only be accurate at energies that are sufficiently low that the product $2kr\equiv|z|$ is less than unity. If this inequality is satisfied, the series on the right-hand side of (2.16) will converge and a reasonably accurate approximation will result if the sum is limited to terms of order z^m with m less than or equal to some small fixed positive integer μ. This requirement limits the m sum to the range $0\leqslant m\leqslant\mu$ and the L sum to $0\leqslant L\leqslant[(\mu-m)/M]$. We shall refer to this as the μth-order-series approximation. If $|\lambda|$ is large and r is of order r_0, ξ may be large. Thus a large number of terms could be required to obtain an accurate approximation to the functions $\mathscr{A}_{L,m}(\xi)$, [cf (2.17)]. As r increases, ξ decreases and fewer terms are required. On the other hand $|z|$ will increase, so that larger values of μ will be required to obtain an accurate solution. For very large values of r, $|z|$ will be greater than unity so the μth-order approximation with μ finite will not suffice. On the other hand, large r means small ξ so only a few terms in the series expansion for $\mathscr{A}_{L,m}(\xi)$ will be required. Since the running index in the series for $\mathscr{A}_{L,m}(\xi)$ is j, which corresponds to the index that labels the order of Born approximation, one sees that for large r a reasonable approximation to $f_0(k,r)$ can be obtained by summing the first ν terms of the Born series for $f_0(k,r)$ [cf. (2.1)]. We shall call this the νth-order Born approximation. In terms of Eq. (2.16) and (2.17) this is formally obtained by limiting the sum on j in (2.17) to $0\leqslant j\leqslant\nu-L$, and limiting the sum on L in (1.6) to $0\leqslant L\leqslant\nu$, while letting the m sum run from 0 to ∞, or until convergence is obtained.

As r is continually decreased from some large initial value, $|V(r)|$ will continually increase, so that higher and higher order ν will be required to obtain accurate approximations to $f_0(k,r)$. Alternatively, if the order ν is fixed, the accuracy of the approximation will decrease as r decreases.

From the above discussion one sees that a μth-order series approximation, with μ fixed, will provide an accurate result when r is small, but the accuracy will deteriorate as r is increased, while the accuracy of the νth-order Born calculatin, with ν fixed, will be accurate in the complementary region, being excellent when r is large and poor when r is small. An intermediate approximation scheme can be obtained by adding to the first ν terms of the Born series all of the terms of the μth-order series approximation that do not arise from the first ν Born terms. We shall call this the (ν,μ)th-order combined approximation.

To test the proposed approximation schemes, we have calculated $f_0(k,r)$ using several of them. For these calculations employed the following physical parameters:

$$m=939\ \text{MeV/c}^2,$$
$$r_0=1\ \text{fm},\quad k=5\times10^4\ \text{fm}^{-1}. \tag{4.1}$$

The value of k given corresponds to a kinetic energy of 5.17 eV. In addition the radius was assigned values from 1 to 10^4 fm, λ was assigned values of 10 and -10, and N was set equal to integral values of 3–6. As a crude measure of the error in the calculation of $f_0(k,r)$, we computed the quantity

$$\text{ERROR}=\left|\left(k^2+\frac{d^2}{dr^2}-V(r)\right)f_0^{\text{App}}(k,r)\right|, \tag{4.2}$$

where $f_0^{\text{App}}(k,r)$ is the approximating wavefunction whose accuracy is being measured. ERROR $=0$ for the exact solution. This test of the accuracy of the approxmation schemes is illustrated in Figs. 1–6. The approximation schemes tested where the μth-order series methods with $\mu=2$ and 3, the first-order Born approximation ($\nu=1$) and the $(\mu,1)$th-order combined approximation, again with $\mu=2$ and 3. Figures 1–4 show plots of the error, as defined in (4.2), vs r for each of

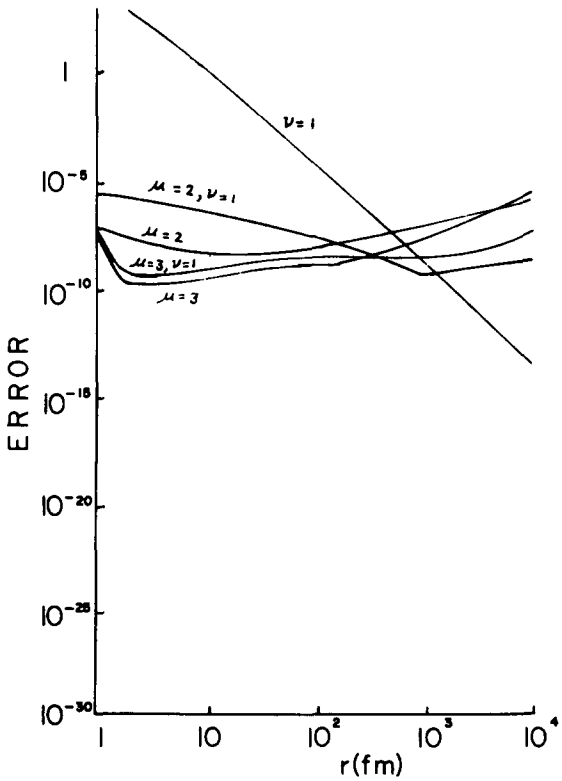


FIG. 1. A plot of ERROR vs r for five different approximation schemes for the case of an attractive van der Waals potential with parameters $N = 3$, $r_0 = 1$ fm, $\lambda = 10$, and a wavenumber $k = 5\times10^{-4}$ fm^{-1}. The definition of the function "ERROR" and a description of each of the approximation schemes is given in the text.

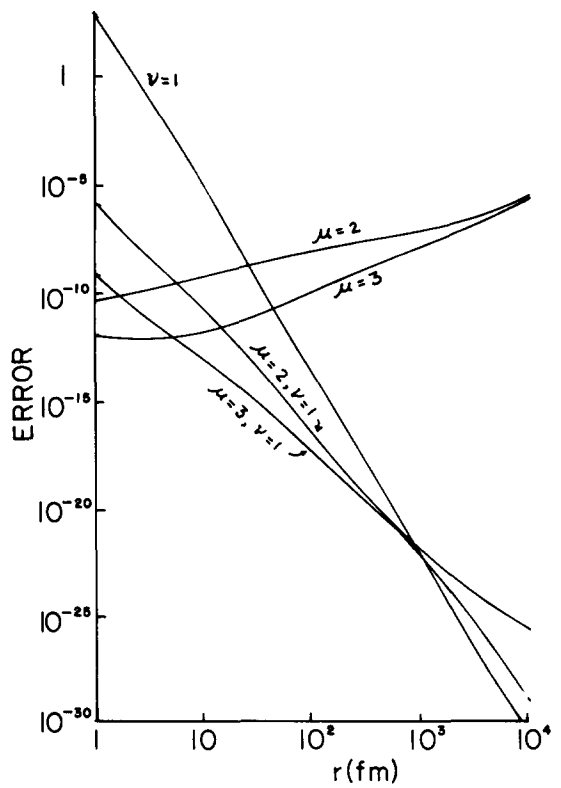


FIG. 3. ERROR vs r for the same approximation schemes, for the case $N = 5$, $r_0 = 1$ fm, $\lambda = 10$, and $k = 5\times10^{-4}$ fm^{-1}.

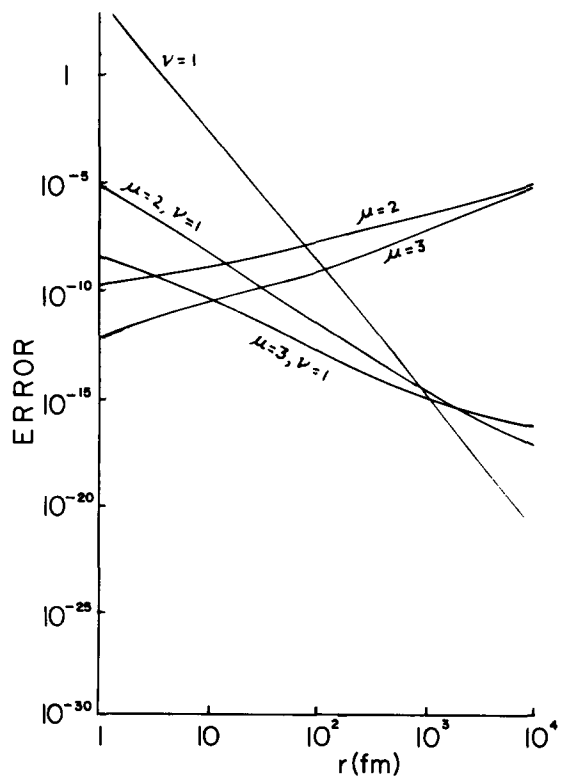


FIG. 2. ERROR vs r for the same set of approximation schemes for the case $N = 4$, $r_0 = 1$ fm, $\lambda = 10$, and $k = 5\times10^{-4}$ fm^{-1}.

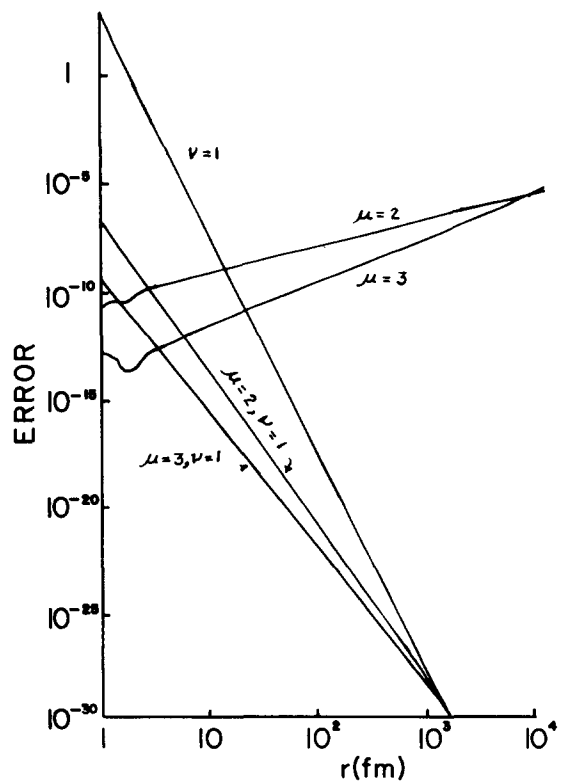


FIG. 4. ERROR vs r for the same approximation schemes, for the case $N = 6$, $r_0 = 1$ fm, $\lambda = 10$, and $k = 5\times10^{-4}$ fm^{-1}.

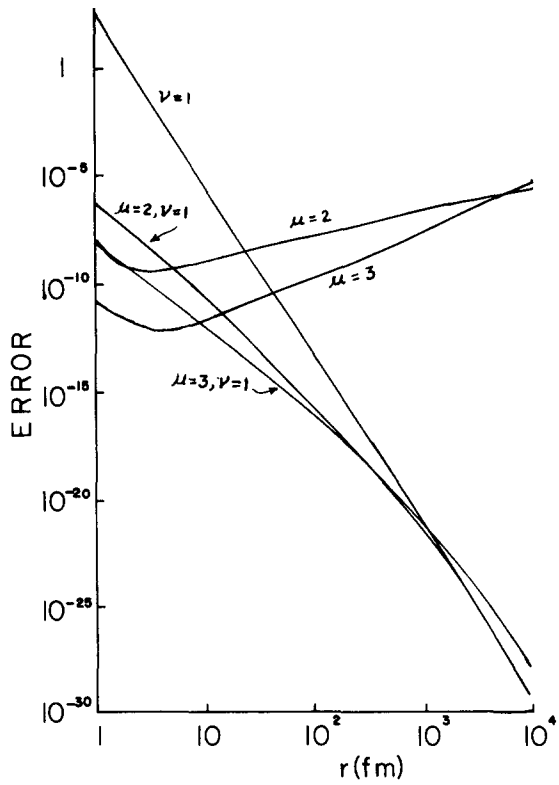


FIG. 5. ERROR vs r for the same approximation schemes, for a repulsive potential with $N=5$, $r_0=1$ fm, $\lambda=-10$, and k once again equal to 5×10^{-4} fm^{-1}.

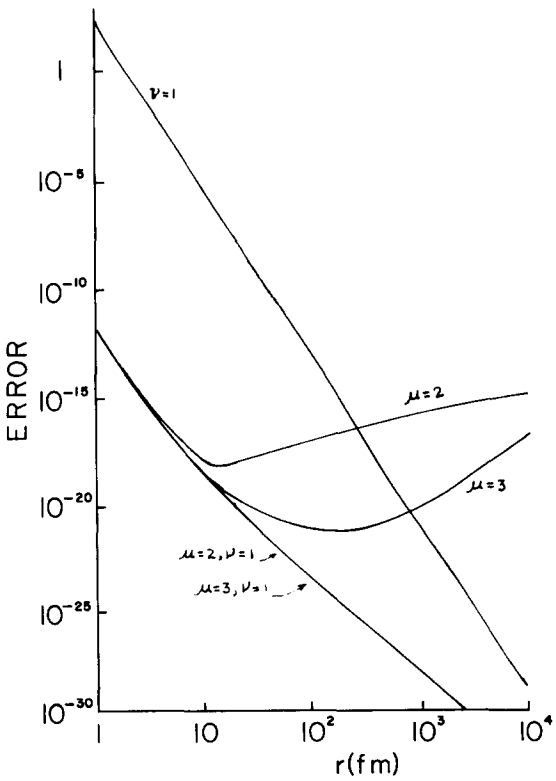


FIG. 6. ERROR vs r for the same set of schemes, for the potential considered in Fig. 3, but with $k=5\times10^{-7}$ fm^{-1}.

the approximations considered. The wavenumber and strength parameter were fixed at $k=5\times10^{-4}$ fm^{-1} and $\lambda=10$, corresponding to an attractive potential, in each of the calculations, while the potential parameter N varied from $N=3$ in Fig. 1 up to $N=6$ in Fig. 4; Figs. 5 and 6 are again plots of ERROR vs r but with the parameters $k=5\times10^{-4}$ fm, $\lambda=-10$, and $N=5$ in Fig. 5 and $k=5\times10^{-7}$ fm, $\lambda=10$, and $N=5$ in Fig. 6. These figures show, as was expected, that the μth-order series methods are most accurate for small r with the accuracy improving with increasing μ, and that the first-order Born series calculation is very accurate at large r. They also show that the combined approximation is quite accurate over the entire range of r considered, whereas the μth-order series approximation is not very accurate at large r and the first-order Born approximation breaks down at small r. One also sees that the accuracy of the various approximations increase with increasing N. This simply results from the fact that for a fixed value of λ the magnitude of the potential decreases as N increases for all radii greater than 1 fm. Comparison of Figs. 3 and 5 shows that the sign of the potential does not have much effect on the accuracy of any of the approximation schemes, while a comparison of Figs. 3 and 6 shows that the accuracy of the μth-order series approximation schemes increase as k decreases when r is held fixed, which was again expected since these approximations are based on a series expansion in $z=2kr$, which converges rapidly when $|z|$ is smaller than unity. These last two results were also found to hold for $N=3$, 4, and 6.

Using bounds on the magnitudes of the terms of the Born series [cf. (2.1) and (2.2)] which are given in Ref. 3, one finds that the first-order Born approximations is reasonably accurate if $|\xi|$ is small compared with unity. For $|\lambda|=10$ this implies that the first-order Born approximation should be accurate if r is larger than r_N, where $r_N\sim100$, 15, 10, and 3 fm for $N=3$–6, respectively. Since the series approximation is accurate for $r<r_s\equiv(2k)^{-1}$, one should get reasonable agreement between the values of $f_0(k,r)$ calculated by means of a μth-order series calculation for $\mu=2$ or 3 and a first-order Born calculation, when r lies between r_N and r_s. An examination of the calculated values of $f_0(k,r)$ revealed that this was indeed the case.

Figure 7 shows the dependence of $|V(r)|$ on r for $N=3$–6 and $|\lambda|=10$. The two horizontal lines correspond to the kinetic energy $E=\hbar^2k^2/2m$ for $k=5\times10^{-7}$ and 5×10^{-4} fm^{-1}. When $|V(r)|\ll E$, as it is, for example, for $\lambda=10$ and $N=3$ when $r\gg3$ fm if $E=5.17$ eV, or when $r\gg250$ fm if $E=5.17\times10^{-6}$ fm, then the potential is negligible, and the zeroth-order Born approximation should be reasonably accurate, i.e., $f_0(k,r)\sim e^{-ikr}$. Again, if one checks the calculated values with this simple approximation, it turns out to be very accurate over the range indicated.

The behavior of $f_0(k,r)$ at small values of r, or large value of $|\lambda|$, was discussed at the end of Sec. III. In Fig. 8, we have plotted $f_0(k,r)$ vs r for $1\leqslant r\leqslant10$ fm, for $\lambda=-10$, and $N=3$–6. The increase of $|f_0(k,r)|$ as r tends to zero is quite evident. For an attractive potential $f_0(k,r)$ oscillates very rapidly as r tends to zero, so in Figs. 9(a)–9(d) we have plotted $f_0(k,r)$ against the more natural variable ξ for the $\lambda=+10$ case

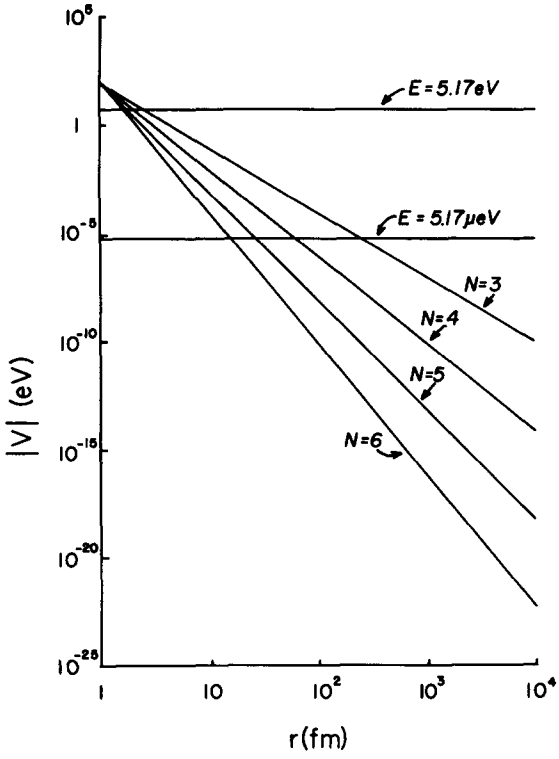


FIG. 7. $|V(r)|$ vs r for various potentials. The horizontal lines at 5.17 eV and 5.17×10^{-6} eV are the kinetic energies corresponding to $k = 5\times10^{-4}\,\mathrm{fm}^{-1}$ and $5\times10^{-7}\,\mathrm{fm}^{-1}$, respectively.

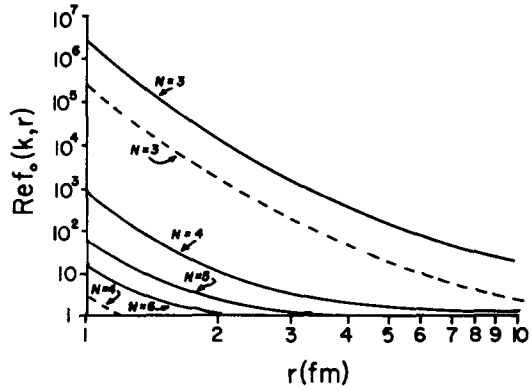


FIG. 8. A plot of $f_0(k,r)$ vs r at wavenumber $k = 5\times10^{-4}\,\mathrm{fm}^{-1}$ for a set of repulsive potentials with $3\leq N\leq 6$, $\lambda = -10$, and $r_0 = 1$ fm. The solid curves correspond to $\mathrm{Re}[f_0(k,r)]$ and the dashed curves to $\mathrm{Im}[f_0(k,r)]$. The imaginary part of $f_0(k,r)$ lies below the range of values plotted when $N > 4$.

[cf. (2.18) and (3.15)], again with $N = 3$–6. The range of ξ values included in these figures corresponds to a range of r of $10^2\leq r\leq 10^3$ fm for $N = 3$, $10\leq r\leq 30$ fm for $N = 4$, $5\leq r\leq 10$ fm for $N = 5$, and $3\leq r\leq 6$ fm for $N = 6$. The points on the graph correspond to the computed value of $f_0(k,r)$ for the more accurate approximation scheme at each r. These *best* computed values are only slightly different from the values predicted by Eq. (3.15) for the $\lambda = 10$ case or (3.16) for the

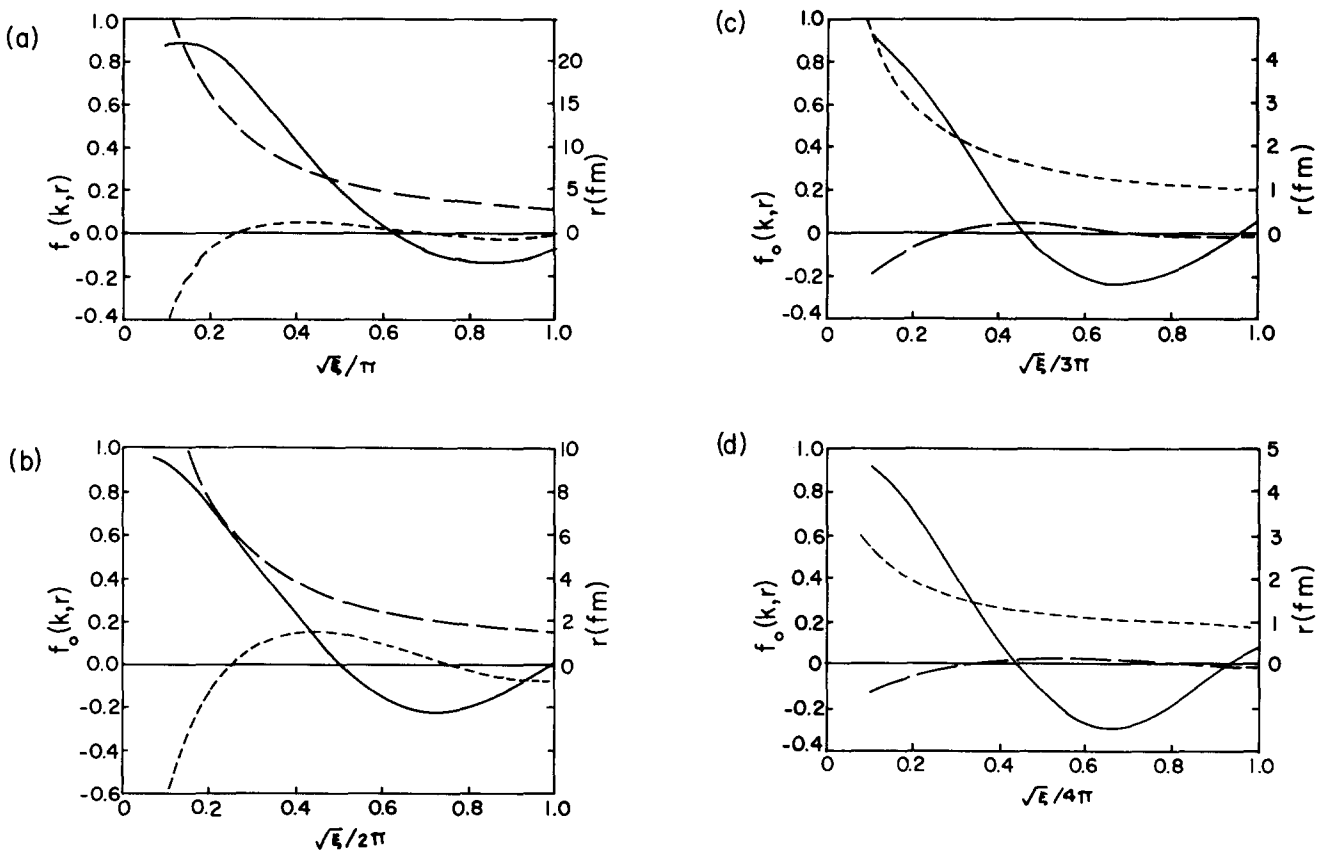


FIG. 9. $f_0(k,r)$ vs $\sqrt{\xi}/N\pi$ at wavenumber $k = 5\times10^{-4}\,\mathrm{fm}^{-1}$ for a set of attractive potentials with $\lambda = +10$, $r_0 = 1$ fm, and (a) $N = 3$, (b) $N = 4$, (c) $N = 5$, and (d) $N = 6$. The solid line corresponds to $\mathrm{Re}[f_0(k,r)]$, the short dashed line to $\mathrm{Im}[f_0(k,r)]$ times unity for (a), and times 100 for (b), (c), and (d). The long dashed curve is a plot of r vs $\sqrt{\xi}/N\pi$, with the abscissa for r at the right-hand edge of the graphs.

$\lambda = -10$ case. For values of r greater than the range of values shown in Fig. 8 or beyond the range of r implicit in Figs. 9(a)–9(d), the function $f_0(k,r)$ is quite accurately approximated by $\exp(-ikr)$.

V. THE s-WAVE SCATTERING AMPLITUDE

To determine the scattering amplitude, one needs the regular solution $\phi_0(k,r)$ to the s-wave Schrödinger equation, as well as the two irregular solutions $f_0(\pm k,r)$. [Note that the second irregular solution $f_0(-k,r)$ can be obtained by the same approximations used for $f_0(k,r)$, all one needs to do is change the sign of k.] If one requires that

$$\lim_{r\to 0}\{\phi_0(k,r)\} = 1, \tag{5.1}$$

then

$$\phi_0(k,r) = [f_0(-k)f_0(k,r) - f_0(k)f_0(-k,r)]/2ik, \tag{5.2}$$

where $f_0(\pm k)$ are the Jost functions. Applying rd/dr to (5.2) and solving the resulting pair of equations, (5.2) and the derivative equation for the ratio of the Jost functions yields

$$e^{2i\delta_0} \equiv \frac{f_0(k)}{f_0(-k)} = \frac{\mathscr{L}(k,r)f_0(k,r) - rdf_0(k,r)/dr}{\mathscr{L}(k,r)f_0(-k,r) - rdf_0(-k,r)/dr}, \tag{5.3}$$

where δ_0 is the s-wave phase shift and $\mathscr{L}(k,r)$ is r times the logarithmic derivative of $\phi_0(k,r)$, i.e.,

$$\mathscr{L}(k,r) = \frac{rd(\ln\phi_0(k,r))}{dr}. \tag{5.4}$$

Although the right-hand side of (5.3) appears to depend on r, it in fact does *not*, and therefore it may be evaluated at any convenient radius. If one assumes that $V(r)$ is real and that it behaves better than r^{-2} near the origin, then $\phi_0(k,r)$ and hence $\mathscr{L}(k,r)$ will be real and entire functions of k^2 for every fixed r[3]. Hence for small k one can write

$$\mathscr{L}(k,r) = \sum_{n=0}^{\infty} k^{2n}\mathscr{L}_n(r). \tag{5.5}$$

Using (5.3), one finds that the s-wave scattering amplitude is given by

$$A_0(k) = e^{i\delta_0}\sin\delta_0/k = \mathscr{N}(k,r)/[\mathscr{D}(k,r) - ik\mathscr{N}(k,r)], \tag{5.6a}$$

where

$$\mathscr{N}(k,r) = \frac{\mathscr{L}(k,r)[f_0(k,r) - f(-k,r)] - rd[f_0(k,r) - f_0(-k,r)]/dr}{2ik} \tag{5.6b}$$

and

$$\mathscr{D}(k,r) = \frac{\mathscr{L}(k,r)[f_0(k,r) + f_0(-k,r)] - rd[f_0(k,r) + f_0(-k,r)]/dr}{2}. \tag{5.6c}$$

Since $f_0(-k_0,r)$ is the complex conjugate of $f_0(k,r)$ as long as $V(r)$ is real,[3] which we shall assume to be the case, the $\mathscr{N}$ and $\mathscr{D}$ will be real functions of k and r. Equations (5.6a)–(5.6c) hold for all values of r. If we take $r > R$ [cf. Eq.(1.6)], and choose k so that $k \ll r^{-1}$, then we may use (2.16) to evaluate the functions $f_0(\pm k,r)$ and $df_0(\pm k,r)/dr$.

For potentials that vanish faster than any inverse power of r as $r\to\infty$, such as a cutoff potential, or one tends exponentially to zero, only the $L = 0$ term appears in the low-energy expansion of $f_0(k,r)$ given by (2.16). It then follows that $\mathscr{N}$ and $\mathscr{D}$ are even functions of k, as will be the cross section which is given by

$$\sigma = 4\pi(\mathscr{N}(k,r))^2/[(\mathscr{D}(k,r))^2 + k^2(\mathscr{N}(k,r))^2] \tag{5.7}$$

for small values of k where the s wave dominates. The $L = 1$ terms of (2.16) give rise to two distinct sets of terms for both $\mathscr{N}$ and $\mathscr{D}$. The first of these sets, like those arising from the $L = 0$ term, will be proportional to even powers of k, while the second set will either be proportional to odd powers of k or to even powers of k times $\ln k$. It is this second set of terms that distinguishes the low-energy cross section of a potential with a van der Waals tail from the cross section of a similar potential that vanishes faster than any inverse power of r. The contribution of the term of (2.16) with $L > 1$ can be neglected compared to the leading terms arising from $L \leqslant 1$ as k tends to zero.

To examine these statements more closely, we split $\mathscr{N}$ and $\mathscr{D}$ in the following fashion:

$$\mathscr{N} = \mathscr{N}_\alpha + \mathscr{N}_\beta, \quad \mathscr{D} = \mathscr{D}_\alpha + \mathscr{D}_\beta, \tag{5.8}$$

where $\mathscr{N}_\alpha$ and $\mathscr{D}_\alpha$ consist of all the terms in the low energy expansion of $\mathscr{N}$ and $\mathscr{D}$, respectively, that are even functions of k, and $\mathscr{N}_\beta$ and $\mathscr{D}_\beta$ consist of all the remaining terms. From (2.16), its derivative with respect to r and Eq. (5.5), one obtains

$$\begin{aligned}\mathscr{N}_\alpha(k,r) = 2r\sum_{m=0}^{[(M-1)/2]} k^{2m}\Bigg\{\sum_{n=0}^{m}\mathscr{L}_{m-n}(r)(-4r^2)^n \\ \times \mathscr{A}_{0,2n+1}(\xi) \\ -(-4r^2)^m\Big((2m+1)\mathscr{A}_{0,2m+1}(\xi) - M\mathscr{B}_{0,2m+1}(\xi)\Big)\Bigg\} \\ + k^{M-1}\mathscr{F}(r)(-1)^{(M-1)/2}\ln(2r/r_0)\mathscr{P}_{M,0} + O(k^{M-1}),\end{aligned} \tag{5.9a}$$

$$\begin{aligned}\mathscr{N}_\beta(k,r) = k^{M-1}\mathscr{F}(r)\{\tfrac{1}{2}(-1)^{M/2}\pi\mathscr{P}_{M,E} \\ + (-1)^{(M-1)/2}\ln(kr_0)\mathscr{P}_{M,O}\} \\ + o(k^{M-1}),\end{aligned} \tag{5.9b}$$

$$\begin{aligned}\mathscr{D}_\alpha(k,r) = \sum_{m=0}^{[M/2]} k^{2m}\Bigg\{\sum_{n=0}^{m}\mathscr{L}_{m-n}(r)(-4r^2)^n\mathscr{A}_{0,2n}(\xi) \\ -(-4r^2)^m\Big(2m\mathscr{A}_{0,2m}(\xi) - M\mathscr{B}_{0,2m}(\xi)\Big)\Bigg\} \\ + k^M\mathscr{F}(r)(-1)^{M/2}\ln(2r/r_0)\mathscr{P}_{M,E} + o(k^M),\end{aligned} \tag{5.9c}$$

$$\mathscr{D}_\beta(k,r) = k^M \mathscr{F}(r)\{(-1)^{M/2}\ln(kr_0)\mathscr{P}_{M,E} + (-1)^{(M+1)/2}\pi\mathscr{P}_{M,O}/2\} + o(k^M), \tag{5.9d}$$

with

$$\mathscr{F}(r) = mc\lambda(2r_0)^{M+1}\hbar^{-1}\{\mathscr{L}_0(r)\mathscr{A}_{1,0}(\xi) + M\mathscr{B}_{1,0}(\xi)\}, \tag{5.10}$$

while $\mathscr{P}_{M,E}$ and $\mathscr{P}_{M,O}$ are projection operators onto even and odd values of M,

$$\mathscr{P}_{M,O} = \{1+(-1)^M\}/2, \quad \mathscr{P}_{M,E} = 1 - \mathscr{P}_{M,O} \tag{5.11}$$

and

$$\mathscr{B}_{L,m}(\xi) = \sum_{j=1}^{\infty} j\xi^j b_{LM+m}^{(L+j)}(L). \tag{5.12}$$

With these expressions the scattering amplitude becomes

$$A_0(k) = A_0^{SR}(k) + A_0^{LR}(k), \tag{5.13}$$

where

$$A_0^{SR}(k) \simeq \mathscr{N}_\alpha(k,r)/\{\mathscr{D}_\alpha(k,r) - ik\mathscr{N}_\alpha(k,r)\}. \tag{5.14}$$

Since $\mathscr{N}_\alpha$ and $\mathscr{D}_\alpha$ are real functions of k^2, the differential cross section $|A_0^{SR}(k)|^2$ obtained by ignoring $A_0^{LR}(k)$ will be an even function of k. As this is a property of the low-energy differential cross section of a short range potential (one vanishing faster than any inverse power of r as $r\to\infty$), we have labeled this component of the scattering amplitude SR for short range, even though A_0^{SR} does contain some contribution from the van der Waals potential. The second term $A_0^{LR}(k)$ is given by

$$A_0^{LR}(k) \simeq \mathscr{N}_\beta(k,r)/\mathscr{D}_\alpha(0,r). \tag{5.15}$$

This amplitude has a k dependence that provides important information about the long-range van der Waals tail. For the power law $V(r) \propto r^{-N}$ it will be shown that $A_0^{LR}(k)$ behaves like $k^{N-3}\ln k$ for odd values of N, and like k^{N-3} (which is an odd power of k) for even values of N. In all cases the differential cross section will contain a term that can *not* be parameterized by a power series in k^2. The k dependence of this term changes with N and hence it provides a means of testing for the presence of a van der Waals tail with a specific value of N. From (5.9b), (5.9c), and (5.15) one has

$$A_0^{LR}(k) = \frac{4mc\lambda r_0^2(2kr_0)^{N-3}}{\hbar}\left\{\frac{\mathscr{L}_0(k)\mathscr{A}_{1,0}(\xi) + M\mathscr{B}_{1,0}(\xi)}{\mathscr{L}_0(k)\mathscr{A}_{0,0}(\xi) + M\mathscr{B}_{0,0}(\xi)}\right\} \times((-1)^{M/2}\pi\mathscr{P}_{M,E}/2 + (-1)^{(M-1)/2}\ln(kr_0)\mathscr{P}_{M,O}). \tag{5.16}$$

Next, from (3.11) it follows that

$$b_M^{(j+1)}(1) = (-1)^M b_0^{(j)}(0)/(M+1)! = (-1)^N b_0^{(j)}(0)/(N-1)!.$$

Hence, by (2.17) and (5.12) one finds that the term in braces on the right-hand side of (5.16) equals $(-1)^N/(N-1)!$, which then gives us

$$A_0^{LR}(k) = -\frac{4mc\lambda r_0^2(2kr_0)^{N-3}}{(N-1)!\hbar}(\tfrac{1}{2}(-1)^{N/2}\pi\mathscr{P}_{M,E} - (-1)^{(N-1)/2}\ln(kr_0)\mathscr{P}_{M,O}). \tag{5.17}$$

This in turn leads to the following low-energy differential cross sections:

$$\frac{d\sigma}{d\Omega} \approx |A_0^{SR}(0) - 2mc\lambda r_0^2\ln(kr_0)/\hbar|^2 \quad \text{for } N=3, \tag{5.18a}$$

$$\frac{d\sigma}{d\Omega} \approx |A_0^{SR}(k)|^2 + K(2kr_0)^{N-3}(-1)^{N/2}\pi/(N-1)!, \tag{5.18b}$$

for N even, $N\geqslant 4$, and

$$\frac{d\sigma}{d\Omega} \approx |A_0^{SR}(k)|^2 + 2K(2kr_0)^{N-3}(-1)^{(N+1)/2}\ln(kr_0)/(N-1)! \tag{5.18c}$$

for N odd, $N\geqslant 5$, where

$$K = -4mc\lambda\hbar^{-1}r_0^2 A_0^{SR}(0). \tag{5.18d}$$

The characteristic van der Waals amplitude, which we have called $A_0^{LR}(k)$, will only arise if the van der Waals interaction extends out to infinite radius, for, if we truncate the potential in any radius, *no matter how large*, then $\mathscr{N}$ and $\mathscr{D}$ can be expanded in a power series in k^2, and the amplitude $A_0^{LR}(k)$ does not occur. Thus, we can associated $A_0^{LR}(k)$ with the behavior of the potential as r tends to infinity. Since $V(r)$ vanishes in this limit, one is not surprised that $A_0^{LR}(k)$ is first order in λ.

To add to our insight into this matter, and to determine the angular dependence of the differential cross section, we consider the following approach. Let us define two potentials V_1 and V_2 by

$$V_1(r) = V(r)U(R-r), \quad V_2(r) = V(r)U(r-R), \tag{5.19}$$

where $U(X)$ is the unit step function, $U(X) = 1$, for $X\geqslant 0$, and $U(X) = 0$ for $X<0$. Clearly, $V(r) = V_1(r) + V_2(r)$. We assume that R is very large so that $V_2(r)$ is a very weak potential. Then, if one calculates the T matrix operator for the interaction V, retaining only first-order terms in the interaction V_1, he gets[8]

$$T = T_1 + V_2 + V_2G_0T_1 + T_1G_0V_2 + T_1G_0V_2G_0T_1, \tag{5.20}$$

where

$$T_1 = V_1 + V_1G_0T_1$$

and G_0 is the free-particle Green's function. Since the interaction V_1 is only strong at small values of r and the Green's function falls off rapidly with the distance $|\mathbf{r}-\mathbf{r}'|$, the three terms of (5.20) that arise from coupling between the interaction V_1 at a point $\mathbf{r}$ with the interaction V_2 at a point $\mathbf{r}'$ will be negligibly small, so one need only retain the two leading terms in (5.20). If one takes the matrix elements of the T matrix operator between initial and final plane wave states and multiplies the resulting expression by $(-m/2\pi\hbar^2)$, he obtains the scattering amplitude. Let us divide the scattering amplitude A into two parts; the first one, A^{SR}, will consist of the matrix element of T_1 and those terms arising from the matrix element of V_1 that are proportional to even powers of q, where q is the momentum transfer; the second part which we designate A^{LR} will consist of all those terms in the low-energy expansion of the matrix element V_1 that were not

included in A^{SR}. This procedure yields

$$A^{LR} = 2mc\lambda\hbar^{-1}r_0^2 g_N(qr_0) \tag{5.21}$$

for the characteristic scattering amplitude arising from the van der Waals tail of the potential, where

$$g_N(x) = (-1)^{(N-1)/2}x^{N-3}\ln(x)/(N-2)! \quad (N \text{ odd}),$$
$$g_N(x) = (-1)^{(N-2)/2}\pi x^{N-3}/2 \quad (N \text{ even}).$$

The differential cross section then becomes

$$\frac{d\sigma}{d\Omega} \approx |a + 2mc\lambda r_0^2\hbar^{-1}\ln(qr_0)|^2 \quad \text{for } N = 3, \tag{5.22a}$$

where a is the scattering length, and

$$\frac{d\sigma}{d\Omega} \approx |A^{SR}(E,\theta)|^2 + 4mc\lambda\hbar^{-1}r_0^2(\mathrm{Re}A^{SR}(0,\theta))g_N(qr_0) \tag{5.22b}$$

for $N > 3$. Since $A^{SR}(E,\theta)$ is the sum of the scattering amplitude of a finite range potential and other terms proportional to an even power of q, $|A^{SR}|^2$, and $\mathrm{Re}A^{SR}$ can reasonably be parameterized as follows:

$$|A^{SR}(E,\theta)|^2 \simeq a_0 + a_1 E + a_2 E\cos\theta, \tag{5.23}$$
$$\mathrm{Re}A^{SR}(0,\theta) \simeq a_3,$$

where the a_n are constants and $a_3^2 \leqslant a_0$.

If one projects out the s-wave component of $A^{LR}(E,\theta)$, he gets the amplitude $A_0^{LR}(k)$ of (5.17), as required, although our derivation of Eqs. (5.21)–(5.23) is far from rigorous the arguments employed seem to be reasonable and they lead to the correct s-wave dependence. Furthermore, the s-wave amplitude, which has been carefully derived, dominates at low energy and the arguments used to justify the neglect of the last three terms of Eq. (4.18), which was essential in our nonrigorous derivation, should be much stronger for angular momentum components with $l > 0$, since the centrifugal barrier greatly decreases the probability of multiple scattering between the two regions of space $r \ll R$ and $r > R$.

VI. COMMENTS AND CONCLUSIONS

If k is taken to be a complex number in Eqs. (1.1) and (1.3), the steps leading to the construction of the series expansion for $f_0(k,r)$ given by Eqs. (2.16)–(2.19) remain valid. From (2.16) it can be seen that $f_0(k,r)$ has logarithmic branch points at $k = 0$ and $k = \infty$. If one joins these two branch points by a cut running along the positive imaginary axis of the complex k plane or, equivalently, the negative real z axis of the complex z plane, then $f_0(k,r)$ will satisfy the standard symmetry relation[3]

$$f_0(-k^*,r) = f_0^*(k,r) \tag{6.1}$$

on the first Riemann sheet given by $-3\pi/2 < \arg k \leqslant \pi/2$. If k is a pure-real or pure-imaginary number, the energy $E = \hbar^2k^2/2m$ will be real, a condition that physical solutions to (1.1) must satisfy. A solution $f_0(k,r)$ for which k is a positive or a negative real number corresponds to an incoming or an outgoing spherical wave, respectively. Equation (5.2) states that a physical solution with positive energy is a linear combination of an incoming and an outgoing spherical wave. The negative energy solutions, if they exist, correspond to s-wave bound states. These solutions must be normalizable; hence k must be a negative imaginary number. Equation (6.1) guarantees that $f_0(k,r)$ will be real when k lies along the negative imaginary axis. In addition to the continuum of positive energy scattering solutions and the bound states, (1.1) may also have resonance solutions. The resonance solutions[9] are regular at the origin and are proportional to $f_0(k,r)$ when $r > R$, as was the case for the bound state solutions, but in the case of the resonance solutions k will be complex with a positive imaginary part. Note that the resonance solutions are not normalizable since they increase exponentially in magnitude as $r \to \infty$.

Although we have referred to the μth-order series approximation scheme based on (2.16) as a low-energy approximation scheme, a more precise term would be a threshold approximation scheme, since the condition that $|k|$ is small implies that the energy $\hbar^2k^2/2m$ lies close to the threshold at zero energy. From the discussion at the beginning of this section one can see that a finite series approximation scheme can be used to construct any wavefunction $f_0(k,r)$ no matter if k is real, imaginary, or even complex as long as $|kr|$ is small. Hence these approximations can be employed to construct the external wavefunctions over a range of r values $R < r \ll (2|k|)^{-1}$, for bound states, resonances, and scattering states as long as $|k|$ is small.

In Sec. V we saw that a μth-order series approximation with $\mu = 2$ or 3 provided excellent approximations for $f_0(k,r)$ at low energies and relatively small radii, $|kr| \lesssim 0.1$. This claim is supported by the fact that the zeroth-order series expansion provides an exact solution to $f_0(k,r)$ when $k = 0$ and the additional facts that, in the calculations reported in Sec. IV with $k \leqslant 5\times10^{-4}$ fm, the error as given by (4.2) was small for both $\mu = 2$ and 3 when $|kr| \lesssim 0.1$, the error decreased with μ, and finally the change in the calculated values of $f_0(k,r)$ in going from $\mu = 2$ to 3 was never greater than 0.1% for $|kr| \lesssim 0.1$, indicating that the series was converging rapidly with increasing μ.

The advantage of having an approximation formula for $f_0(k,r)$ for relatively small values of r is twofold: First, in matching the internal to the external wavefunction to determine the Jost functions for scattering states, or the energies of bound and resonant states, one normally has to numerically integrate the radial Schrödinger equation from the origin out to the matching radius, knowing the value of $f_0(\pm k,r)$ at relatively small values of r reduces the number of steps in the numerical integration; second, the approximation provides one with an explicit expression for the leading λ and N dependence of $f_0(k,r)$. This enables one to analytically determine the contribution of the van der Waals tail to such things as low-energy scattering, as was done in Sec. V, or to the binding energy of levels which lie just below threshold. Such a general analysis is not possible, or at least very difficult, if one employs simple numerical solutions to (1.1).

Since (2.16) can not be employed until $V(r)$ reaches its asymptotic form, $r \geqslant R$, the condition $|k|r \ll 1$ may not apply in the asymptotic region, even when $|k|$ is quite small. In such cases the (μ,ν)th-order combined approximation or the νth-order Born approximation might be usefully employed.

ACKNOWLEDGMENTS

I would like to thank Dr. M. K. Sundaresan and Dr. J. E. Hardy for helpful discussions.

APPENDIX A

In this appendix we shall show that the integral on the right-hand side of Eq. (2.7) may be taken over any simple path between the points z and $+\infty$ that does not encircle the origin. To establish this fact, we need two theorems.

Theorem 1: $|F^{(j)}(z)|$ and $|dF^{(j)}(z)/dz|$ are finite for every z in the domain D_y, where D_y consists of all z such that $\mathrm{Re}z \geqslant 0$ and $|z| \geqslant y > 0$ for some fixed positive real number y.

Proof: If the integral is taken along the path C_z, one has

$$|F^{(j)}(z)| \leqslant 2\int_{|z|}^{\infty} |x|^{-N} |F^{(j-1)}(x)|\, d\,|x|.$$

Evaluating this integral for $j = 1$, using (2.6) on the right-hand side, and then iterating $j-1$ times yields

$$\begin{aligned} |F^{(j)}(z)| &\leqslant 2^j \left(\int_{|z|}^{\infty} x^{-N}\, dx\right)^j / j! \\ &= \{2|z|^{1-N}/(N-1)\}^j / j! \\ &\leqslant \{2y^{1-N}/(N-1)\}^j / j!. \end{aligned} \tag{A1}$$

This proves that $|F^{(j)}(z)|$ is uniformly bounded in D_y. It then follows that $dF^{(j)}(z)/dz$ is also uniformly bounded in D_y, since

$$\begin{aligned} |dF^{(j)}(z)/dz| &\leqslant \left|\int_z^{\infty} \exp(z-x) x^{-N} F^{(j-1)}(x)\, dx\right| \\ &\leqslant 2^j \int_{|z|}^{\infty} |y|^{-N} \left(\int_{|y|}^{\infty} x^{-N}\, dx\right)^{j-1} / (j-1)!\, dy \\ &\leqslant \{2y^{1-N}/(N-1)\}^j / j!\,. \end{aligned} \tag{A2}$$

Theorem 2: $F^{(j)}(z)$ is analytic throughout D_y.

Proof: Clearly $F^{(0)}(z) \equiv 1$ is analytic throughout D_y. If one assumes that $F^{(j)}(z)$ is analytic throughout D_y for some $j > 1$, it follows that the integrand on the right-hand side of (2.7) and its partial derivatrive with respect to z are analytic functions of both z and ζ' for every z and ζ' in D_y. In addition, the limits of integration are analytic functions of z. These facts taken together with (A1) and (A2) imply[10] that $F^{(j+1)}(z)$ is an analytic function of z through D_y. This completes the induction proof.

As a corollary to these two theorems, one finds that

$$|F(\Lambda,z)| \leqslant \sum_{j=0}^{\infty} |\Lambda^j F^{(j)}(z)| \leqslant \exp\{2|\Lambda|y^{1-N}/(N-1)\}; \tag{A3}$$

therefore the infinite sum converges uniformly in D_y and since each term is analytic throughout D_y, $F(z)$ is also an analytic function of z throughout D_y for each fixed value of Λ.

Let us replace the contour C_z in (2.7) by a new contour $C(R)$ described as follows: $C(R)$ consists of two segments $C_1(R)$, which is straight line segment running from $\zeta' = z$ to $\zeta' = z_R \equiv zR/|z|$, where R is a positive real number $\geqslant |z|$, and a second segment $C_2(R)$ which is a circular arc of radius R running from $\zeta' = z_R$ to $\zeta' = R$. Next we let $R \to \infty$ and observe that $C_1(\infty) = C_z$, while from (A1) it follows that the integral vanishes along the line segment $C_2(R)$ as $R \to \infty$. Hence, the contour C_z can be replaced by $C(\infty)$. Finally, since the integrand is analytic throughout D_y, one can deform the contour so that it follows any simple path from z to $+\infty$ in D_y.

APPENDIX B

In this appendix we shall evaluate a few of the constants of integration $C_n^{(j)}$ [cf. (2.13a) and (2.13b)]. We begin by sketching the general procedure for evaluating these coefficients. To obtain the coefficients $C_n^{(j)}$, one must evaluate the integral

$$I_n^{(j)}(z) = \int_z^{\infty} g_n^{(j-1)}(x)\, dx, \tag{B1}$$

where

$$g^{(j-1)}(z) = (e^{-z}\delta_{n,1} + \delta_{n,2}) z^{-N} F^{(j-1)}(z). \tag{B2}$$

If one can find an indefinite integral of $g_n^{(j-1)}(z)$, call it $G_n^{(j)}(z)$, whose asymptotic behavior as $|z| \to \infty$ is known, then

$$I_n^{(j)}(z) = \alpha_n^{(j)} - G_n^{(j)}(z), \tag{B3}$$

where $\alpha_n^{(j)}$ is a constant of integration, and, furthermore, the function $\mathcal{F}^{(j)}(z)$, defined by (2.14a) or (2.15b), can only differ from the series expansion of $G_n^{(j)}(z)$ by a constant, i.e.,

$$G_n^{(j)}(z) = \mathcal{F}_n^{(j)}(z) + \gamma_n^{(j)}, \tag{B4}$$

where $\gamma_n^{(j)}$ is the only z-independent term in the series expansions of $G_n^{(j)}(z)$. Therefore, by (2.15) and (B4) one has

$$C_n^{(j)} = \lim_{|z| \to \infty} G_n^{(j)}(z) - \gamma_n^{(j)} = \alpha_n^{(j)} - \gamma_n^{(j)}. \tag{B5}$$

This procedure is easily carried out when $j = 1$, but for large values of j it is hard to find functions $G_n^{(j)}(z)$ with the desired properties. To overcome this difficulty, we assume that the potential is proportional to z^{β} rather than z^{-N}, where β differs from minus N by an infinitesimal amount. By making this change the indefinite integrals become sums of terms which are powers of z times generalized hypergeometric functions. Both series expansions and asymptotic forms of these functions are known[11] so that one can construct the functions $G_n^{(j)}(z)$ by letting $\beta \to -N$ after the indefinite integrals are obtained. To illustrate the procedure in its simplest form, we first employ it to evaluate $C_1^{(1)}$ and $C_1^{(2)}$:

$$\begin{aligned} I_1^{(1)}(z) &= \lim_{\beta \to -N} \int_z^{\infty} e^{-u} u^{\beta}\, du \\ &= \alpha_1^{(1)} - \lim_{\beta \to -N} \left\{ \frac{z^{1+\beta}}{1+\beta}\, {}_1F_1(1+\beta, 2+\beta; -z) \right. \\ &\quad \left. + \frac{(-1)^N}{(N-1)!} \left(\frac{1}{\beta+N} + \psi^{(N)}\right)\right\}. \end{aligned} \tag{B6}$$

The constant of integration inside the braces has been included to cancel the pole at $\beta = -N$ in ${}_1F_1$. From the asymptotic form[11] of ${}_1F_1$, one then has

$$\lim_{|z| \to \infty} I_1^{(1)}(z) = \alpha_1^{(1)} + \lim_{|z| \to \infty} \frac{e^{-z}}{z^N} + \frac{(-1)^N \psi(N)}{(N-1)!} = 0$$

so that

$$\alpha_1^{(1)} = (-1)^{N-1} \psi(N)/(N-1)!, \tag{B7}$$

and from the series expansion of ${}_1F_1$ one has

$$G_1^{(1)}(z) = \lim_{\beta \to -N} \left(\frac{z^{1+\beta}}{1+\beta} {}_1F_1(1+\beta, 2+\beta, -z) + \frac{(-1)^N}{(N-1)!} \frac{1}{\beta+N} \right)$$
$$= \frac{(-1)^{N-1}}{(N-1)!} \ln z + \sum_{\substack{n=0 \\ n \neq N-1}}^{\infty} \frac{(-1)^n z^{n+1-N}}{n!(n+1-N)}; \tag{B8}$$

therefore, $\gamma_1^{(1)} = 0$ and

$$C_1^{(1)} = \alpha_1^{(1)} = (-1)^{N-1} \psi(N)/(N-1)!. \tag{B9}$$

To evaluate $C_2^{(1)}$, we examine $I_2^{(1)}(z)$

$$I_2^{(1)}(z) = \lim_{\beta \to -N} \int_z^{\infty} u^{\beta} \, du$$
$$= \alpha_2^{(1)} - \lim_{\beta \to -N} \frac{z^{\beta+1}}{\beta+1} = \alpha_2^{(1)} + \frac{z^{1-N}}{N-1}.$$

Since $I_2^{(1)}(z) \to 0$ as $|z| \to \infty$ one sees that $\alpha_2^{(1)} = 0$. Hence the "series" expansion of $I_2^{(1)}(z)$ has only one term and no z-independent terms. Thus,

$$C_2^{(1)} = 0. \tag{B10}$$

We must next consider the $j = 2$ case

$$I_1^{(2)}(z) = \int_z^{\infty} e^{-u} u^{-N} F^{(1)}(u) \, du = \alpha_1^{(2)} - G_1^{(2)}(z)$$
$$\equiv \alpha_1^{(2)} - \lim_{\beta \to 0} \left\{ \frac{(-1)^{N-1}}{(N-1)!} \left(\frac{1}{\beta+N} + \psi(N) \right) \frac{z^{\beta+1}}{\beta+1} + \frac{1}{(N-1)(2N-2)!(\beta+N)} \right]$$
$$+ \frac{z^{2\beta+2}}{2(\beta+1)^2} ({}_2F_2(\beta+1, 2\beta+2; \beta+2, 2\beta+3; -z)$$
$$- {}_1F_1(\beta+1, \beta+2; -z)) \Big\}, \tag{B11}$$

where ${}_2F_2$ is a generalized hypergeometric function[11] and a constant of integration proportional to $(\beta+N)^{-1}$ has been included to cancel a pole of ${}_2F_2$. From the asymptotic forms of ${}_1F_1$ and ${}_2F_2$ given in Ref. 11 and properties of gamma functions,[4] one finds

$$\lim_{|z| \to \infty} I_1^{(2)}(z) = \alpha_1^{(2)} + \frac{2\psi(2N-1)}{(N-1)^2(2N-2)!} = 0. \tag{B12}$$

Employing the series expansion of ${}_1F_1$ and ${}_2F_2$ and letting $\beta \to -N$, one finds that

$$G_1^{(2)}(z) = -\frac{z^{2-2N}}{N-1} \left\{ \frac{(-z)^{N-1}}{(N-1)!} (-\ln z + \psi(N)) - \sum_{\substack{n=0 \\ n \neq N-1}}^{\infty}{}' \frac{(-z)^n}{n!(n+1-N)} - 2 \left(\frac{-(-z)^{2N-2}}{(2N-2)!} \ln z - \sum_{\substack{n=0 \\ n \neq 2N-2}}^{\infty}{}' \frac{(-z)^n}{n!(n+2-2N)} \right) \right\}. \tag{B13}$$

The $n = 2N-2$ term of the first sum is independent of z when multiplied by z^{2-2N}; therefore,

$$\gamma_1^{(2)} = (N-1)^{-2}/(2N-2)!. \tag{B14}$$

Applying (B12) then gives

$$C_1^{(2)} = \alpha_1^{(2)} - \gamma_1^{(2)} = -2(N-1)\psi(2N-1) \div (N-1)^2(2N-2)!. \tag{B15}$$

From (B11)–(B14) and the series expansion of the exponential integral function,[4] one also finds

$$I_1^{(2)}(z) = z^{2-2N} [E_N(z) - 2E_{2N-1}(z)]/(N-1). \tag{B16}$$

By following the same procedure, a rather lengthy calculation yields

$$C_2^{(2)} = \frac{(-1)^{N-1}}{[(N-1)!]^2} \left\{ \frac{[\psi(N)]^2 - \psi^{(1)}(N)}{2} + \frac{\pi^2}{3} \right\}, \tag{B17}$$

where $\psi^{(1)}(N)$ is a polygamma function[4]:

$$\psi^{(1)}(N) = \pi^2/6 - \sum_{j=1}^{N-1} j^{-2}. \tag{B18}$$

[1] G. Feinberg and J. Sucher, Phys. Rev. D **20**, 1717 (1979).
[2] P. J. S. Watson and W. J. Romo, Carleton University preprint.
[3] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966); J. Math. Phys. **1**, 319 (1960).
[4] *Handbook of Mathematical Functions*, edited by M. Abramowitz and A. Stegun (National Bureau of Standards, Washington, D.C., 1965).
[5] The notation $[X]$ stands for the greatest integer less than or equal to X; for example, in (2.12a), $[m/M] \equiv$ (greatest integer $\leq m/M$).
[6] The $M = 1$ case is more complicated and a closed form solution for the coefficients shall not be presented.
[7] We use the following notation: $f(x) = o(x)$ as $x \to \infty$ (or 0) means that $f(x)/x$ vanishes as $x \to \infty$ (or 0); $f(x) = O(x)$ as $x \to \infty$ (or 0) means that $f(x)/x$ is bounded as $x \to \infty$ (or 0).
[8] A. Messiah, *Quantum Mechanics* (Wiley, New York, 1958), Vol. II, pp. 848–51.
[9] W. J. Romo, Nucl. Phys. A **116**, 618 (1968).
[10] E. T. Whittaker and G. N. Watson, *A Course of Modern Analyses* (Cambridge U.P., Cambridge, England, 1969), 4th ed., p. 92.
[11] Y. L. Luke, *Mathematical Functions and Their Approximations* (Academic, New York, 1975), pp. 154–252.

Angular momentum analyticity of Schrödinger energy levels for quark confining potentials

D. P. Datta
Department of Physics, North Bengal University, District Darjeeling, India 734430

S. Mukherjee
International Center for Theoretical Physics, Trieste, Italy and
Department of Physics, North Bengal University, District Darjeeling, India 734430[a]

(Received 8 October 1981; accepted for publication 7 December 1981)

Using the Kato–Rellich perturbation theory for linear operators, the analyticity of the Schrödinger energy levels $E_n(l)$ in the half-plane Re $l > -\frac{1}{2}$ has been proved for a large class of quark-confining potentials.

PACS numbers: 03.65.Ge, 12.40.Qq

1. INTRODUCTION

The problem of the analyticity of the Schrödinger energy levels $E_n(l)$ in the complex l plane for the potential

$$V(r) = r^\alpha, \quad \alpha > 0 \tag{1.1}$$

was recently examined by Grosse and Martin.[1] They showed that $E_n(l)$ are analytic in Re $l > -\frac{1}{2}$. In an earlier paper[2] we extended their results to include potentials which are superpositions of powers of the form

$$V(r) = r + k\beta r^\beta, \tag{1.2}$$

for two nonoverlapping regions of β: (i) for $0.4 < \beta < 1$, $k > 0$, and (ii) $-2 < \beta < 0.25$, k real. The proof of the analyticity for Case (i) followed the approach of Grosse and Martin.[1] For Case (ii), we made use of an argument based on the Kato–Rellich perturbation theory for linear operators. The class of potentials (1.2) is physically interesting because it includes, for example, the conventional charmonium potential

$$V(r) = -k/r + r, \quad k > 0. \tag{1.3}$$

It is, therefore, tempting to look for an argument which will cover the missing range of values of β, viz., $0.25 \leqslant \beta \leqslant 0.4$. In the present paper we present a variation of our earlier method, which proves the analyticity of $E_n(l)$ for the potential (1.2) for $0 \leqslant \beta < 1$. This, in conjunction with our earlier results,[2] now completes the proof for the entire range of values of interest, $-2 < \beta < 1$. The new proof is given in Sec. 2. In Sec. 3, we present an extension of our results for a general superposition of power potentials for suitable ranges of powers.

2. SUPERPOSITION OF TWO POTENTIALS

Consider the Hilbert space $L^2(0, \infty)$ of square integrable functions. Let C_0^∞ be the subspace of infinitely differentiable functions with compact support. Let

$$\dot{H} = -\frac{d^2}{dx^2} + \frac{\lambda}{x^2} + x + k\beta x^\beta, \qquad \lambda = l(l+1), \quad \beta > -2, \tag{2.1}$$

be the minimal operator on C_0^∞. For real l, $\dot{H}$ is essentially self-adjoint if and only if $l \geqslant \frac{1}{2}$. Thus, for the self-adjoint extension of $\dot{H}$ with $l > -\frac{1}{2}$ the natural choice is the Friedrichs extension

$$H(\beta, \lambda) = -\frac{d^2}{dx^2} + \frac{\lambda}{x^2} + x + k\beta x^\beta, \tag{2.2}$$

$$D(H) = \{u \mid u, u' \text{ absolutely continuous} \mid u, u' \in L^2(0, \infty) \mid u(0) = 0 \mid Hu \in L^2(0, \infty)\}. \tag{2.3}$$

To realize the associated closed operator for each complex l with Re $l > -\frac{1}{2}$ we define the minimal operator

$$\dot{H}_0 = -\frac{d^2}{dx^2} + \frac{\lambda}{x^2} + x, \quad D(\dot{H}_0) = C_0^\infty. \tag{2.4}$$

For simplicity, we shall henceforth work on the λ plane cut along $-\frac{1}{4}$ to $-\infty$.

Lemma 1: For each fixed λ in the cut plane, $\dot{H}_0$ is a densely defined sectorial operator.

Proof: Write

$$\dot{H}_0^\alpha = -\frac{d^2}{dx^2} + \frac{\lambda}{x^2} + \alpha x, \quad \alpha > 0 \tag{2.5}$$

$$D(\dot{H}_0^\alpha) = C_0^\infty.$$

Let us define a unitary transformation

$$[U(\eta)f](x) = \eta^{1/2} f(\eta x), \quad \eta > 0. \tag{2.6}$$

For $\eta = \alpha^{1/3}$, this states that $\dot{H}_0^\alpha$ and $\alpha^{2/3}\dot{H}_0$ are unitarily connected[3]:

$$\dot{H}_0^\alpha = U(\alpha^{2/3}\dot{H}_0)U^{-1}. \tag{2.7}$$

An application of Hardy's inequality

$$\int |u'|^2 dx \geqslant \frac{1}{4}\int x^{-2}|u|^2 dx, \quad u(0) = 0 \tag{2.8}$$

gives

$$\begin{aligned} \mathrm{Re}(\dot{H}_0^\alpha u, u) &= \int |u'|^2 dx + \mu \int x^{-2}|u|^2 dx \\ &\quad + \alpha \int x|u|^2 dx, \quad \mu = \mathrm{Re}\,\lambda \\ &\geqslant (\tfrac{1}{4} + \mu)\int x^{-2}|u|^2 dx \\ &\quad + \alpha \int x|u|^2 dx. \end{aligned} \tag{2.9}$$

[a] Permanent address.

 0022-2488/83/030573-03$02.50

For $\mu \geq -\frac{1}{4}$, $\mathrm{Re}(\dot{H}_0^{\alpha} u, u) > 0$ for each α. On the other hand, if $\mu < -\frac{1}{4}$, $\mathrm{Re}(\dot{H}_0^{\alpha} u, u) > 0$ for some α. Hence, in any case Eq. (2.7) shows that $\mathrm{Re}(\dot{H}_0 u, u) > 0$, since a unitary transformation preserves expectation values. Thus the expectation values of $\dot{H}_0$ lie in a sector whose opening angle is strictly less than π. This proves the lemma.

Now for each densely defined sectorial operator T there exists[4] a closed (m-sectorial) extension T_F (the Friedrichs extension). In the present case we have the extension [see, e.g., Ex. 2.17, Chap. VI of Ref. 4]

$$\dot{H}_{0F} = H(0, \lambda), \quad D(\dot{H}_{0F}) = D(H).$$

Lemma 2: Let $u \in D(H)$. Then for any compact set in the λ cut plane the relation

$$\|xu\| \leq \|H(0, \lambda)u\| \tag{2.10}$$

holds.

Proof: As remarked by Simon[3] it will be sufficient to consider $u \in C_0^{\infty}$, since C_0^{∞} is a core of $D(H)$. As quadratic forms on $C_0^{\infty} \otimes C_0^{\infty}$, we can write

$$\begin{aligned} \dot{H}_0^{\alpha} \dot{H}_0^{\alpha +} &= \left(p^2 + \frac{\lambda}{x^2} + \alpha x\right)\left(p^2 + \frac{\bar{\lambda}}{x^2} + \alpha x\right) \\ &= \left(p^2 + \frac{\lambda}{x^2}\right)\left(p^2 + \frac{\bar{\lambda}}{x^2}\right) \\ &\quad + \alpha^2 x^2 + \alpha[p, [p, x]] \\ &\quad + 2\alpha p x p + \frac{2\alpha\mu}{x} \\ &\geq \alpha^2 x^2 + \frac{2\alpha\mu}{x}, \end{aligned} \tag{2.11}$$

since $(p^2 + \lambda/x^2)(p^2 + \bar{\lambda}/x^2)$ is a nonnegative self-adjoint operator. The manipulation with the operators done above are valid on C_0^{∞}. Taking expectation values and making use of the relation (2.7), we get

$$\alpha^{4/3} \|\dot{H}_0 u\|^2 \geq \alpha^2 \|xu\|^2 + 2\alpha\mu \|x^{-1/2} u\|^2. \tag{2.12}$$

Now for each λ in the cut plane and each $u \in C_0^{\infty}$ we can find $\alpha > 0$ such that

$$(\alpha^2 - \alpha^{4/3}) \|xu\|^2 + 2\alpha\mu \|x^{-1/2} u\|^2 \geq 0. \tag{2.13}$$

Hence for α satisfying Eq. (2.13), Eq. (2.12) becomes

$$\alpha^{4/3} \|\dot{H}_0 u\|^2 \geq \alpha^{4/3} \|xu\|^2, \tag{2.14}$$

which holds for any finite λ. The result can now be extended to any $u \in D(H)$.

Lemma 3: $x^{\beta} (0 \leq \beta < 1)$ is $H(0, \lambda)$-bounded with relative bound zero.

Proof: This follows from the fact that one can choose constants $a (0 < a < 1)$ and b (sufficiently large) so that

$$x^{\beta} \leq ax + b, \quad x > 0. \tag{2.15}$$

It follows that

$$\begin{aligned} \|x^{\beta} u\|^2 &\leq a \|xu\|^2 + b \|u\|^2 \\ &\leq a \|H(0, \lambda)u\|^2 + b \|u\|^2, \qquad u \in D(H), \end{aligned} \tag{2.16}$$

where use has been made of Lemma 2. Lemma 3 now follows from the standard definition.

Corollary 1: For each fixed λ in the cut plane, $H(\beta, \lambda)$ is a holomorphic family of type A in β $(0 \leq \beta < 1)$ with compact resolvents.

Proof: Since $H(0, \lambda)$ is closed and x^{β} is $H(0, \lambda)$-bounded with relative bound zero, the result follows from the standard criterion.[5]

A similar result was derived in an earlier paper.[2] It was shown that the operator family $H(\beta, \lambda)$ is homomorphic (of type A) in β in the interval $-2 < \beta < \frac{1}{4}$ for each fixed λ in the cut plane. Hence, by virtue of the unique analytic continuation property, it follows that $H(\beta, \lambda)$ is a holomorphic family (of type A) in β in the entire interval $-2 < \beta < 1$ for any λ in the cut plane. A consequence of this is that every nondegenerate eigenvalue of $H(\beta, \lambda)$ is analytic in β near $\beta = \beta_0$ with $-2 < \beta_0 < 1$. That every eigenvalue of $H(\beta, \lambda)$ is nondegenerate can be proved as in Ref. 2. The nonexistence of any branch point (or natural boundary) of $E_n(l)$ in the half-plane $\mathrm{Re}\, l > -\frac{1}{2}$ then follows by the method given in Ref. 2. Essentially the argument can be summarized as follows. Since the eigenvalue $\epsilon_n(l)$ for the operator $H(0, \lambda)$ does not have such singularities in the half-plane $\mathrm{Re}\, l > -\frac{1}{2}$, the eigenvalue $E_n(l)$ of $H(\beta, \lambda)$ also cannot have these types of singularities because of its analyticity property in β. The eigenvalue $E_n(l)$ is free from isolated poles or essential singularities[1] because of its Herglotz property, $\mathrm{Im}\, E_n(l)/\mathrm{Im}\, l > 0$. We can thus write down our main result:

Theorem 1: The energy level $E_n(l)$ in the potential

$$V(r) = r + k\beta r^{\beta}, \quad k \text{ real}, \quad -2 < \beta < 1$$

is analytic in l in the half-plane $\mathrm{Re}\, l > -\frac{1}{2}$.

3. SUPERPOSITION OF A NUMBER OF POTENTIALS

In proving the analyticity in the cut λ plane of the energy levels E_n for the superposition of a number of power potentials, we make use of a theorem of Kato, which may be stated as follows.

Let A be T-bounded with relative bound $a < 1$. Then any operator which is T-bounded with relative bound b is also $(T + A)$-bounded with relative bound $\leq b(1 - a)^{-1}$. In Ref. 2, we showed that x^{-1} is $H(0, \lambda)$-bounded with relative bound zero. Therefore, if we define

$$\tilde{H}(0, k_1, \lambda) = H(0, \lambda) + k_1 x^{-1}, \quad k_1 \text{ real} \tag{3.1}$$

it is clear that $x^{\beta} (-2 < \beta < 1)$ will be $\tilde{H}(0, k_1, \lambda)$-bounded with relative bound zero for any fixed λ in the cut plane. Hence, we conclude:

For any fixed λ in the cut plane,

$$\tilde{H}(\beta, k_1, \lambda) = H(0, \lambda) + k_1 x^{-1} + k\beta x^{\beta}, \quad k \text{ real} \tag{3.2}$$

is a holomorphic family of type A in β for $-2 < \beta < 1$.

This means that each nondegenerate eigenvalue of $\tilde{H}(\beta, k_1, \lambda)$ is analytic in β in the interval $(-2, 1)$. The nondegeneracy of each eigenvalue $E_n(\beta)$ for fixed real k_1 and λ in the cut plane can be proved by noting that the dimension of the range space of the projection operator $P_n(\beta)$ associated with $E_n(\beta)$ is unity for $-2 < \beta < 1$. Thus the algebraic multiplicity of $E_n(\beta)$ for each β in the interval $-2 < \beta < 1$ is one and hence it is nondegenerate. The β analyticity thus obtained for each energy level now guarantees that the angu-

lar momentum analyticity of $E_n(l)$, which is proved in Theorem 1 for $\beta = 0$, should hold uniformly throughout the interval $-2 < \beta < 1$. For details we refer to Ref. 2.

It is now clear that the above steps repeated a suitable number of times will also prove the analyticity of the energy level $E_n(l)$ in Re $l > -\frac{1}{2}$ for a class of potentials of the form

$$V(r) = r + \sum_i k_i r^{\beta_i} + k\beta r^{\beta}, \quad -2 < \beta_i, \beta < 1. \tag{3.3}$$

It may also be noted that the present method can be used to prove the analyticity of the energy level $E_n(l)$ for potentials

$$V(r) = r^{\alpha} + \sum_i k_i r^{\beta_i} + k\beta r^{\beta}, \ \alpha > 0, \quad 0 \leqslant \beta_i, \beta < \alpha. \tag{3.4}$$

Superposition of terms with powers in the interval $-2 < \beta < 0$ is missing since it is difficult to translate the corresponding argument[2] for arbitrary $\alpha > 0$.

ACKNOWLEDGMENTS

One of us (S. M.) would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Center for Theoretical Physics, Trieste, Italy. We are also thankful to the Department of Atomic Energy, Government of India, for financial support.

[1] H. Grosse and A. Martin, Phys. Rep. **60**, 343 (1980).

[2] D. P. Datta and S. Mukherjee, J. Math. Phys. **23**, 665 (1982), and erratum **24** (4) (1983).

[3] B. Simon, Ann. Phys. (N. Y.) **58**, 76 (1970).

[4] T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1976).

[5] Reference 4, Theorem 2.6, pp. 377–8.

Scattering theory in arrangement channel quantum mechanics

J. W. Evans and D. K. Hoffman
Ames Laboratory[a] *and Department of Chemistry, Iowa State University, Ames, Iowa 50011*

D. J. Kouri[b]
Department of Chemistry, University of Houston—Central Campus, Houston, Texas 77004

(Received 1 June 1982; accepted for publication 17 September 1982)

The time-independent scattering theory associated with the non-self-adjoint matrix Hamiltonians $\underline{\underline{H}}$ of arrangement channel quantum mechanics is presented in detail first using the 3-particle case as an example. A key feature is the biorthogonality of a suitably constructed set of scattering eigenvectors and duals. Channel space Möller operators, $\underline{\underline{S}}$- and $\underline{\underline{T}}$-matrices are defined and a variety of properties investigated including the way multichannel unitarity is imbedded into the theory. Some remarks on the time-dependent theory are also made. A detailed discussion of channel space density matrix scattering theory (of interest, e.g., in reactive kinetic theory) is presented using the Liouville representation. We describe some special cases including the exclusion of breakup and 2×2 choices of three particle $\underline{\underline{H}}$.

PACS numbers: 03.65.Nk, 03.80. + r

1. INTRODUCTION

Inherent difficulties with the standard Lippmann–Schwinger equation approach to many-body scattering theory[1] have recently lead to the development of a variety of alternative approaches.[2] Most of these implement various forms of decomposition of the wavefunction or T-matrices to obtain "well behaved" scattering equations. In this work, we present an analysis of the scattering theory pertaining to the "arrangement channel quantum mechanics" (ACQM) approach.[3] Thus, we consider a system of N distinguishable particles characterized by a Hamiltonian H (with center of mass kinetic energy removed) acting on the N-particle Hilbert space $\mathscr{H}$. The ACQM theory is characterized by a matrix Hamiltonian $\underline{\underline{H}}$ with operator valued components $H_{\alpha\beta}$ where $\alpha,\beta,\cdots$ belong to some subset of arrangement channels (clusterings) for the N particles. Typically, $\underline{\underline{H}}$ is not self-adjoint (or even normal) but its components satisfy the summation condition[4]

$$\sum_{\alpha} H_{\alpha\beta} = H \quad \text{for all } \beta. \tag{1.1}$$

Other constraints will be described later.

The channel space form of the Schrödinger equation becomes

$$(\underline{\underline{H}} - \lambda)\underline{\psi} = \underline{0} \quad \text{(time independent)}, \tag{1.2}$$

$$\left(\underline{\underline{H}} - i\hbar\frac{\partial}{\partial t}\right)\underline{\psi} = \underline{0} \quad \text{(time dependent)}, \tag{1.3}$$

where $\underline{\psi}$ is a vector in the channel space $\mathscr{C} = \oplus_{\alpha}\mathscr{H}$ with components $|\psi_\alpha\rangle \in \mathscr{H}$. Summing over the components of the rhs of these equations and using (1.1), it follows that either $\Sigma_\alpha |\psi_\alpha\rangle \neq 0$ and satisfies the corresponding Hilbert space equation or $\Sigma_\alpha |\psi_\alpha\rangle = 0$. In the context of (1.2), the former are called "physical" eigenvectors for which $\lambda = E \in \mathbb{R}$ and the latter "spurious."

Recently, there have been some extensive investigations into the spectral and semigroup theory for channel space Hamiltonian $\underline{\underline{H}}$.[5,6] Although these are expected to be typically scalar spectral having a complete set of physical and spurious eigenvectors, this as yet has only been demonstrated for the Faddeev case (with some technical assumptions).[6] Consequently, it is appropriate to first consider the time-independent scattering theory where we deal directly with $\underline{\underline{H}}$-eigenvectors (rather than $e^{i/\hbar\underline{\underline{H}}t}$ as in the time-dependent theory). In fact, the original motivation for the introduction of the $\underline{\underline{H}}$ operators was to provide a "well behaved" set of time-independent scattering equations amenable to numerical solution.

If we denote by H_α the α channel Hamiltonian so $H = H_x + V^\alpha$ for all α (where V^α consists of those potentials external to channel α), then it is natural to make the decomposition

$$\underline{\underline{H}} = \underline{\underline{H}}_0 + \underline{\underline{V}}, \tag{1.4}$$

where $[\underline{\underline{H}}_0]_{\alpha\beta} = \delta_{\alpha\beta}H_\alpha$ and $[\underline{\underline{V}}]_{\alpha\beta}$ involve potentials guaranteeing (1.1). The operator $\underline{\underline{V}}$ is typically chosen so that the kernel of the resulting scattering equation

$$\underline{\psi} = \underline{\phi} + \underline{\underline{G}}_0^{\pm}(E)\underline{\underline{V}}\underline{\psi} \tag{1.5}$$

is "well behaved." Here $(E - \underline{\underline{H}}_0)\underline{\phi} = \underline{0}$, $\underline{\underline{G}}_0^{\pm}(E) = \lim_{\epsilon\to 0+}\underline{\underline{G}}_0(E \pm i\epsilon)$, and $\underline{\underline{G}}_0(z) = (z - \underline{\underline{H}}_0)^{-1}$. One usually demands that the kernel $\underline{\underline{G}}_0^{\pm}(E)\underline{\underline{V}}$ is connected after a finite number of iterations[2,5,7] guaranteeing uniqueness of the scattering solutions of (1.5) (to within any normalizable $\underline{\underline{H}}$-eigenvectors of the same "imbedded" eigenvalue). This requirement is further motivated by the "fiber compactness assumption"[7,8] that for reasonable potentials connectivity of some iterate leads to compactness of that or a higher (finite) iterate. This guarantees that standard numerical solution techniques are applicable. In our work we further elucidate the role of these conditions.

In Sec. 2, we investigate the scattering theory of both

[a] Operated for the U. S. Department of Energy by Iowa State University under contract No. W-7405-eng-82. This research was supported by the Office of Basic Energy Sciences.
[b] Supported under National Science Foundation Grant CHE79-19098.

 0022-2488/83/030576-12$02.50

the eigenvectors and their corresponding duals for a general 3-particle, 3- (2-cluster) channel $\underline{\underline{H}}$ (remembering that $\underline{\underline{H}}$ is not self-adjoint). A biorthogonality property is proved and channel space Möller operators defined. The corresponding $\underline{\underline{S}}$- and $\underline{\underline{T}}$-matrices are introduced in Sec. 3 and various properties, including multichannel unitarity, investigated. Some remarks on the corresponding time-dependent theory are made in Sec. 4. A scattering theory for channel space density matrices is presented in Sec. 5 using the Liouville representation. Various expressions for the transition super operator (which is related to the collision operator in the corresponding reactive Boltzmann equations) are provided. In Sec. 6, some modifications of the above 3-particle case and the extension to the N-particle case are discussed.

2. SCATTERING EQUATIONS FOR EIGENVECTORS OF $\underline{\underline{H}}$ AND THEIR DUALS

For purposes of illustration, in Secs. 2–5 we consider a system of three distinguishable particles labeled 1,2,3. Furthermore, we consider only 3×3 channel space Hamiltonians $\underline{\underline{H}}$ with components $H_{\alpha\beta}$ labeled by the *two cluster* arrangement channels, namely (1)(2 3), (2)(1 3), (3)(1 2). The discussion of Secs. 2–5 will go through with virtually no constraints on the choice of $\underline{\underline{V}}$ other than that the required solutions to the scattering equation (1.5) exist. However, the advantage of the "connected" choices is that any nonuniqueness of the scattering solutions must correspond to spatially confined $\underline{\underline{H}}$-eigenvectors of the same eigenvalue as may be seen by iteration of the homogeneous equation. This suggests that *all* the scattering eigenvectors of $\underline{\underline{H}}$ are accounted for by the appropriate inhomogeneous solutions of (1.5) in contrast to the standard Lippmann–Schwinger equations. Clearly, the latter type of nonuniqueness problems will arise for a continuous range of energies if some $H_{\alpha\alpha}$ can support bound states in channels other than α, and in Appendix A an example is given where $H_{\alpha\alpha}=H_\alpha$ but disconnectivity still leads to such nonuniqueness.

Another consequence of a "connected" choice of $\underline{\underline{V}}$ is a partial interpretational property of the components of $\underline{\psi}$. Specifically, it follows that the different two cluster (pair bound state) parts of the Hilbert space wavefunction are contained asymptotically only in the appropriate channel components. This may not hold for a disconnected choice of $\underline{\underline{V}}$ even where $H_{\alpha\alpha}=H_\alpha$ (see Appendix B).

It is convenient to characterize the eigenvectors of H assuming asymptotic completeness. If $\alpha=(i)(j\,k)$ and bound states of $(j\,k)$ exist, then H has weak (scattering) eigenvectors $|\psi_\alpha^\pm\rangle$ where $+/-$ denotes an asymptotically prepared pre-/post-collisional state with $(j\,k)$ bound and (i) free. This asymptotic state is denoted $|\phi_\alpha\rangle$ and is an eigenvector of H_α with the same energy (eigenvalue) as $\psi_\alpha^\pm$. There are also eigenvectors $|\psi_0^\pm\rangle$ where $+/-$ denotes that the three particles are prepared pre-/post-collisionally asymptotically free. The asymptotic state here is a plane wave $|\phi_0\rangle$ of the same energy (eigenvalue) as $|\psi_0^\pm\rangle$. For simplicity, we will suppress state labels in this work. There may also be true 3-particle bound state eigenvectors $|\psi_n\rangle$ of H.

We now consider the weak (scattering) eigenvectors and dual eigenvectors of $\underline{\underline{H}}$. We recall that the components of any eigenvector of $\underline{\underline{H}}$ either sum to an H-eigenvector with the same real eigenvalue (physical) or sum to zero (spurious).[4,5] Also, by taking any eigenvector of H and constructing the corresponding equal component dual channel space vector, one obtains a dual eigenvector of $\underline{\underline{H}}$ with the same eigenvalue.[4,5] The asymptotic 2-cluster scattering eigenvectors $\underline{\psi}_\alpha^\pm$ corresponding to $|\psi_\alpha^\pm\rangle$, with eigenvalue E, satisfy

$$\underline{\psi}_\alpha^\pm=\underline{\phi}_\alpha+\underline{\underline{G}}_0^\pm(E)\underline{\underline{V}}\underline{\psi}_\alpha^\pm, \tag{2.1}$$

where $[\underline{\phi}_\alpha]_\beta=\delta_{\alpha\beta}|\phi_\alpha\rangle$ so $(E-\underline{\underline{H}}_0)\underline{\phi}_\alpha=\underline{0}$. We also have linearly independent eigenvectors $\underline{\psi}_{0(j)}^\pm, j=1,2,3$, with eigenvalue E, satisfying

$$\underline{\psi}_{0(j)}^\pm=\underline{\phi}_{0(j)}^\pm+\underline{\underline{G}}_0^\pm(E)\underline{\underline{V}}\underline{\psi}_{0(j)}^\pm, \tag{2.2}$$

where $(E-\underline{\underline{H}}_0)\underline{\phi}_{0(j)}^\pm=\underline{0}$ and $\underline{\phi}_{0(j)}^\pm\sim\underline{\theta}^j|\phi_0\rangle$ in the breakup region. Here $\underline{\theta}^j$ are linearly independent. We may, for example, choose $\underline{\theta}^j=\underline{e}^j$ where $(\underline{e}^j)_\alpha=\delta_{\alpha j}$ so $\underline{\psi}_{0(j)}^\pm=\underline{\psi}_{j0}^\pm$ in the notation of Ref. 5. Another significant choice is where one of the $\underline{\psi}_{0(j)}^\pm$ is physical and the other two spurious, e.g., $\Sigma_\alpha\theta^j{}_\alpha$ is zero if $j=1,2$ and nonzero if $j=3$ (so the physical scattering eigenvectors of $\underline{\underline{H}}$ are in 1-1 correspondence with those of H).[6] We term such a choice canonical. By considering all three $\underline{\psi}_{0(j)}^\pm$ instead of just $j=3$ (which is sufficient to "represent" the $|\psi_0^\pm\rangle$ physics), the inhomogeneous terms in (2.1) and (2.2) constitute a complete set of $\underline{\underline{H}}_0$-eigenvectors. This will allow definition of corresponding channel space Möller operators, independent of $\underline{\theta}^j$, on the whole of $\mathscr{C}$.

The asymptotic 2-cluster dual scattering eigenvectors $\underline{\zeta}_\alpha^{\pm\prime}$ of $\underline{\underline{H}}$ chosen with all components equal to $\langle\psi_\alpha^\pm|$ satisfy

$$\underline{\zeta}_\alpha^{\pm\prime}=\underline{\zeta}_\alpha'+\underline{\zeta}_\alpha^{\pm\prime}\underline{\underline{V}}\underline{\underline{G}}_0^\mp(E), \tag{2.3}$$

where $[\underline{\zeta}_\alpha']_\beta=\delta_{\alpha,\beta}\langle\phi_\beta|$ so $\underline{\zeta}_\alpha'(E-\underline{\underline{H}}_0)=\underline{0}'$. We also construct linearly independent dual eigenvectors $\underline{\zeta}_{0(j)}^{\pm\prime}, j=1,2,3$ satisfying

$$\underline{\zeta}_{0(j)}^{\pm\prime}=\underline{\zeta}_{0(j)}^{\pm\prime}+\underline{\zeta}_{0(j)}^{\pm\prime}\underline{\underline{V}}\underline{\underline{G}}_0^\mp(E), \tag{2.4}$$

where $\underline{\zeta}_{0(j)}^{\pm\prime}(E-\underline{\underline{H}}_0)=\underline{0}'$ and $\underline{\zeta}_{0(j)}^{\pm\prime}\sim\underline{\Phi}^{j\prime}\langle\phi_0|$ in the breakup region. Here $\underline{\Phi}^{j\prime}$ are biorthogonal to the $\underline{\theta}^j$. If we make the above canonical choice of $\underline{\theta}^j$, then $\underline{\Phi}^{3\prime}$ and $\underline{\zeta}_{0(3)}^{\pm\prime}$ have equal components.

In this work, we assume that all of the above integral equations have suitable solutions which do, in fact, correspond to weak eigenvectors of $\underline{\underline{H}}$. The type of analysis required to prove this rigorously is described for 2-particle scattering in Refs. 9 and 10. If there is any nonuniqueness in these equations, it is assumed a single appropriate solution is chosen. It has been observed previously[5,6] that after explicitly setting the components of $\underline{\zeta}_\alpha^{\pm\prime}$ equal, (2.3) reduce to LS–GT equations.[11]

Standard manipulations can be performed for the above equations. For example, any solutions $\underline{\psi},\underline{\zeta}'$ of

$$\underline{\psi}=\underline{\phi}+\underline{\underline{G}}_0^\pm(E)\underline{\underline{V}}\underline{\psi},\quad \underline{\zeta}'=\underline{\zeta}'+\underline{\zeta}'\underline{\underline{V}}\underline{\underline{G}}_0^\mp(E), \tag{2.5}$$

where $(E-\underline{\underline{H}}_0)\underline{\phi}=\underline{0}$, $\underline{\zeta}'(E-\underline{\underline{H}}_0)=\underline{0}'$ satisfy

$$\underline{\psi}=(\underline{\underline{1}}+\underline{\underline{G}}^\pm(E)\underline{\underline{V}})\underline{\phi},\quad \underline{\zeta}'=\underline{\zeta}'(\underline{\underline{1}}+\underline{\underline{V}}\underline{\underline{G}}^\mp(E)), \tag{2.6}$$

respectively. Here $\underline{\underline{G}}^\pm(E)=\lim_{\epsilon\to 0+}\underline{\underline{G}}(E\pm i\epsilon)$, $\underline{\underline{G}}(z)=(z-\underline{\underline{H}})^{-1}$ and we have the standard resolvent equations

$$\underline{G}(z) = \underline{G}_0(z) + \underline{G}_0(z)\underline{V}\underline{G}(z) = \underline{G}_0(z) + \underline{G}(z)\underline{V}\underline{G}_0(z). \quad (2.7)$$

It is natural to ask whether the scattering eigenvectors and duals constructed above are biorthogonal in a generalized "delta-function" sense. This is readily verified if the dual vector has equal components. It is also anticipated to be true in general since the corresponding inhomogeneous terms in the integral equations are constructed to be biorthogonal. If the scattering eigenvectors of $\underline{H}$ described above together with all physical and spurious normalizable eigenvectors of $\underline{H}$ form a basis for the channel space (so $\underline{H}$ is scalar spectral), then biorthogonality is immediately verifiable.[4,5] However, it is possible to show that biorthogonality of the scattering eigenvectors holds *without* this assumption (Appendix C).

We now define and investigate the properties of the channel space Möller operator from a time-independent perspective. Firstly, $\underline{\Omega}^{\pm}$ are defined on the entire channel space by

$$\underline{\psi}_{\alpha}^{\pm} = \underline{\Omega}^{\pm}\underline{\phi}_{\alpha}, \quad \underline{\psi}_{0(j)}^{\pm} = \underline{\Omega}^{\pm}\underline{\phi}_{0(j)}^{\pm}, \quad (2.8)$$

so

$$\underline{\Omega}^{\pm} = \sum_{\alpha}^{\int} \underline{\psi}_{\alpha}^{\pm}\underline{\xi}'_{\alpha} + \sum_{0(j)}^{\int} \underline{\psi}_{0(j)}^{\pm}\underline{\xi}_{0(j)}^{\pm\prime}, \quad (2.9)$$

where the $\int$ also include a sum/integral over state labels. Any ambiguity in (2.8) due to normalizable $\underline{H}$-eigenvectors imbedded into the continuum will not affect $\underline{\Omega}^{\pm}$ in (2.9) regarded as operators on $\mathscr{C}$ (since the imbedded eigenvalues are a set of measure zero). We denote $\mathscr{P}_{\text{scatt}}^{\pm} = \text{Range}(\underline{\Omega}^{\pm})$, the subspace spanned by the corresponding scattering eigenvectors and expect that $\mathscr{P}_{\text{scatt}}^{+} = \mathscr{P}_{\text{scatt}}^{-} = \mathscr{P}_{\text{scatt}}$ for reasonable systems.[9,10] Also $\mathscr{P}_{\text{scatt}}^{\pm} = \text{Range}(\underline{\mathscr{P}}_{\text{scatt}}^{\pm})$ where

$$\underline{\mathscr{P}}_{\text{scatt}}^{\pm} = \sum_{\alpha}^{\int} \underline{\psi}_{\alpha}^{\pm}\underline{\zeta}_{\alpha}^{\pm\prime} + \sum_{0(j)}^{\int} \underline{\psi}_{0(j)}^{\pm}\underline{\zeta}_{0(j)}^{\pm\prime} \quad (2.10)$$

are $\underline{H}$-invariant ($[\underline{\mathscr{P}}_{\text{scatt}}^{\pm}, \underline{H}] = 0$), nonorthogonal projection operators (from biorthogonality). From (2.6) and (2.7) one obtains expressions for $\underline{\Omega}^{\pm}$ in terms of the energy dependent operators $\underline{\Omega}^{\pm}(E)$. Explicitly,

$$\underline{\Omega}^{\pm}\underline{\phi}_{\beta} = \underline{\Omega}^{\pm}(E_{\beta})\underline{\phi}_{\beta}, \quad \underline{\Omega}^{\pm}\underline{\phi}_{0(j)}^{\pm} = \underline{\Omega}^{\pm}(E_{0(j)})\underline{\phi}_{0(j)}^{\pm}, \quad (2.11)$$

where $\underline{\Omega}^{\pm}(E) = \lim_{\epsilon\to 0+} \underline{\Omega}(E \pm i\epsilon)$ and

$$\underline{\Omega}(z) = \underline{1} + \underline{G}(z)\underline{V} = \underline{1} + \underline{G}_0(z)\underline{V}\underline{\Omega}(z). \quad (2.12)$$

Möller operators $\underline{\tilde{\Omega}}^{\pm}$ associated with the dual scattering eigenvectors are defined on $\mathscr{P}_{\text{scatt}}^{\pm}$ by

$$\underline{\zeta}_{\alpha}^{\pm\prime} = \underline{\xi}_{\alpha}^{\pm\prime}\underline{\tilde{\Omega}}^{\pm}, \quad \underline{\zeta}_{0(j)}^{\pm\prime} = \underline{\xi}_{0(j)}^{\pm\prime}\underline{\tilde{\Omega}}^{\pm}. \quad (2.13)$$

Also defining $\underline{\tilde{\Omega}}^{\pm}$ to be zero on Range $(\underline{1} - \underline{\mathscr{P}}_{\text{scatt}}^{\pm})$, we have

$$\underline{\tilde{\Omega}}^{\pm} = \sum_{\alpha}^{\int} \underline{\phi}_{\alpha}\underline{\zeta}_{\alpha}^{\pm\prime} + \sum_{0(j)}^{\int} \underline{\phi}_{0(j)}\underline{\zeta}_{0(j)}^{\pm\prime}. \quad (2.14)$$

Of course, Range $(\underline{\tilde{\Omega}}^{\pm})$ is the full channel space. From (2.6) and (2.7) one obtains expressions for $\underline{\tilde{\Omega}}^{\pm}$ in terms of the energy dependent operators $\underline{\tilde{\Omega}}^{\pm}(E)$. Explicitly,

$$\underline{\xi}'_{\alpha}\underline{\tilde{\Omega}}^{\pm} = \underline{\xi}'_{\alpha}\underline{\tilde{\Omega}}^{\pm}(E_{\alpha}), \quad \underline{\xi}_{0(j)}^{\pm\prime}\underline{\tilde{\Omega}}^{\pm} = \underline{\xi}_{0(j)}^{\pm\prime}\underline{\tilde{\Omega}}^{\pm}(E_{0(j)}), \quad (2.15)$$

where $\underline{\tilde{\Omega}}^{\pm}(E) = \lim_{\epsilon\to 0+} \underline{\tilde{\Omega}}(E \pm i\epsilon)$ and

$$\underline{\tilde{\Omega}}(z) = \underline{1} + \underline{V}\underline{G}(z) = \underline{1} + \underline{\tilde{\Omega}}(z)\underline{V}\underline{G}_0(z). \quad (2.16)$$

It is readily verified, from biorthogonality, that

$$\underline{\tilde{\Omega}}^{\pm}\underline{\Omega}^{\pm} = \underline{I}, \quad (2.17)$$

$$\underline{\Omega}^{\pm}\underline{\tilde{\Omega}}^{\pm} = \underline{\mathscr{P}}_{\text{scatt}}^{\pm}. \quad (2.18)$$

In closing this section, we remark that clearly $\underline{\Omega}^{\pm}$ and $\underline{\tilde{\Omega}}^{\pm}$ are independent of the choice of $\{\underline{\theta}^{j}, \underline{\Phi}^{j\prime}\}$ and satisfy the intertwining relations[3]

$$\underline{H}\underline{\Omega}^{\pm} = \underline{\Omega}^{\pm}\underline{H}_0, \quad \underline{\tilde{\Omega}}^{\pm}\underline{H}\underline{\mathscr{P}}_{\text{scatt}}^{\pm} = \underline{H}_0\underline{\tilde{\Omega}}^{\pm}. \quad (2.19)$$

3. CHANNEL SPACE $\underline{S}$- and $\underline{T}$-MATRICES

For our discussion of $\underline{S}$-matrices, we invoke the assumption that $\underline{\mathscr{P}}_{\text{scatt}}^{+} = \underline{\mathscr{P}}_{\text{scatt}}^{-}$, which is anticipated to hold for systems of interest. (In the 2-particle scattering theory, this question is analyzed using Möller operators and the corresponding identity termed "weak asymptotic completeness."[12]) We now define channel space scattering operators by[3]

$$\underline{S}^{\pm} = \underline{\tilde{\Omega}}^{\mp}\underline{\Omega}^{\pm}. \quad (3.1)$$

It then follows immediately from (2.17) and (2.18) that

$$\underline{S}^{+}\underline{S}^{-} = \underline{S}^{-}\underline{S}^{+} = \underline{I}. \quad (3.2)$$

We shall show that although $\underline{S}^{\pm}$ are not unitary (if the breakup channel is open), the identity (3.2) does incorporate the relationships typically associated with "unitarity" for the multichannel reactive scattering problem.[13]

First, we make a connection between the matrix elements of $\underline{S}^{\pm}$ and those of the various corresponding Hilbert space operators $S_{\alpha\beta}, S_{\alpha 0}^{\pm}$ $S_{0\beta}^{\pm}$, and $S_{00}^{\pm}$ using the obvious notation. Consider first

$$\begin{aligned}(\underline{S}^{\pm})_{\alpha\beta} &= (\underline{\xi}'_{\alpha}, \underline{S}^{\pm}\underline{\phi}_{\beta}) \\ &= (\underline{\zeta}_{\alpha}^{\mp\prime}, \underline{\psi}_{\beta}^{\pm}) \\ &= \langle \psi_{\alpha}^{\mp} | \psi_{\beta}^{\pm} \rangle \\ &= S_{\alpha\beta}^{\pm}, \end{aligned} \quad (3.3)$$

where again state labels have been suppressed and we have used the equal component property of $\underline{\zeta}_{\alpha}^{\mp\prime}$ and the summation property of $\underline{\psi}_{\beta}^{\pm}$. Further, it follows that

$$(\underline{S}^{\pm})_{\alpha\beta} = (\underline{S}^{\mp})_{\beta\alpha}{}^{*}. \quad (3.4)$$

For the other cases, it is convenient to use a *canonical* choice of $\underline{\theta}^{j}$ (and thus $\underline{\psi}_{0(j)}^{\pm}$) with $j = 3$ (say) physical. One may similarly verify that

$$\begin{aligned}(\underline{S}^{\pm})_{\alpha 0(j)} &= S_{\alpha 0}^{\pm}\delta_{j,3}, \\ (\underline{S}^{\pm})_{0(3)\beta} &= S_{0\beta}^{\pm}, \\ (\underline{S}^{\pm})_{0(3)0(j)} &= S_{00}^{\pm}\delta_{j,3}. \end{aligned} \quad (3.5)$$

The matrix elements with $0(k), k \neq 3$, as the bra are not simply determined. We have the additional relationships

$$(\underline{S}^{\pm})_{\alpha 0(3)} = (\underline{S}^{\mp})_{0(3)\alpha}{}^{*}(\underline{S}^{\pm})_{0(3)0(3)'} = (\underline{S}^{\mp})_{0(3)'0(3)}{}^{*}, \quad (3.6)$$

where in the second, we have indicated the interchange of state labels.

It may now be readily verified that $\underline{S}^{\pm}$ are *not* unitary for if this were the case it would require, for example, that when $\underline{\theta}^j = \underline{e}^j$,

$$(\underline{\xi}^{\pm}_{0(j)}, \underline{S}^{\pm}\underline{\phi}_\alpha) = (\underline{\xi}_\alpha, \underline{S}^{\mp}\underline{\phi}^{\mp}_{0(j)})^* = S^{\mp *}_{\alpha,0} \quad \text{for } j = 1,2,3, \tag{3.7}$$

which contradicts (3.6) noting $\underline{\xi}^{\pm}_{0(3)}{}'$ in that expression corresponds to $\Sigma_j \underline{\xi}^{\pm}_{0(j)}{}'$ here.

The multichannel reactive scattering "unitarity" relationships are now readily obtained. Firstly, using (3.2), (3.3), and (3.5) and thus a canonical choice of $\underline{\theta}^j$, we obtain

$$\begin{aligned}\delta_{\alpha,\beta} &= (\underline{\xi}'_\alpha, \underline{S}^{\pm}\underline{S}^{\mp}\underline{\phi}_\beta) \\ &= \sum_\gamma (\underline{\xi}'_\alpha, \underline{S}^{\pm}\underline{\phi}_\gamma)(\underline{\xi}'_\gamma, \underline{S}^{\mp}\underline{\phi}_\beta) \\ &\quad + \sum_{0(j)} (\underline{\xi}'_\alpha, \underline{S}^{\pm}\underline{\phi}^{\pm}_{0(j)})(\underline{\xi}^{\pm}_{0(j)}, \underline{S}^{\mp}\underline{\phi}_\beta) \\ &= \sum_\gamma S^{\pm}_{\alpha\gamma} S^{\mp}_{\gamma\beta} + \sum_0 S^{\pm}_{\alpha 0} S^{\mp}_{0\beta}, \end{aligned} \tag{3.8}$$

where only $j = 3$ contributes in the second term. Similarly one obtains

$$\begin{aligned} 0 &= (\underline{\xi}'_\alpha, \underline{S}^{\pm}\underline{S}^{\mp}\underline{\phi}^{\mp}_{0(3)}) \\ &= \sum_\gamma S^{\pm}_{\alpha\gamma} S^{\mp}_{\gamma 0} + \sum_{0'} S^{\pm}_{\alpha 0'} S^{\mp}_{0'0}, \end{aligned} \tag{3.9}$$

$$\begin{aligned} 0 &= (\underline{\xi}^{\mp}_{0(3)}{}', \underline{S}^{\pm}\underline{S}^{\mp}\underline{\phi}_\beta) \\ &= \sum_\gamma S^{\pm}_{0\gamma} S^{\mp}_{\gamma\beta} + \sum_{0'} S^{\pm}_{00'} S^{\mp}_{0'\beta}, \end{aligned} \tag{3.10}$$

$$\begin{aligned} \delta_{0,0'} &= (\underline{\xi}^{\mp}_{0(3)}{}', \underline{S}^{\pm}\underline{S}^{\mp}\underline{\phi}^{\mp}_{0(3)'}) \\ &= \sum_\gamma S^{\pm}_{0\gamma} S^{\mp}_{\gamma 0'} + \sum_{0''} S^{\pm}_{00''} S^{\mp}_{0''0'}. \end{aligned} \tag{3.11}$$

We remind the reader that for the canonical choice $\underline{\Phi}^{3\prime} = (1,1,1)$ but $\underline{\theta}^3$ can be any vector satisfying $\Sigma_\alpha \theta^3_\alpha = 1$.

Finally in this section, we make the connection between the $\underline{S}$-matrices described above and corresponding channel space $\underline{T}$-matrices. Consider first the $(\underline{S}^{\pm})_{\alpha\beta}$ matrix elements. Now since

$$\begin{aligned} \underline{\xi}^{\mp}_\alpha{}' &= \underline{\xi}'_\alpha(\underline{1} + \underline{V}\underline{G}^{\pm}(E_\alpha)) = \underline{\xi}^{\pm}_\alpha{}' + \underline{\xi}'_\alpha \underline{V}(\underline{G}^{\pm}(E_\alpha) \\ &\quad - \underline{G}^{\mp}(E_\alpha)), \end{aligned} \tag{3.12}$$

we have

$$\begin{aligned} (\underline{S}^{\pm})_{\alpha\beta} &= (\underline{\xi}^{\mp}_\alpha{}', \underline{\psi}^{\pm}_\beta) \\ &= (\underline{\xi}^{\pm}_\alpha{}', \underline{\psi}^{\pm}_\beta) + (\underline{\xi}'_\alpha, \underline{V}(\underline{G}^{\pm}(E_\alpha) \\ &\quad - \underline{G}(E^{\mp}_\alpha))\underline{\psi}^{\pm}_\beta) \\ &= \delta_{\alpha,\beta} \mp 2\pi i \delta(E_\alpha - E_\beta)(\underline{\xi}'_\alpha, \underline{V}\underline{\psi}^{\pm}_\beta), \end{aligned} \tag{3.13}$$

where we have used the formal identity from (A3),

$$(\underline{G}^{\pm}(E) - \underline{G}^{\mp}(E))\underline{\psi}^{\pm}_\beta = \mp 2\pi i \delta(E - E_\beta)\underline{\psi}^{\pm}_\beta. \tag{3.14}$$

Equation (3.13) motivates the definition of channel space $\underline{T}$-matrices,

$$\underline{T}^{\pm} = \underline{V}\underline{\Omega}^{\pm}, \tag{3.15}$$

so then

$$(\underline{S}^{\pm})_{\alpha\beta} = \delta_{\alpha\beta} \mp 2\pi i \delta(E_\alpha - E_\beta)(\underline{T}^{\pm})_{\alpha\beta}. \tag{3.16}$$

Instead by decomposing $\underline{\psi}^{\pm}_\beta$ in (3.13), analogous to (3.12), one obtains

$$(\underline{S}^{\pm})_{\alpha\beta} = \delta_{\alpha\beta} \mp 2\pi i \delta(E_\alpha - E_\beta)(\underline{\tilde{T}}^{\pm})_{\alpha\beta}, \tag{3.17}$$

where

$$\underline{\tilde{T}}^{\pm} = \underline{\tilde{\Omega}}^{\mp}\underline{V}. \tag{3.18}$$

A similar analysis again implementing a canonical choice $\underline{\theta}^j$, with $j = 3$ physical, shows

$$(\underline{S}^{\pm})_{\alpha 0(3)} = \mp 2\pi i \delta(E_\alpha - E_0)(\underline{T}^{\pm})_{\alpha 0(3)}, \tag{3.19}$$

$$(\underline{S}^{\pm})_{0(3)\beta} = \mp 2\pi i \delta(E_0 - E_\beta)(\underline{\tilde{T}}^{\pm})_{0(3)\beta}, \tag{3.20}$$

where in (3.19), we have decomposed $\underline{\xi}^{\pm}_\alpha{}'$ and in (3.20), $\underline{\psi}^{\pm}_\beta$, analogous to (3.12).

From (2.11), (2.12), (2.15), and (2.16) one obtains expressions for $\underline{T}^{\pm}$ and $\underline{\tilde{T}}^{\pm}$ in terms of energy dependent operators $\underline{T}^{\pm}(E)$. Explicitly, for any choice of $\underline{\theta}^j$,

$$\underline{T}^{\pm}\underline{\phi}_\beta = \underline{T}^{\pm}(E_\beta)\underline{\phi}_\beta, \quad \underline{T}^{\pm}\underline{\phi}^{\pm}_{0(j)} = \underline{T}^{\pm}(E_{0(j)})\underline{\phi}^{\pm}_{0(j)}, \tag{3.21}$$

$$\underline{\xi}'_\alpha \underline{\tilde{T}}^{\pm} = \underline{\xi}'_\alpha \underline{T}^{\pm}(E_\alpha), \quad \underline{\xi}^{\mp}_{0(j)}{}'\underline{\tilde{T}}^{\pm} = \underline{\xi}^{\mp}_{0(j)}{}'\underline{T}^{\pm}(E_{0(j)}), \tag{3.22}$$

where $\underline{T}^{\pm}(E) = \lim_{\epsilon \to 0+} \underline{T}(E \pm i\epsilon)$ and[3,14,15]

$$\begin{aligned} \underline{T}(z) &= \underline{V} + \underline{V}\underline{G}(z)\underline{V} = \underline{V} + \underline{V}\underline{G}_0(z)\underline{T}(z) \\ &= \underline{V} + \underline{T}(z)\underline{G}_0(z)\underline{V}. \end{aligned} \tag{3.23}$$

From (3.21) and (3.22) it follows that $\underline{T}^{\pm}$ and $\underline{\tilde{T}}^{\pm}$ coincide on the $\underline{H}_0$-energy-shell [consistent with (3.16) and (3.17)].

A reactive optical theorem is readily derived in a channel space setting from the identities

$$\begin{aligned} \underline{T}(E + i\epsilon) - \underline{T}(E - i\epsilon) &= \underline{V}(\underline{G}(E + i\epsilon) - \underline{G}(E - i\epsilon))\underline{V} \\ &= \underline{V}\underline{G}(E \pm i\epsilon)(-2i\epsilon)\underline{G}(E \mp i\epsilon)\underline{V} \\ &= \underline{T}(E \pm i\epsilon)\underline{G}_0(E \pm i\epsilon)(-2i\epsilon) \\ &\quad \times G_0(E \mp i\epsilon)\underline{T}(E \mp i\epsilon) \\ &= \underline{T}(E \pm i\epsilon)(\underline{G}_0(E + i\epsilon) \\ &\quad - G_0(E - i\epsilon))\underline{T}(E \mp i\epsilon). \end{aligned} \tag{3.24}$$

Thus, formally we can write

$$\underline{T}^{+}(E) - \underline{T}^{-}(E) = -2\pi i \underline{T}^{\pm}(E)\delta(E - \underline{H}_0)\underline{T}^{\mp}(E). \tag{3.25}$$

Finally, we remark that it follows from (3.16), (3.17), (3.19), (3.20), and (3.3) and (3.5) that we have agreement on the $\underline{H}_0$-energy-shell of the matrix elements of $\underline{T}^{\pm}, \underline{\tilde{T}}^{\pm}$ and those of the corresponding Hilbert space T-matrices for any choice of $\underline{H}$.

4. TIME-DEPENDENT SCATTERING THEORY FOR VECTORS AND DUAL VECTORS

Some significant complications occur in the time-dependent channel space scattering theory not present in treatments involving self-adjoint Hamiltonians. In solving the equations

$$i\hbar \frac{\partial}{\partial t}\underline{\psi} = \underline{H}\underline{\psi}, \tag{4.1}$$

$$-i\hbar\frac{\partial}{\partial t}\underline{\zeta}' = \underline{\zeta}'\underline{H}, \tag{4.2}$$

one must consider the sense in which $i\underline{H}$ generates a time evolution group (remembering $\underline{H}$ is typically not self-adjoint or even normal). It has been shown, with some weak constraints, that $-\underline{H}$ generates a holomorphic semigroup $e^{-z\underline{H}}$ in the *open* right half-plane. However, at present, it has only been proved that this semigroup extends to the imaginary z axis to yield a time-evolution group $e^{-i/\hbar\underline{H}t}$ when $\underline{V}$ is strictly bounded.[5] In contrast, we indicate below that, with some mild technical assumptions, a functional calculus for $\underline{H}$ (and thus $e^{-i/\hbar\underline{H}t}$) can always be constructed on a suitable subspace.

Define $\underline{P}^{\pm}_{\text{scatt}}$ to be the projection operator associated with $\underline{\psi}^{\pm}_{\alpha}$ and $\underline{\psi}^{\pm}_{0(3)}$ and their biorthogonal duals for a canonical choice of $\underline{\theta}^{j}$ with $j=3$ physical. If any bound states $|\psi_n\rangle$ of H are imbedded into $\underline{H}$-eigenvectors $\underline{\psi}_n$ and if $\underline{\zeta}'_n$ are the corresponding equal component duals, define $\underline{P}_{\text{bnd}} = \Sigma_n \underline{\psi}_n \underline{\zeta}'_n$. Further define $\underline{\mathscr{Q}}^{\pm}_{\text{scatt}} = \underline{\mathscr{P}}^{\pm}_{\text{scatt}} - \underline{P}^{\pm}_{\text{scatt}}$, the projection operator associated with spurious scattering solutions; $\underline{P}^{\pm} = \underline{P}_{\text{bnd}} + \underline{P}^{\pm}_{\text{scatt}}$, the physical projection operator of Refs. 4–6 and $\underline{\mathscr{P}}^{\pm} = \underline{P}_{\text{bnd}} + \underline{\mathscr{P}}^{\pm}_{\text{scatt}}$. All of these are $\underline{H}$-invariant. If we assume the eigenvectors associated with $\underline{\mathscr{P}}^{\pm}$ form part of a basis for $\mathscr{C}$, then it must be generalized Besselian from asymptotic considerations (i.e., convergent linear combinations are L^2) and is assumed generalized Hilbertian (i.e., all L^2 linear combinations are convergent).[16] This, in particular, guarantees that the projection operators and Möller operators (of Sec. 2) are bounded. Furthermore, $\underline{H}$ is real eigenvalue scalar spectral on Range $(\underline{\mathscr{P}}^{\pm})$ (and can be regarded as $*$-self-adjoint there where the involution $*$ is associated with the conjugate linear duality mapping taking eigenvectors to dual vectors). Previous corresponding results were stated only for Range $(\underline{P}^{\pm})$.[5] In particular, we have

$$\begin{aligned} e^{+i/\hbar\underline{H}t}\underline{\mathscr{P}}^{\pm} = e^{+i/\hbar\underline{\mathscr{P}}^{\pm}\underline{H}t} &= \sum_n e^{+i/\hbar E_n t}\underline{\psi}_n\underline{\zeta}'_n \\ &+ \sum_{\alpha}^{\mathrm{f}} e^{+i/\hbar E_\alpha t}\underline{\psi}^{\pm}_{\alpha}\underline{\zeta}^{\pm\prime}_{\alpha} \\ &+ \sum_{0(j)}^{\mathrm{f}} e^{+i/\hbar E_{0(j)}t}\underline{\psi}^{\pm}_{0(j)}\underline{\zeta}^{\pm\prime}_{0(j)} \end{aligned} \tag{4.3}$$

which, from the Hilbertian basis assumption, may be shown uniformly bounded on $\mathscr{C}$. If $\underline{\mathscr{Q}} = \underline{I} - \underline{\mathscr{P}}^{\pm}$ is zero as in the Faddeev case[6] or if the corresponding subspace is spanned by normalizable spurious eigenvectors of $\underline{H}$, then $\underline{H}$ is scalar spectral on the whole channel space[5] [so, e.g., (4.3) extends accordingly]. Even where this is not the case a functional calculus on the whole channel space may still be available.[17] If there are any complex eigenvalue spurious solutions, then $e^{i/\hbar\underline{H}t}$ will not be uniformly bounded.[5]

Now we give an heuristic discussion of the scattering behavior of vector wave pulses with various asymptotic clusterings. Firstly, let $\underline{\psi}^{\pm}_{\alpha}(t)$ be a wave pulse satisfying (4.1) with channel α stable asymptotic clustering $\underline{\phi}_{\alpha}(t)$ as $t\to\mp\infty$. Clearly $\underline{\phi}_{\alpha}(t)$ is evolved by $\underline{H}_0$ and $(\underline{\phi}_{\alpha}(t))_{\beta} = \delta_{\alpha\beta}|\phi_{\alpha}(t)\rangle$. Thus for suitably well-behaved potentials, we expect that

$$\underline{\psi}^{\pm}_{\alpha}(t) \overset{\mathrm{s}}{\sim} \underline{\phi}_{\alpha}(t) \quad \text{as } t\to\mp\infty, \tag{4.4}$$

i.e.,

$$e^{-i/\hbar(\underline{\mathscr{P}}^{\pm})\underline{H}t}\underline{\psi}^{\pm}_{\alpha}(0) - e^{-i/\hbar\underline{H}_0 t}\underline{\phi}_{\alpha}(0) \overset{\mathrm{s}}{\to} \underline{0} \quad \text{as } t\to\mp\infty, \tag{4.5}$$

where "s" indicates strong convergence and where $\underline{\mathscr{P}}^{\pm}$ (or $\underline{\mathscr{P}}^{\pm}_{\text{scatt}}$) may be inserted explicitly without change. Similarly let $\underline{\psi}^{\pm}_{0(j)}(t)$ be a wave pulse satisfying (4.1) corresponding to the three particles asymptotically free. The asymptotic behavior prior to sequential or simultaneous interaction between all three particles is characterized by $\underline{\phi}^{\pm}_{0(j)}(t)$ evolved by $\underline{H}_0$. Also $\underline{\phi}^{\pm}_{0(j)}(t) \sim \underline{\theta}^{j}|\phi_0(t)\rangle$ in the asymptotically free region (i.e., prior to interaction between *any* particles) where $|\phi_0(t)\rangle$ is evolved by the pure kinetic energy operator. Thus for suitable potentials, we expect that

$$\underline{\psi}^{\pm}_{0(j)}(t) \overset{\mathrm{s}}{\sim} \underline{\phi}^{\pm}_{0(j)}(t) \quad \text{as } t\to\mp\infty, \tag{4.6}$$

i.e.,

$$e^{-i/\hbar(\underline{\mathscr{P}}^{\pm})\underline{H}t}\underline{\psi}^{\pm}_{0(j)}(0) - e^{-i/\hbar\underline{H}_0 t}\underline{\phi}^{\pm}_{0(j)}(0) \overset{\mathrm{s}}{\to} 0 \quad \text{as } t\to\mp\infty. \tag{4.7}$$

From (4.5) and (4.7) we conclude that

$$\underline{\Omega}^{\pm} = \operatorname*{s-lim}_{t\to\mp\infty} e^{+i/\hbar\underline{\mathscr{P}}^{\pm}\underline{H}t}e^{-i/\hbar\underline{H}_0 t}, \tag{4.8}$$

and if in addition $i\underline{H}$ generates a *uniformly bounded* time evolution group, then also[3,5]

$$\underline{\Omega}^{\pm} = \operatorname*{s-lim}_{t\to\mp\infty} e^{+i/\hbar\underline{H}t}e^{-i/\hbar\underline{H}_0 t}. \tag{4.9}$$

A rigorous analysis of the conditions on the potentials required for convergence of (4.9) can be made using a modified Cook's method incorporating the uniform boundedness assumption followed by a stationary phase analysis.[5,9]

It is appropriate here to introduce the channel space interaction picture[3] defined by

$$\underline{\psi}_I(t) = e^{+i/\hbar\underline{H}_0 t}\underline{\psi}(t). \tag{4.10}$$

Then, for example,

$$\underline{\psi}^{\pm}_{\alpha}(0) = \underline{\Omega}(t)\underline{\psi}^{\pm}_{\alpha I}(t) = \underline{\Omega}^{\pm}\underline{\phi}_{\alpha}(0), \tag{4.11}$$

$$\underline{\psi}^{\pm}_{0(j)}(0) = \underline{\Omega}(t)\underline{\psi}^{\pm}_{0(j)I}(t) = \underline{\Omega}^{\pm}\underline{\phi}^{\pm}_{0(j)}(0), \tag{4.12}$$

where $\underline{\Omega}(t) = e^{+i/\hbar(\underline{\mathscr{P}}^{\pm})\underline{H}t}e^{-i/\hbar\underline{H}_0 t}$, and since

$$\underline{\psi}^{\pm}_{\alpha I}(t) \overset{\mathrm{s}}{\to} \underline{\phi}_{\alpha}(0), \quad \underline{\psi}^{\pm}_{0(j)I}(t) \overset{\mathrm{s}}{\to} \underline{\phi}^{\pm}_{0(j)}(0) \quad \text{as } t\to\mp\infty \tag{4.13}$$

for suitable potentials, it then follows that

$$\underline{\Omega}^{\pm} = \operatorname*{s-lim}_{t\to\mp\infty} \underline{\Omega}(t) \tag{4.14}$$

provided $\underline{\Omega}(t)$ is uniformly bounded (in agreement with the above analysis).

We consider now the corresponding aspects of the scattering of dual vector wave pulses. If $\underline{\zeta}'_{\alpha}(t)$ is a wave pulse satisfying (4.2) with channel α stable asymptotic clustering, then clearly

$$(\underline{\zeta}^{\pm\prime}_{\alpha}(t))_{\beta} = \langle\psi^{\pm}_{\alpha}(t)| \overset{\mathrm{s}}{\sim} \langle\phi_{\alpha}(t)| \text{ for all } \beta \text{ as } t\to\mp\infty. \tag{4.15}$$

In terms of the interaction picture which here takes the form

$$\underline{\zeta}'_I(t) = \underline{\zeta}'(t)e^{-i/\hbar \underline{H}_0 t}, \tag{4.16}$$

we have

$$(\underline{\zeta}'_{\alpha I}(t))_\beta \overset{s}{\sim} \langle \phi_\alpha(t)|e^{-i/\hbar H_\beta t} \overset{w}{\to} \delta_{\alpha\beta}\langle \phi_\alpha(0)| \quad \text{as } t \to \mp\infty. \tag{4.17}$$

Using obvious notation (4.17) becomes

$$\underline{\zeta}'_{\alpha I}(t) \overset{w}{\to} \underline{\xi}_\alpha(0) \quad \text{as } t \to \mp\infty. \tag{4.18}$$

Next let $\underline{\zeta}^{\pm}_{0(j)}(t)$ be a wave pulse satisfying (4.2) analogous to $\underline{\phi}^{\pm}_{0(j)}(t)$ with asymptotically free particles and let $\underline{\xi}^{\pm}_{0(j)}(t)$ be the corresponding $\underline{H}_0$-evolved asymptotic form analogous to $\underline{\phi}^{\pm}_{0(j)}(t)$. Then in contrast to (4.18) we expect that

$$\underline{\zeta}^{\pm}_{0(j)}(t) \overset{s}{\sim} \underline{\xi}^{\pm}_{0(j)}(t),$$

so

$$\underline{\zeta}^{\pm}_{0(j)I}(t) \overset{s}{\to} \underline{\xi}^{\pm}_{0(j)}(0) \quad \text{as } t \to \mp\infty, \tag{4.19}$$

from the same intuitive reasoning as in the $\underline{\phi}^{\pm}_{0(j)}(t)$ case. Now

$$\underline{\zeta}^{\pm}_{\alpha}(0) = \underline{\zeta}^{\pm}_{\alpha I}(t)\cdot e^{+i/\hbar \underline{H}_0 t}e^{-i/\hbar(\underline{\mathscr{P}}^{\pm})\underline{H}t} = \underline{\xi}^{\pm}_{\alpha}(0)\underline{\tilde{\Omega}}^{\pm}, \tag{4.20}$$

$$\underline{\zeta}^{\pm}_{0(j)}(0) = \underline{\zeta}^{\pm}_{0(j)I}(t)\cdot e^{+i/\hbar \underline{H}_0 t}e^{-i/\hbar(\underline{\mathscr{P}}^{\pm})\underline{H}t} = \underline{\xi}^{\pm}_{0(j)}(0)\underline{\tilde{\Omega}}^{\pm}, \tag{4.21}$$

which suggests (but does *not* prove) that

$$\underline{\tilde{\Omega}}^{\pm} = \text{w-lim}_{t\to\mp\infty} e^{+i/\hbar \underline{H}_0 t}e^{-i/\hbar(\underline{\mathscr{P}}^{\pm})\underline{H}t} \tag{4.22}$$

if the limit exists. In fact, this identity can *only* be satisfied if the partial componentwise interpretational property holds for the $\underline{\psi}$'s. In general the limit on the rhs is "less" than $\underline{\tilde{\Omega}}^{\pm}$ corresponding to outgoing 2-cluster flux lost to the wrong channels (see Appendix D).

5. SCATTERING THEORY FOR DENSITY MATRICES AND THE LIOUVILLE REPRESENTATION

The time-dependent Liouville/Von Neumann equation for a channel space density matrix $\underline{\varrho}(t)$ has the form[4]

$$i\hbar\frac{\partial}{\partial t}\underline{\varrho}(t) = [\underline{H}, \underline{\varrho}(t)]. \tag{5.1}$$

It is convenient to introduce the super-operators

$$\underline{\mathscr{L}}_0 = [\underline{H}_0, \ ], \ \underline{\mathscr{V}} = [\underline{V}, \ ],$$
$$\underline{\mathscr{L}} = \underline{\mathscr{L}}_0 + \underline{\mathscr{V}} = [\underline{H}, \ ]. \tag{5.2}$$

A further useful decomposition is illustrated by[18,19]

$$\underline{\mathscr{L}}_0 = \underline{\mathscr{L}}_0 - \underline{\mathscr{L}}_0^*, \quad \text{where } \underline{\mathscr{L}}_0 = \underline{H}_0 \otimes \underline{I},$$
$$\underline{\mathscr{L}}_0^* = \underline{I} \otimes \underline{H}_0^*,$$
$$\underline{\mathscr{V}} = \underline{\mathscr{V}} - \underline{\mathscr{V}}^*, \quad \text{where } \underline{\mathscr{V}} = \underline{V} \otimes \underline{I},$$
$$\underline{\mathscr{V}}^* = \underline{I} \otimes \underline{V}^*, \tag{5.3}$$

etc., using the notation $(\underline{A} \otimes \underline{B}^*)\underline{C} = \underline{A}\underline{C}\underline{B}$ (cf. Eu[20]). Provided $\underline{\varrho}(0) = \underline{\mathscr{P}}^{\pm}\underline{\varrho}(0)\underline{\mathscr{P}}^{\pm}$, the solution to (5.1) can be formally written as

$$\underline{\varrho}(t) = e^{-i/\hbar \underline{\mathscr{L}} t}(\underline{\mathscr{P}}^{\pm} \otimes \underline{\mathscr{P}}^{\pm *})\underline{\varrho}(0)$$
$$= e^{-i/\hbar(\underline{\mathscr{P}}^{\pm})\underline{H}t}\underline{\varrho}(0)e^{+i/\hbar(\underline{\mathscr{P}}^{\pm})\underline{H}t}. \tag{5.4}$$

In discussing density matrix scattering theory, we shall make use of the interaction picture

$$\underline{\varrho}_I(t) = e^{+i/\hbar \underline{\mathscr{L}}_0 t}\underline{\varrho}(t) = e^{+i/\hbar \underline{H}_0 t}\underline{\varrho}(t)e^{-i/\hbar \underline{H}_0 t}. \tag{5.5}$$

Consider a time-dependent scattering problem with packet-like, trace-class solution $\underline{\varrho}^{\pm}(t)$. The asymptotic condition corresponding to (4.13), (4.18), and (4.19) for collisions where the particle clustering is asymptotically resolved (e.g., the chemical composition of a reactive gas is typically precollisionally resolved in the Boltzmann regime) becomes

$$\underline{\varrho}_I^{\pm}(t) \to \underline{\varrho}_{I\text{as}}(0). \tag{5.6}$$

Here $\underline{\varrho}_{I\text{as}}(0)$ is diagonal with respect to the channel indices *of necessity* for the partially bound channels and can be so chosen for the breakup channel. The sense of the limit is described below. Formal manipulation of (5.6) yields

$$\underline{\varrho}^{\pm}(0) = \underline{\varrho}_I^{\pm}(0) = \underline{\Omega}^{\pm}\underline{\varrho}_{I\text{as}}(0), \tag{5.7}$$

where the Möller super-operators $\underline{\Omega}^{\pm}$ are given by

$$\underline{\Omega}^{\pm} = \lim_{t\to\mp\infty} e^{+i/\hbar \underline{\mathscr{L}} t}(\underline{\mathscr{P}}^{\pm} \otimes \underline{\mathscr{P}}^{\pm *})e^{-i/\hbar \underline{\mathscr{L}}_0 t}. \tag{5.8}$$

Jauch *et al.*[21] argue, in a Hilbert space scattering theory setting, that a physical statement of the asymptotic condition should involve density matrices with a limit in terms of a certain physical topology (which in their case corresponds to a trace class norm). The limit in (5.6) and (5.8) should be regarded appropriately.[5] If the potentials are such that $\underline{\Omega}^{\pm}$ and $\underline{\tilde{\Omega}}^{\pm}$ exist (a stronger condition than the existence of $\underline{\Omega}^{\pm}$, Ref. 21), then

$$\underline{\Omega}^{\pm}\underline{A} = \underline{\Omega}^{\pm}\underline{A}\underline{\tilde{\Omega}}^{\pm}. \tag{5.9}$$

From the Möller super-operators (5.8), we can define transition super-operators $\underline{\mathscr{T}}^{\pm}$, analogous to (3.15) and (3.18), by

$$\underline{\mathscr{T}}^{\pm} = \underline{\mathscr{V}}\underline{\Omega}^{\pm}. \tag{5.10}$$

Thus from (5.2) and (5.9), we have

$$\underline{\mathscr{T}}^{\pm}\underline{A} = \underline{T}^{\pm}\underline{A}\underline{\Omega}^{\pm} - \underline{\Omega}^{\pm}\underline{A}\underline{\tilde{T}}^{\mp}$$
$$= \underline{T}^{\pm}\underline{A} - \underline{A}\underline{\tilde{T}}^{\mp} + \underline{T}^{\pm}\underline{A}\underline{\tilde{T}}^{\mp}\underline{G}_0^{\mp}(E_b)$$
$$- \underline{G}_0^{\pm}(E_k)\underline{T}^{\pm}\underline{A}\underline{\tilde{T}}^{\mp}, \tag{5.11}$$

where E_b/E_k are the energies of the $\underline{H}_0$-eigen bras/kets of $\underline{A}$. This form is familiar from comparison with collision operators in quantum Boltzmann equations where typically the latter version is used (with some cancellation due to energy diagonality of density matrices).[22–26]

We now mimic the discussion of Snider and Sanctuary[23] to provide scattering equations for the above super-operators. For motivational and notational convenience we first present scattering equations for the channel space $\underline{T}$'s of the form

$$\underline{T}^{\pm} = \underline{V} + \underline{V}\underline{\mathscr{G}}_0^{\pm}(\underline{T}^{\pm}), \tag{5.12}$$

where the super-operator $\underline{\mathscr{G}}_0^{\pm}(\ )$ is defined by

$$\underline{\mathscr{G}}_0^{\pm}(\underline{\mathscr{A}}) = \int \underline{G}_0^{\pm}(E)\underline{A}\, d\underline{P}_0(E)$$
$$= \lim_{\epsilon\to 0+} \int\int \frac{1}{E-E' \pm i\epsilon}\, d\underline{P}_0(E')\underline{A}\, d\underline{P}_0(E)$$
$$= \lim_{\epsilon\to 0+} \int\int \frac{1}{E-E' \pm i\epsilon}\, d\underline{\mathscr{P}}_0(E',E)\underline{A}$$
$$= \lim_{\epsilon\to 0+} (\pm i\epsilon - \underline{\mathscr{L}}_0)^{-1}(\underline{A})$$
$$= \underline{\mathbf{G}}_0(\pm i\epsilon)(\underline{A}), \tag{5.13}$$

where $\{\underline{P}_0(E)\}$ is the spectral family of projectors for $\underline{H}_0$, $d\underline{\mathscr{P}}_0(E',E) = d\underline{P}_0(E') \otimes d\underline{P}_0(E)$, and $\underline{\mathbf{G}}_0(z) = (z - \underline{\mathscr{L}}_0)^{-1}$. Thus $\underline{G}_0^{\pm}(E)$ are the spectral components of $\underline{\mathscr{G}}_0^{\pm}$. Clearly

$$\underline{T}^{\pm} = \int \underline{T}^{\pm}(E) d\underline{P}_0(E). \tag{5.14}$$

Returning to the case of the abstract super-operators, $\underline{\mathscr{T}}^{\pm}$ satisfy

$$\underline{\mathscr{T}}^{\pm} = \underline{\mathscr{V}} + \underline{\mathscr{V}}\underline{\mathbf{G}}_0^{\pm}(\underline{\mathscr{T}}^{\pm}), \tag{5.15}$$

where the (super-) super-operator $\underline{\mathbf{G}}_0^{\pm}$ satisfies

$$\underline{\mathbf{G}}_0^{\pm}(\underline{\mathscr{A}}) = \lim_{\epsilon\to 0+} \int\int\int\int \frac{1}{\delta E - \delta E' \pm i\epsilon}\, d\underline{\mathscr{P}}_0(E'_k,E'_b) \times \underline{\mathscr{A}}\, d\underline{\mathscr{P}}_0(E_k,E_b)$$
$$= \lim_{\epsilon\to 0+} \int\int (\delta E - \underline{\mathscr{L}}_0 \pm i\epsilon)^{-1}\underline{\mathscr{A}}\, d\underline{\mathscr{P}}_0(E_k,E_b)$$
$$= \lim_{\epsilon\to 0+} \int\int \underline{\mathbf{G}}_0(\delta E \pm i\epsilon)\underline{\mathscr{A}}\, d\underline{\mathscr{P}}_0(E_k,E_b) \tag{5.16}$$

and $\delta E = E_k - E_b$, etc. Thus $\underline{\mathbf{G}}_0^{\pm}(\delta E)$ are the spectral components of $\underline{\mathbf{G}}_0^{\pm}$.

It is appropriate to introduce frequency (energy difference) dependent transition super-operators $\underline{\mathscr{T}}^{\pm}(w) = \lim_{\epsilon\to 0+} \underline{\mathscr{T}}(w \pm i\epsilon)$ where $\underline{\mathscr{T}}(z)$ satisfies

$$\underline{\mathscr{T}}(z) = \underline{\mathscr{V}} + \underline{\mathscr{V}}\underline{\mathbf{G}}_0(z)\underline{\mathscr{T}}(z) = \underline{\mathscr{V}} + \underline{\mathscr{T}}(z)\underline{\mathbf{G}}_0(z)\underline{\mathscr{V}}. \tag{5.17}$$

Thus

$$\underline{\mathscr{T}}^{\pm} = \int\int \underline{\mathscr{T}}^{\pm}(\delta E) d\underline{\mathscr{P}}_0(E_k,E_b), \tag{5.18}$$

so if $\underline{\mathscr{L}}_0\underline{A} = w\underline{A}$ then $\underline{\mathscr{T}}^{\pm}(\underline{A}) = \underline{\mathscr{T}}^{\pm}(w)\underline{A}$. Various other relationships can be obtained for the $\underline{\mathscr{T}}(z)$'s, $\underline{\Omega}(z)$'s, and appropriate Green functions by standard manipulations (the Hilbert space analogs of some of these are given in Ref. 23 and a more complete list presented in Appendix C). We remark that in the treatment of the scattering theory for energy diagonal density matrices, as with application to reactive Boltzmann equations, $\underline{\mathscr{T}}^{\pm}$ reduce to $\underline{\mathscr{T}}^{\pm} = \lim_{\epsilon\to 0+} \underline{\mathscr{T}}(\pm i\epsilon)$.

Under certain conditions the analysis of Fano[19] may be adapted to provide a contour integral representation for the transition super-operator $\underline{\mathscr{T}}(z)$. Firstly, we note that if $\underline{\mathscr{G}}_0(z) = (z - \underline{\mathscr{L}}_0)^{-1}$, $\underline{\mathscr{G}}_0^*(z) = (z - \underline{\mathscr{L}}_0^*)^{-1}$ then

$$\underline{\mathbf{G}}_0(z) = (z - \underline{\mathscr{L}}_0 + \underline{\mathscr{L}}_0^*)^{-1} = \underline{\mathscr{G}}_0(z + \underline{\mathscr{L}}_0^*)$$
$$= -\underline{\mathscr{G}}_0^*(\underline{\mathscr{L}}_0 - z). \tag{5.19}$$

If we define $\underline{\mathscr{L}}(z), \underline{\mathscr{L}}^*(z)$ as the solutions of

$$\underline{\mathscr{L}}(z) = \underline{\mathscr{V}} + \underline{\mathscr{V}}\underline{\mathscr{G}}_0(z)\underline{\mathscr{L}}(z) = \underline{\mathscr{V}} + \underline{\mathscr{L}}(z)\underline{\mathscr{G}}_0(z)\underline{\mathscr{V}}, \tag{5.20}$$

$$\underline{\mathscr{L}}^*(z) = \underline{\mathscr{V}}^* + \underline{\mathscr{L}}^*(z)\underline{\mathscr{G}}_0^*(z)\underline{\mathscr{V}}^* = \underline{\mathscr{V}}^* + \underline{\mathscr{V}}^*\underline{\mathscr{G}}_0^*(z)\underline{\mathscr{L}}^*(z), \tag{5.21}$$

then, from (5.19), it follows that

$$\underline{\mathscr{L}}(z + \underline{\mathscr{L}}_0^*) = \underline{\mathscr{V}} + \underline{\mathscr{V}}\underline{\mathbf{G}}_0(z)\underline{\mathscr{L}}(z + \underline{\mathscr{L}}_0^*) = \cdots, \tag{5.22}$$

$$\underline{\mathscr{L}}^*(\underline{\mathscr{L}}_0 - z) = \underline{\mathscr{V}}^* - \underline{\mathscr{L}}^*(\underline{\mathscr{L}}_0 - z)\underline{\mathbf{G}}_0(z)\underline{\mathscr{V}}^* = \cdots. \tag{5.23}$$

Furthermore, since $\underline{\mathscr{G}}_0(z) = \underline{G}_0(z) \otimes \underline{I}$, $\underline{\mathscr{G}}_0^*(z) = \underline{I} \otimes \underline{G}_0^*(z)$ we conclude also that (cf. Eu[20])

$$\underline{\mathscr{A}}(z + \underline{\mathscr{L}}_0^*) = \underline{T}(z + \underline{\mathscr{L}}_0^*) \otimes \underline{I}, \tag{5.24}$$

$$\underline{\mathscr{A}}^*(\underline{\mathscr{L}}_0 - z) = \underline{I} \otimes \underline{T}^*(\underline{\mathscr{L}}_0 - z). \tag{5.25}$$

Secondly, we make use of the fact that since $\underline{\mathscr{L}}_0 = \underline{\mathscr{L}}_0 - \underline{\mathscr{L}}_0^*$ represents a decomposition of $\underline{\mathscr{L}}_0$ into commuting parts, one may show

$$\underline{\mathbf{G}}_0(z) = \frac{1}{2\pi i}\int_{-\infty-i\eta}^{+\infty+i\eta} d\zeta\, \underline{\mathscr{G}}_0(\zeta)\underline{\mathscr{G}}_0^*(\zeta - z), \tag{5.26}$$

where $\eta \in (0, \mathrm{Im}\, z)$. The proof parallels that Hugenholtz[27] for operators with discrete spectra but uses an extension of the Dunford functional calculus for the unbounded self-adjoint operator $\underline{H}_0$.

At this stage we assume that $\underline{\mathscr{L}}(z), \underline{\mathscr{L}}^*(z)$ are analytic off the real z axis (which, in particular, requires that $\underline{H}$ have no complex eigenvalue spurious solutions). Then using the results listed above, the analysis of Fano[19] may be modified to show that

$$\underline{\mathscr{T}}(z) = \underline{\mathscr{L}}(z + \underline{\mathscr{L}}_0^*) - \underline{\mathscr{L}}^*(\underline{\mathscr{L}}_0 - z) + \frac{1}{2\pi i}\int_{-\infty+i\eta}^{+\infty+i\eta} d\zeta\, [\underline{\mathscr{G}}_0(\zeta) - \underline{\mathscr{G}}_0^*(\zeta - z)]\underline{\mathscr{L}}(\zeta) \times \underline{\mathscr{L}}^*(\zeta - z)[\underline{\mathscr{G}}_0(\zeta) - \underline{\mathscr{G}}_0^*(\zeta - z)]. \tag{5.27}$$

The analysis of Eu[20] may be adapted to give an alternative derivation of (5.27) in which the requirements that $\sigma(\underline{H}) \in \mathbb{R}$ and

$$\frac{1}{2\pi i}\oint_C d\zeta\, (\zeta - \underline{H})^{-1} = \underline{0} \quad \text{if } C \cap \{\mathrm{Im}\, \zeta = 0\} = \varnothing$$
$$= \underline{1}$$
if counterclockwise C "encloses" $\{\mathrm{Im}\, \zeta = 0\}$ (5.28)

are clear.

A somewhat different algebraic expression for $\underline{\mathscr{T}}(z)$ may also be obtained from an adaption of Eu's work as follows.[20] Define

$$\underline{\mathscr{L}}_1(z) = \underline{\mathscr{V}} + \underline{\mathscr{V}}\underline{\mathbf{G}}_0(z)\underline{\mathscr{T}}(z), \tag{5.29}$$

$$\underline{\mathscr{L}}_2(z) = \underline{\mathscr{V}}^* + \underline{\mathscr{V}}^*\underline{\mathbf{G}}_0(z)\underline{\mathscr{T}}(z), \tag{5.30}$$

so that

$$\underline{\mathscr{T}}(z) = \underline{\mathscr{L}}_1(z) - \underline{\mathscr{L}}_2(z). \tag{5.31}$$

Substituting (5.31) into (5.29) and (5.30) and using (5.22) and (5.23) to achieve appropriate rearrangements, one obtains

$$\begin{pmatrix} \underline{\mathscr{L}}_1(z) \\ \underline{\mathscr{L}}_2(z) \end{pmatrix} = \begin{pmatrix} \underline{\mathscr{L}}(z + \underline{\mathscr{L}}_0^*) \\ \underline{\mathscr{L}}^*(\underline{\mathscr{L}}_0 - z) \end{pmatrix} + \begin{pmatrix} \underline{0} & -\underline{\mathscr{L}}(z + \underline{\mathscr{L}}_0^*) \\ \underline{\mathscr{L}}^*(\underline{\mathscr{L}}_0 - z) & \underline{0} \end{pmatrix} \times \underline{\mathbf{G}}_0(z)\begin{pmatrix} \underline{\mathscr{L}}_1(z) \\ \underline{\mathscr{L}}_2(z) \end{pmatrix}. \tag{5.32}$$

Iterating (5.32) once decouples the equations for $\underline{\mathscr{L}}_1$ and $\underline{\mathscr{L}}_2$ leading to the formal solutions

$$\underline{\mathscr{L}}_1(z) = (\underline{1} + \underline{\mathscr{L}}(z + \not{4}_0^*)\underline{\mathbf{G}}_0(z)\underline{\mathscr{L}}^*(\not{4}_0 - z)\underline{\mathbf{G}}_0(z))^{-1} \times (\underline{\mathscr{L}}(z + \not{4}_0^*) - \underline{\mathscr{L}}(z + \not{4}_0^*)\, G_0(z)\mathscr{L}^*(\not{4}_0 - z))$$

$$\underline{\mathscr{L}}_2(z) = (\underline{1} + \underline{\mathscr{L}}^*(\not{4}_0 - z)\underline{\mathbf{G}}_0(z)\underline{\mathscr{L}}(z + \not{4}_0^*)\underline{\mathbf{G}}_0(z))^{-1} \times (\underline{\mathscr{L}}^*(\not{4}_0 - z) + \underline{\mathscr{L}}^*(\not{4}_0 - z)\underline{\mathbf{G}}_0(z)\underline{\mathscr{L}}(z + \not{4}_0^*)). \quad (5.33)$$

It is elucidating to compare (5.33) with (5.11) acting on an eigenket $\underline{\phi}_\alpha \underline{\xi}'_\beta$ of $\underline{\mathscr{L}}_0$. Retaining only terms quadratic in the $\underline{T}$'s in (5.33), one obtains

$$\begin{aligned}\underline{\mathscr{T}}^{\pm}(E_\alpha - E_\beta)\underline{\phi}_\alpha \underline{\xi}'_\beta &\approx \underline{T}^{\pm}(E_\alpha)\underline{\phi}_\alpha \underline{\xi}'_\beta - \underline{\phi}_\alpha \underline{\xi}'_\beta \underline{T}^{\mp}(E_\beta) \\ &+ \underline{T}^{\pm}(E_\alpha - E_\beta + E_b)\underline{\phi}_\alpha \underline{\xi}'_\beta \underline{T}^{\mp}(E_\beta)\underline{G}_0^{\mp}(E_\beta) \\ &- \underline{G}_0^{\pm}(E_\alpha)\underline{T}^{\pm}(E_\alpha)\underline{\phi}_\alpha \underline{\xi}'_\beta \underline{T}^{\mp}(E_k - E_\alpha + E_\beta), \quad (5.34)\end{aligned}$$

where E_b/E_k are the energies of the $\underline{H}_0$-eigen bras/kets of this operator. In contrast, (5.11) shows that

$$\begin{aligned}\underline{\mathscr{T}}^{\pm}\underline{\phi}_\alpha \underline{\xi}'_\beta &= \underline{\mathscr{T}}^{\pm}(E_\alpha - E_\beta)\underline{\phi}_\alpha \underline{\xi}'_\beta \\ &= \underline{T}^{\pm}(E_\alpha)\underline{\phi}_\alpha \underline{\xi}'_\beta - \underline{\phi}_\alpha \underline{\xi}'_\beta \underline{T}^{\mp}(E_\beta) \\ &+ \underline{T}^{\pm}(E_\alpha)\underline{\phi}_\alpha \underline{\xi}'_\beta \underline{T}^{\mp}(E_\beta)\underline{G}_0^{\mp}(E_\beta) \\ &- \underline{G}_0^{\pm}(E_\alpha)\underline{T}^{\pm}(E_\alpha)\underline{\phi}_\alpha \underline{\xi}'_\beta \underline{T}^{\mp}(E_\beta). \quad (5.35)\end{aligned}$$

Thus, except where $E_\alpha = E_k$, $E_\beta = E_b$, different energies appear in the quadratic terms of (5.34) and (5.35). The difference must be accounted for in the higher order terms of (5.33). Olmsted *et al.*[28] have further investigated this question (for the Hilbert space analog).

6. RESTRICTIONS, MODIFICATIONS, AND EXTENSIONS

First we consider a restriction of the analysis of the 3-particle problem in previous sections to the case where the breakup channel is strictly closed. Such systems considered here incorporate a true 3-body potential which guarantees that the total potential becomes unbounded in the breakup region (see Fig. 1). The asymptotic $|\phi_\alpha\rangle$ for each arrangement channel α include an infinite number of bound states. Furthermore, there are now no scattering solutions $|\psi_0^{\pm}\rangle$.

For a real system with tightly bound pairs, it may be possible to add a fictitious 3-body potential to strictly exclude breakup at all energies without significantly affecting the nonreactive and rearrangement collision dynamics of tightly bound reactants for a significant range of energies (of course, artificial higher energy bound states are introduced in all channels).

We expect there to be scattering solutions $\underline{\psi}_\alpha^{\pm}$ as before but the association with integral equations must be reexamined. For type 1B potentials, all infinite wall potentials must appear on diagonal [otherwise (1.1) requires off-diagonal $\underline{H}$ components with regions of $-\infty$ potential]. This allows construction of incoming spurious solutions with asymptotic parts in the "wrong" channels and allows spurious parts in outgoing parts of physical scattering solutions (thus destroying interpretation). Consequently, we confine our attention to type 1A potentials and BKLT choice of $\underline{H}$ where only appropriate channel potentials (infinitely deep wells) appear on diagonal. Application of the techniques of Appendix B demonstrates that the components of the scattering solution have the desired interpretational features (the appearance of channel Hamiltonians on diagonal in itself is *not* enough to guarantee the desired interpretational property as may be seen by analysis of the disconnected example of Appendix A). The $\underline{H}_0$-eigenvectors $\{\underline{\phi}_\alpha\}$ by themselves form a basis for $\mathscr{C}$ so $\underline{\Omega}^{\pm}$ are still defined on the whole space and thus, from (3.1) and (3.4), $\underline{S}^{\pm}$ are *unitary* (in the usual sense for operators on the Hilbert space $\mathscr{C}$).

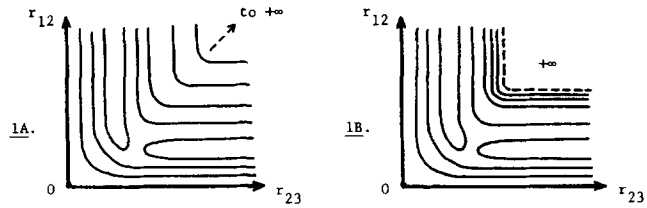
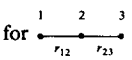


FIG. 1. Potential plots for  collinear configurations.

Consider next a three particle system where there are no (1 2) bound states. We examine a corresponding 2×2 channel space Hamiltonian $\underline{H}$, where the components are labeled by $\alpha,\beta, \cdots = (1)(2\ 3), (2)(1\ 3)$, having the decomposition $\underline{H} = \underline{H}_0 + \underline{V}$ with $(\underline{H}_0)_{\alpha\beta} = \delta_{\alpha\beta}H_\beta$. Forming the corresponding integral equation, the discussion of Secs. 2–5 may be readily modified to demonstrate the existence of scattering solutions $\underline{\psi}_\alpha^{\pm}$ with $\alpha = (1)(2\ 3), (2)(1\ 3)$, and $\underline{\psi}_{0(j)}^{\pm}$ with $j = 1,2$ corresponding to $|\psi_\alpha^{\pm}\rangle$ and $|\psi_0^{\pm}\rangle$, respectively. The corresponding inhomogeneous terms are given by the $\underline{H}_0$-eigenvectors $(\underline{\phi}_\alpha)_\beta = \delta_{\alpha\beta}|\phi_\alpha\rangle$ and $\underline{\phi}_{0(j)}^{\pm} \sim \underline{\theta}^j|\phi_0\rangle$ in the breakup region where $\underline{\theta}^j, j = 1,2$ are linearly independent. Although no choice of 2×2 $\underline{H}$ is connected, one might expect that provided $H_\beta \not\subset H_{\alpha\alpha}$ (i.e., $V_\beta \not\subset H_{\alpha\alpha}$) for $\alpha \neq \beta$, the partial interpretational property will be satisfied (or a full interpretational property if breakup and the physically unstable channel are excluded as above). This may be verified by application of the methods of Appendix B showing specificially that outgoing i cluster bound state contributions are contained only in the ith component ($i = 1,2$). For certain choices of $\underline{H}$ there may exist spurious wavelike solutions associated with unphysical Hamiltonians of the type discovered in Appendix A. These will not interfere with the interpretational property (see Appendix F).

Finally, we describe the extension of arrangement channel (space) scattering theory to the case of N distinguishable particles. The following characterization of scattering eigenvectors and duals analogous to Sec. 2 has been given by Evans.[5] Let $|\phi_{\alpha,\alpha'}^{\pm}\rangle$ denote the scattering eigenvector of H_α with asymptotic clustering α' finer than α (if $\alpha = \alpha'$ drop the $\pm$). Set $[\underline{\phi}_{\alpha,\alpha'}^{\pm}]_\beta = \delta_{\alpha\beta}|\phi_{\alpha,\alpha'}^{\pm}\rangle$. If $\underline{H} = \underline{H}_0 + \underline{V}$ where $(\underline{H}_0)_{\alpha\beta} = \delta_{\alpha\beta}H_\alpha$ (where α,β range over a subset of arrangement channels of interest), then the scattering eigenvectors $\underline{\psi}_{\alpha,\alpha'}^{\pm}$ corresponding to $|\psi_\alpha^{\pm}\rangle$ (which generalize $\underline{\psi}_\alpha^{\pm}, \underline{\psi}_{j,0}^{\pm}$ of the 3-particle case) satisfy

$$\underline{\psi}_{\alpha,\alpha'}^{\pm} = \underline{\phi}_{\alpha,\alpha'}^{\pm} + \underline{G}_0^{\pm}(E)\underline{V}\underline{\psi}_{\alpha,\alpha'}^{\pm}. \quad (6.1)$$

A discussion of the "canonical" choice as well as the corre-

sponding dual scattering eigenvectors is given in Ref. 5. The analysis of Appendix A extends to prove biorthogonality of these scattering eigenvectors and duals. Since $\underline{\phi}^{\pm}_{\alpha,\alpha'}$ form a complete set of $\underline{H}_0$-eigenvectors on the channel space $\mathscr{C}$, the Möller operators $\underline{\Omega}^{\pm}, \underline{\tilde{\Omega}}^{\pm}$ associated with these scattering solutions may be defined so that $\mathrm{dom}(\underline{\Omega}^{\pm}) = \mathscr{C}$, range $(\underline{\tilde{\Omega}}^{\pm}) = \mathscr{C}$. A development parallel to that of Secs. 2–5 is readily obtained.

Connected choices of $\underline{H}$ again have special features. Suppose $\alpha,\beta, \cdots$ include all 2-cluster channels and possibly some others, then the corresponding BKLT choice is connected. Thus any nonuniqueness in (6.1) corresponds to spatially confined $\underline{H}$-eigenvectors. Furthermore, the techniques of Appendix B may be applied to show that a partial interpretational property again holds. Specifically, the 2-cluster parts of the Hilbert space scattering wavefunction are contained only in the appropriate channel components. More generally, any channel component contains parts of the scattering wavefunction corresponding to that and finer clusterings [cf., Ref. 29]. For the case where only 2-cluster channels are retained, Kouri *et al.*[3] demonstrate the agreement between corresponding channel and Hilbert space scattering matrix elements by manipulating scattering equations and using Lippmann's identity (rather than via the more succinct wavefunction approach described previously).

Consider the case where the N particle system is confined to 2-cluster channels (a potential generalizing that of type 1A excludes breakup). Then a connected BKLT choice of $\underline{H}$ guarantees the full interpretational property for the scattering eigenvector components. Furthermore, the corresponding $\underline{S}^{\pm}$ are *unitary* on $\mathscr{C}$. Both these features are crucial in our development, from a channel space perspective, of a kinetic theory for a dilute reactive gaseous system (where breakup and recombination are excluded).[26] In the latter we also consider an N-particle, m-cluster system where breakup and recombination are excluded by inclusion of a suitable potential. Again a BKLT choice with components labeled by m-cluster channels will guarantee the desired interpretational properties. Furthermore, these properties are preserved for suitable energies if a spatially confining potential is added (of course, here, if the energy is too high, then the assignment of clustering inside the container becomes fuzzy). These results follow from the type of analysis described in Appendix B.

For comparison with the 3-particle discussion, we finally consider the general N-particle problem where one or more of the 2-cluster channels do not correspond to a stable molecule. If the connected $\underline{H}$ includes all 2-cluster channels except some of these, then the partial interpretational property for the 2-cluster parts of the scattering wavefunction is satisfied. The scattering equations may, however, exhibit homogeneous spurious wavelike solutions as seen in Appendix F. If no 2-cluster channels are stable, then a BKLT choice of $\underline{H}$ with components labeled by stable 3-cluster (and possibly finer) channels will exhibit a partial interpretational property with respect to the 3-cluster part of the scattering wavefuntion. There are further obvious extensions of these results.

APPENDIX A

We shall use the notation $i = (i)(j\,k)$, $i = 1,2,3$ where $\{i,j,k\} = \{1,2,3\}$ and $V_i = V_{jk}$ for the potential internal to channel i so $H_i = T + V_i$, where T is the kinetic energy. Further, $H = H_i + V^i$ for all i where $V^i = V_j + V_k$ assuming there are no true 3-body forces. We consider the channel space Hamiltonian

$$\underline{H} = \begin{pmatrix} H_1 & V^2 & V^3 \\ 0 & H_2 & 0 \\ V^1 & 0 & H_3 \end{pmatrix}. \tag{A1}$$

This does not correspond to a connected choice of $\underline{V}$ since

$$(\underline{G}_0\underline{V})^{2n} = \begin{pmatrix} (G_1V^3G_3V^1)^n & 0 & 0 \\ 0 & 0 & 0 \\ 0 & (G_3V^1G_1V^3)^{n-1}(G_3V^1G_1V^2) & (G_3V^1G_1V^3)^n \end{pmatrix}, \quad n \geqslant 1,$$

$$(\underline{G}_0\underline{V})^{2n+1} = \begin{pmatrix} 0 & (G_1V^3G_3V^1)^n(G_1V^2) & (G_1V^3)(G_3V^1G_1V^3)^n \\ 0 & 0 & 0 \\ (G_3V^1)(G_1V^3G_3V^1)^n & 0 & 0 \end{pmatrix}, \quad n \geqslant 0, \tag{A2}$$

and the four corner elements each contains terms of the form $(G_0V_2)^m$ (after expansion of G_1,G_3).

In Sec. 2, we discuss the solutions of the inhomogeneous scattering equation (1.5) associated with the decomposition $\underline{H} = \underline{H}_0 + \underline{V}$. To obtain other scattering solutions of $\underline{H}$ in (A1) which correspond to homogeneous solutions of (1.5), we make the decomposition $\underline{H} = \underline{H}_2 + \underline{V}^2$, where

$$\underline{H}_2 = \begin{pmatrix} T & 0 & V_2 \\ 0 & H_2 & 0 \\ V_2 & 0 & T \end{pmatrix}, \quad \underline{V}^2 = \begin{pmatrix} V_1 & V_1+V_3 & V_1 \\ 0 & 0 & 0 \\ V_3 & 0 & V_3 \end{pmatrix}. \tag{A3}$$

A complete set of $\underline{H}_2$ eigenvectors is given by

$$\left\{ \begin{pmatrix} 0 \\ |\phi\rangle \\ 0 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} |\phi\rangle \\ 0 \\ |\phi\rangle \end{pmatrix}, \begin{pmatrix} |\hat{\phi}\rangle \\ 0 \\ -|\phi\rangle \end{pmatrix} \right\}, \tag{A4}$$

where $\{|\phi\rangle\} = \{|\phi_2\rangle, |\phi_2^{\pm}\rangle\}(\{|\hat{\phi}\rangle\} = \{|\hat{\phi}_2\rangle|\hat{\phi}_2^{\pm}\rangle\})$ constitute a complete set of H_2 ($H_2 = T - V_2$) partially bound and scattering eigenvectors. These may be taken as the asymptotic parts $\underline{\phi}^{(\pm)}$ of $\underline{H}$ scattering eigenvectors $\underline{\psi}^{\pm}$ which satisfy

$$\underline{\psi}^{\pm} = \underline{\phi}^{(\pm)} + \underline{G}_2^{\pm}(E)\underline{V}^2\underline{\psi}^{\pm}. \tag{A5}$$

Choosing $|\phi\rangle = |\phi_2^{\pm}\rangle, |\hat{\phi}\rangle = |\hat{\phi}_2^{\pm}\rangle$ in (A4) generates

from (A5) the scattering solutions

$$\underline{\psi}^{\pm}_{2,0}, \quad \tfrac{1}{2}(\underline{\psi}^{\pm}_{1,0}+\underline{\psi}^{\pm}_{3,0}),(\underline{\psi}^{\pm}_{1,0}-\underline{\psi}^{\pm}_{3,0}) = \begin{pmatrix} |\phi_2^{\pm}\rangle \\ 0 \\ -|\hat{\phi}_2^{\pm}\rangle \end{pmatrix},$$

respectively. If there are any $|\phi_2\rangle$, then choosing

$$\underline{\phi} = \begin{pmatrix} 0 \\ |\phi_2\rangle \\ 0 \end{pmatrix}$$

in (A5) generates $\underline{\psi}^{\pm}_{(2)(1\,3)}$ discussed previously, and choosing

$$\underline{\phi} = \frac{1}{2}\begin{pmatrix} |\phi_2\rangle \\ 0 \\ |\phi_2\rangle \end{pmatrix}$$

generates another physical scattering solution corresponding to $|\psi_2^{\pm}\rangle$ which is a homogeneous solution of (1.5). We can, of course, keep one of these as the physical representative and replace the second by a difference (generating a new spurious wavelike $\underline{H}$-eigensolution). If any $|\hat{\phi}_2\rangle$ exist, then choosing

$$\underline{\phi} = \begin{pmatrix} |\hat{\phi}_2\rangle \\ 0 \\ -|\hat{\phi}_2\rangle \end{pmatrix}$$

generates another new spurious wavelike $\underline{H}$-eigensolution

$$\underline{\psi}^{s} = \begin{pmatrix} |\hat{\phi}_2\rangle \\ 0 \\ -|\hat{\phi}_2\rangle \end{pmatrix}$$

which is a homogeneous solution of (1.5). The discussion of the spectral theory and functional calculus for ACQM Hamiltonians of Ref. 5 must be slightly modified for this case. In particular, if $\underline{H}$ in (A1) is scalar spectral, then the newly discovered wavelike eigensolutions must be included in the basis.

APPENDIX B

We investigate here the componentwise interpretational features of the scattering eigenvectors $\underline{\psi}^{\pm}$ of $\underline{H}$ first for a connected choice of $\underline{V}$. Making a formal expansion of the solution of $\underline{\psi}^{\pm} = \underline{\phi}^{\pm} + \underline{G}_0^{\pm}(E)\underline{V}\underline{\psi}^{\pm}$ yields

$$|\psi_j^{\pm}\rangle = |\phi_j^{\pm}\rangle + G_j^{\pm}(E)\left[\underline{V}\sum_{n=0}^{\infty}(\underline{G}_0^{\pm}(E)\underline{V})^n\underline{\phi}\right]_j \tag{B1}$$

using the component labeling notation of Appendix A. The appearance of $G_j^{\pm}(E)$ allows $|\psi_j^{\pm}\rangle$ to support channel j clustering [i.e., $(i\,k)$ bound states where $\{i\,j\,k\} = \{1\,2\,3\}$]. The key feature of the connectivity assumption is that it guarantees we cannot resum the second term in (B1) to obtain a contribution of the form $G_k^{\pm}(E)\cdots$ where $k \neq j$. Here we assume this means that $|\psi_j^{\pm}\rangle$ cannot support channel k clustering.

Another approach uses the differential form of the channel space Schrödinger equation. We simply look at the form of $\underline{H}$ in each asymptotic tube and check which components of $\underline{\psi}$ can support the corresponding bound states, e.g., in the 2-tube

$$\underline{H}_B = \begin{pmatrix} H_1 & V^2 & 0 \\ 0 & H_2 & V^3 \\ V^1 & 0 & H_3 \end{pmatrix} \sim \begin{pmatrix} T & 0 & 0 \\ 0 & H_2 & V_2 \\ V_2 & 0 & T \end{pmatrix}. \tag{B2}$$

Thus $(\underline{H}_B - E)\underline{\psi} = 0$ in the 2-tube becomes

$$\begin{aligned} &T|\Psi_1\rangle = E|\psi_1\rangle, \\ &T|\psi_3\rangle + V_2|\psi_1\rangle = E|\psi_3\rangle, \\ &H_2|\psi_2\rangle + V_2|\psi_3\rangle = E|\psi_2\rangle, \end{aligned} \tag{B3}$$

so $|\psi_1\rangle$ and thus $|\psi_3\rangle$ cannot support channel 2-clustering (unlike $|\psi_2\rangle$). Both these techniques carry over to the N-particle case.

Application of both approaches to the example of Appendix A suggests how the interpretational property breaks down. If

$$\underline{\phi}^{\pm} = \begin{pmatrix} 0 \\ |\phi_2\rangle \\ 0 \end{pmatrix}$$

then

$$\underline{\psi}^{\pm} = \begin{pmatrix} \sum_{n>0}(G_1V^3G_3V^1)^n(G_1V^2)|\phi_2\rangle \\ |\phi_2\rangle \\ \sum_{n>1}(G_3V^1G_1V^3)^{n-1}G_3V^1G_1V^2|\phi_2\rangle \end{pmatrix}, \tag{B4}$$

demonstrating that any post collision 2-channel bound pair part of the wavefunction is distributed between $|\psi_1^{\pm}\rangle$ and $|\psi_3^{\pm}\rangle$. Accordingly, we can resum the expressions for these components to obtain a contribution $G_2^{\pm}\cdots$. This behavior is obvious from the differential form of the $\underline{H}$ eigenvector equation and the structure of $\underline{H}_2$.

APPENDIX C

A proof of the biorthogonality of the scattering eigenvectors and duals of $\underline{H}$ is given here using biorthogonality of the corresponding inhomogeneous terms in their integral equations (eigenvectors and duals of $\underline{H}_0$). As mentioned in the text, this property is directly verifiable where the dual eigenvector has equal components. However, a proof covering all cases is given here. Accordingly, we implement a general notation where $\underline{\psi}_J^{\pm}, \underline{\zeta}_J^{\pm\prime}$ denote any scattering eigenvector and corresponding dual of $\underline{H}$ associated with energy eigenvalue E_J. The corresponding inhomogeneous terms are denoted by $\underline{\phi}_J^{\pm}, \underline{\xi}_J^{\pm\prime}$ (the $\pm$ here can be dropped for $\underline{\psi}_\alpha^{\pm}, \underline{\zeta}_\alpha^{\pm\prime}$).

Our aim is thus to prove that

$$(\underline{\zeta}_J^{\pm\prime}, \underline{\psi}_K^{\pm}) = \delta_{J,K} \tag{C1}$$

which should be interpreted as

$$(\underline{\zeta}_J^{\pm\prime}, \int dK\, a_K \underline{\psi}_K^{\pm}) = a_J \tag{C2}$$

for suitable a_K. Here $\int dK$ denotes sum/integration over the state labels and $\delta_{J,K}$ is the corresponding Kronecker/Dirac delta function. The following restricted spectral representation result for $\underline{H}$, derived from the Plemelj relations, is required here:

$$\underline{G}^{\pm}(E)\int dK\, a_K \underline{\psi}_K^{+(-)}$$
$$= \int dK\, a_K \left(\mp i\pi\delta(E-E_K) + \frac{\mathscr{P}}{E-E_K}\right)\underline{\psi}_K^{+(-)}. \quad \text{(C3)}$$

Here $\mathscr{P}$ represents the Cauchy principal value integral. A corresponding representation for $\underline{G}_0^{\pm}(E)$ comes from the standard theory since $\underline{H}_0$ is self-adjoint.

As a preliminary to proving (C2), we consider

$$(\underline{\xi}_K^{\pm\prime}, \underline{V}\underline{G}^{\mp}(E_K)\int dJ\, a_J \underline{\psi}_J^{\pm})$$
$$= \int dJ\, a_J \left\{ \pm i\pi(\underline{\xi}_K^{\pm\prime}, \underline{V}\underline{\psi}_J^{\pm})\delta(E_K - E_J) + \left(\underline{\xi}_K^{\pm\prime}, \underline{V}\frac{\mathscr{P}}{(E_K-E_J)}\underline{\psi}_J^{\pm}\right)\right\}$$
$$= -\int dJ\, a_J (\underline{\xi}_K^{\pm\prime}, \underline{G}_0^{\pm}(E_J)\underline{V}\underline{\psi}_J^{\pm}). \quad \text{(C4)}$$

Thus finally,

$$(\underline{\xi}_K^{\pm\prime}, \int dJ\, a_J \underline{\psi}_J^{\pm})$$
$$= (\underline{\xi}_K^{\pm\prime}, \{\underline{1} + \underline{V}\underline{G}^{\mp}(E_K)\}\int dJ\, a_J \underline{\psi}_J^{\pm})$$
$$= \int dJ\, a_J (\underline{\xi}_K^{\pm\prime}, \{\underline{1} - \underline{G}_0^{\pm}(E_J)\underline{V}\}\underline{\psi}_J^{\pm})$$
$$= \int dJ\, a_J (\underline{\xi}_K^{\pm\prime}, \underline{\phi}_J^{\pm})$$
$$= a_K \quad \text{(C5)}$$

as required. We have used (2.6), (C4), (2.1), and (2.2).

APPENDIX D

We first give an example demonstrating how breakdown of the componentwise interpretational property implies that

$$\underline{\Omega}^{-1}(t) \not\to \underline{\tilde{\Omega}}^{\pm}. \quad \text{(D1)}$$

We take the Hamiltonian of Appendix A and suppose that (1 3) bound states exist. The outgoing 2-channel bound state components in $\underline{\psi}^{+}_{(2)(1\,3)}$ were shown to be distributed between the 1 and 3 components. Consequently, if $\underline{\xi}_2'(t) = (0, \langle\phi_2(0)|e^{i/\hbar H_2 t}, 0)$ then

$$0 \sim (\underline{\xi}_2'(t), \underline{\psi}^{+}_{(2)(1\,3)}(t))$$
$$= (\underline{\xi}_2'(0), \underline{\Omega}^{-1}(t)\underline{\psi}^{+}_{(2)(1\,3)}(0))$$
$$\not\to (\underline{\xi}_2'(0), \underline{\tilde{\Omega}}^{-}\underline{\psi}^{+}_{(2)(1\,3)}(0))$$
$$= S^{+}_{2,2} \quad \text{as } t\to\infty. \quad \text{(D2)}$$

Specifically, one can show in this case that

$$w\text{-}\lim_{t\to+\infty}\underline{\Omega}^{-1}(t) = (\underline{1} - \sum_2 \underline{\phi}_2\underline{\xi}_2')\underline{\tilde{\Omega}}^{-}. \quad \text{(D3)}$$

For general 3-particle, 3-channel $\underline{H}$, if any outgoing i channel bound state flux is missing from the i component of $\underline{\psi}^{+}_{(j)(i\,k)}(0)$, then $\lim_{t\to+\infty}(\underline{\xi}_i(t), \underline{\psi}^{+}_{(j)(i\,k)}(t))$ will be less than $S^{+}_{i,j}$.

APPENDIX E

There are many relations satisfied by the super-operators of Sec. 5 analogous to standard scattering theory identities. Some of these are presented here. Suppose that

$$\underline{H}_0\underline{A} = E\underline{A}, \quad \underline{A}\underline{H}_0 = E'\underline{A}, \quad \text{(E1)}$$

so that $\underline{\mathscr{L}}_0\underline{A} = (E-E')\underline{A}$. Then clearly

$$\underline{\Omega}^{\pm}\underline{A} = \underline{\Omega}^{\pm}(E,E')\underline{A} = \underline{Q}^{\pm}(E)\underline{A}\underline{\tilde{Q}}^{\pm}(E'), \quad \text{(E2)}$$

so

$$\underline{\Omega} = \int\int \underline{\Omega}^{\pm}(E,E')d\underline{\mathscr{P}}_0(E,E'). \quad \text{(E3)}$$

On the other hand if $\underline{A}$ satisfies

$$\underline{\mathscr{L}}_0\underline{A} = w\underline{A} \quad \text{(E4)}$$

[which does not imply separate relationships analogous to (E1) are satisfied] then standard Lippmann–Schwinger type manipulations give

$$\underline{\Omega}^{\pm}\underline{A} = \underline{\Omega}^{\pm}(w)\underline{A}, \quad \text{(E5)}$$

where $\underline{\Omega}^{\pm}(w) = \lim_{\epsilon\to 0+}\underline{\Omega}^{\pm}(w\pm i\epsilon)$ and $\underline{\Omega}(z)$ satisfies

$$\underline{\Omega}(z) = \underline{1} + \underline{\mathbf{G}}(z)\underline{\mathscr{V}} = \underline{1} + \underline{\mathbf{G}}_0(z)\underline{\mathscr{V}}\underline{\Omega}(z), \quad \text{(E6)}$$

where $\underline{\mathbf{G}}(z) = (z-\underline{\mathscr{L}})^{-1}$ so

$$\underline{\mathbf{G}}(z) = \underline{\mathbf{G}}_0(z) + \underline{\mathbf{G}}_0(z)\underline{\mathscr{V}}\underline{\mathbf{G}}(z) = \underline{\mathbf{G}}_0(z) + \underline{\mathbf{G}}(z)\underline{\mathscr{V}}\underline{\mathbf{G}}_0(z). \quad \text{(E7)}$$

Thus

$$\underline{\Omega} = \int\int \underline{\Omega}^{\pm}(E-E')d\underline{\mathscr{P}}_0(E,E'). \quad \text{(E8)}$$

The frequency dependent transition super-operator of (5.17) may also be defined by

$$\underline{\mathscr{T}}(z) = \underline{\mathscr{V}}\underline{\Omega}(z), \quad \text{(E9)}$$

so $\underline{\Omega}(z) = \underline{1} + \underline{\mathbf{G}}_0(z)\underline{\mathscr{T}}(z)$ and $\underline{\mathscr{V}}\underline{\mathbf{G}}(z) = \underline{\mathscr{T}}(z)\underline{\mathbf{G}}_0(z)$. Besides (5.17), $\underline{\mathscr{T}}(z)$ also satisfies

$$\underline{\mathscr{T}}(z) = \underline{\mathscr{V}} + \underline{\mathscr{V}}\underline{\mathbf{G}}(z)\underline{\mathscr{V}}. \quad \text{(E10)}$$

APPENDIX F

For the 3-particle system where no (1 2) bound states exist, we consider the following 2×2 choices of $\underline{H}$. Firstly, if

$$\underline{H} = \begin{pmatrix} H_3 + V_1 & V_1 \\ V_2 & H_3 + V_2 \end{pmatrix} \quad \text{(F1)}$$

and $G_3(z) = (z-H_3)^{-1}$ then

$$|\psi(n)_k\rangle = G_3(E_n)V_k|\psi(n)\rangle, \quad k=1,2,$$
$$|\psi^{\pm}_{\alpha k}\rangle = G_3^{\pm}(E)V_k|\psi^{\pm}_{\alpha}\rangle, \quad k=1,2, \ \alpha=(1)(2\,3),(2)(1\,3). \quad \text{(F2)}$$

Further choosing $\underline{\theta}^{j} = \underline{e}^{j}$ so $\underline{\psi}^{\pm}_{0(j)} = \underline{\psi}_{j,0}$ (Ref. 5) and setting $(E-H_3)|\phi_3^{\pm}\rangle = 0$, $t^{\pm}_{k/3} = V_k + V_k G_3^{\pm} t^{\pm}_{k/3}$, one obtains

$$[\underline{\psi}^{\pm}_{j,0}]_k = \delta_{k,j}|\phi_3^{\pm}\rangle + G_3^{\pm}t^{\pm}_{k/3}|\phi_3^{\pm}\rangle + G_3^{\pm}t^{\pm}_{k/3}G_3^{\pm}V_k|\psi_0^{\pm}\rangle \quad \text{(F3)}$$

where $\{k,\hat{k}\} = \{1,2\}$.

Secondly, if

$$\underline{H} = \begin{pmatrix} T+V_1+V_3 & V_1+V_3 \\ V_2 & T+V_2 \end{pmatrix} \quad \text{(F4)}$$

then

$$|\psi(n)_k\rangle = G_0(E_n)(V_k + \delta_{k,1} V_3)|\psi(n)\rangle, \quad k = 1,2,$$
$$|\psi^{\pm}_{\alpha k}\rangle = G_0^{\pm}(E)(V_k + \delta_{k,1} V_3)|\psi^{\pm}_{\alpha}\rangle, \quad k = 1,2, \quad \alpha = (1)(2\,3),(2)(1\,3), \tag{F5}$$

and

$$\underline{\psi}^{\pm}_{j,0} = \underline{e}^{j}|\phi_0\rangle + G_0^{\pm}\begin{pmatrix} t^{\pm}_{1+3} \\ t^{\pm}_{2} \end{pmatrix}|\phi_0\rangle + G_0^{\pm}\begin{pmatrix} t^{\pm}_{1+3} G_0^{\pm} V_2 \\ t^{\pm}_{2} G_0^{\pm}(V_1 + V_3) \end{pmatrix}|\psi_0^{\pm}\rangle. \tag{F6}$$

From (F2) and (F5) we see that for a canonical choice of $\underline{\theta}^j$, the spurious wavelike eigenvectors span $\zeta = \{\underline{\psi} : \Sigma_{i=1}^{2} |\psi_i\rangle = 0\}$. Thus, with the usual technical assumptions, both these Faddeev-like $\underline{H}$ are scalar spectral. Furthermore, the partial interpretational property is satisfied for both and the scattering equations with kernel $G_0^{\pm}\underline{V}$ have unique solutions.

Finally, we consider the BKLT-type choice

$$\underline{H} = \begin{pmatrix} H_1 & V_1 + V_3 \\ V_2 + V_3 & H_2 \end{pmatrix}. \tag{F7}$$

In addition to scattering solutions analogous to those described above, others may exist which are associated with the decomposition

$$\underline{H} = \underline{H}_3 + \underline{V}^3$$

where

$$\underline{H}_3 = \begin{pmatrix} T & V_3 \\ V_3 & T \end{pmatrix}, \quad \underline{V}^3 = \begin{pmatrix} V_1 & V_1 \\ V_2 & V_2 \end{pmatrix}. \tag{F8}$$

A complete set of $\underline{H}_3$ eigenvectors is given by

$$\left\{\frac{1}{2}\begin{pmatrix} |\phi\rangle \\ |\phi\rangle \end{pmatrix}, \begin{pmatrix} |\hat{\phi}\rangle \\ -|\hat{\phi}\rangle \end{pmatrix}\right\}, \tag{F9}$$

where $\{|\phi\rangle\} = \{|\phi^{\pm}_{3}\rangle\}(\{|\hat{\phi}\rangle\} = \{|\hat{\phi}_3\rangle, |\hat{\phi}^{\pm}_{3}\rangle\})$ constitute complete sets of H_3 ($\hat{H}_3 = T - V_3$) scattering and partially bound eigenvectors. These may be taken as the asymptotic part $\underline{\phi}^{(\pm)}$ of $\underline{H}$ scattering eigenvectors which satisfy

$$\underline{\psi}^{\pm} = \underline{\phi}^{(\pm)} + \underline{G}_3^{\pm}(E)\underline{V}^3\underline{\psi}^{\pm}. \tag{F10}$$

This generates scattering solutions

$$\frac{1}{2}(\underline{\psi}^{\pm}_{1,0} + \underline{\psi}^{\pm}_{2,0}), \quad \underline{\psi}^{\pm} = \begin{pmatrix} |\hat{\phi}_3\rangle \\ -|\hat{\phi}_3\rangle \end{pmatrix},$$
$$\underline{\psi}^{\pm} = \underline{\psi}^{\pm}_{1,0} - \underline{\psi}^{\pm}_{2,0} = \begin{pmatrix} |\hat{\phi}^{\pm}_{3}\rangle \\ -|\hat{\phi}^{\pm}_{3}\rangle \end{pmatrix},$$

respectively. Clearly

$$\underline{\psi}^{\pm} = \begin{pmatrix} |\hat{\phi}_3\rangle \\ -|\hat{\phi}_3\rangle \end{pmatrix}$$

are homogeneous solutions of (1.5). The partial interpretational property is not destroyed by the existence of such solutions since they do not mix with the outgoing parts of other scattering solutions. This is verified by taking a Neumann expansion for the latter and observing that it is not possible to resum any subset of terms to obtain a contribution $\hat{G}^{\pm}_{3} \cdots$ where $\hat{G}_3(z) = (z - \hat{H}_3)^{-1}$. The structure of this Hamiltonian is discussed in more detail in Ref. 30.

Remark: Suppose now that (1 2) bound states do exist and consider the $\underline{H}$-eigenvectors $\underline{\psi}^{\pm}_{(3)(1\,2)}$ for the above choices of $\underline{H}$. For (F1) we may choose the position of the asymptotic part in either (of the 2) components, for (F4) only in component 1, and for (F7) only distributed equally in both components.

[1] S. Weinberg, Phys. Rev. **B 133**, 232 (1964).
[2] K. L. Kowalski, in *Few Body Systems and Nuclear Forces, Lecture Notes in Physics*, Vol. 87 (Springer, New York, 1978).
[3] D. J. Kouri, H. Krüger, and F. S. Levin, Phys. Rev. D **15**, 1156 (1977). The ACQM approach resulted from the original work of Baer and Kouri, and later work by Kouri and Levin, and Tobocman (see Ref. 14).
[4] D. K. Hoffman, D. J. Kouri, and Z. H. Top, J. Chem. Phys. **70**, 4640 (1979).
[5] J. W. Evans, J. Math. Phys. **22**, 1672 (1981).
[6] J. W. Evans and D. K. Hoffman, J. Math. Phys. **22**, 2858 (1981).
[7] W. N. Polyzou and E. F. Redish, Ann. Phys. **119**, 1 (1979).
[8] W. N. Polyzou, Ph.D. Thesis, University of Maryland, 1979.
[9] M. Reed and B. Simon, *Methods of Modern Mathematical Physics III: Scattering Theory* (Academic, New York, 1979).
[10] W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, 1977).
[11] W. Glöckle, Nucl. Phys. A **141**, 620 (1970); **158**, 257 (1970); W. Tobocman, Phys. Rev. C **11**, 43 (1975).
[12] Page 3 of Ref. 9.
[13] J. M. Jauch and J.-P. Marchland, Helv. Phys. Acta **39**, 325 (1966).
[14] M. Baer and D. J. Kouri, Phys. Rev. A **4**, 1924 (1971); J. Chem. Phys. **56**, 1958 (1972) and J. Math. Phys. **14**, 1637 (1973); D. J. Kouri and F. S. Levin, Bull. Am. Phys. Soc. **19**, 489 (1974); Phys. Rev. A **10**, 1616 (1974); Phys. Lett. B **50**, 421 (1974); Nucl. Phys. A **250**, 127 (1975), **253**, 395 (1975); W. Tobocman, Phys. Rev. C **9**, 2466 (1974).
[15] L. D. Faddeev, *Mathematical Aspects of the Three Body Problem in the Quantum Scattering Theory* (Davey, New York, 1965).
[16] I. Singer, *Bases in Banach Spaces I* (Springer, New York, 1970).
[17] Appendix C of Ref. 6.
[18] U. Fano, in *Lectures on the Many-Body Problem*, edited by E. R. Cainiello (Academic, New York, 1964), Vol. 2, p. 217.
[19] U. Fano, Phys. Rev. **131**, 259 (1963).
[20] B. C. Eu, J. Chem. Phys. **63**, 298 (1975).
[21] J. M. Jauch, B. Misra, and A. G. Gibson, Helv. Phys. Acta **41**, 513 (1968).
[22] R. F. Snider, J. Chem. Phys. **32**, 1051 (1960).
[23] R. F. Snider and B. C. Sanctuary, J. Chem. Phys. **55**, 1555 (1971).
[24] J. T. Lowry and R. F. Snider, J. Chem. Phys. **61**, 2320 and 2330 (1974).
[25] B. C. Eu, J. Chem. Phys. **63**, 303 (1975).
[26] J. W. Evans, D. K. Hoffman, and D. J. Kouri, "The Reactive Quantum Boltzmann Equations from an Arrangement Channel Space Representation and BBGKY Hierarchy," preprint, Ames Laboratory and Department of Chemistry, Iowa State University, 1982.
[27] N. M. Hugenholtz, Physica **23**, 481 (1957).
[28] R. D. Olmsted, J. C. Rainwater, and R. F. Snider, J. Chem. Phys. **64**, 5100 (1976).
[29] V. Vanzani, in *Few Body Nuclear Physics* (International Atomic Energy Commission, Vienna, 1978).
[30] J. W. Evans, "Spectral Theory of Arrangement Channel Q. M. Hamiltonians: Perturbation and Spurious Multiplier Techniques," preprint, Ames Laboratory and Department of Chemistry, Iowa State University, 1982.

Partitioning "opposite" bounds for potential scattering phase shifts

M. A. Abdel-Raouf[a)]
Fakultät für Physik, Universität Freiburg, 7800 Freiburg, West Germany

(Received 3 August 1982; accepted for publication 5 November 1982)

It is shown that the partitioning technique (PT) provides us with upper (lower) bounds for the scattering phase shifts under the same conditions where Schwinger's variational method (SVM) leads to lower (upper) bounds.

PACS numbers: 03.65.Nk, 03.80. + r, 34.10. + x

The interest in using Schwinger's variational method (SVM) for treating electron–atom and electron–molecule scattering problems has increased tremendously during the last few years (see, e.g., Watson and McKoy,[1] Takatsuka *et al.*,[2] Nesbet,[3] and references therein). In particular, questions have been reconsidered about the quality of the calculated phase shifts in comparison with those obtained by other variational methods and the possibility of deriving upper and lower bounds. It is well known[4–6] that Schwinger's phase shifts δ_S are lower (upper) bounds for the exact ones δ_{ex} if the scattering potential is definite positive (negative) and $|\delta_{ex}| < \Pi$. Sugar and Blankenbecler[7] tried, on the one hand, to relax the last condition. On the other hand, they indicated that, while calculations of opposite bounds should be practically more difficult, they become accessible if one proceeds in analogous manner to the method of intermediate operators (see, e.g., Basley and Fox[8] and Abdel-Raouf[9]). Actually, additional formulations of opposite bounds were established many years ago by Hahn[10] on a similar basis to the Temple-like formulae which are frequently used in bound-state problems.[9,11–13]

In this note, we show that our representation of the partitioning technique (Abdel-Raouf,[14] and henceforth referred to as I) can be employed for obtaining opposite bounds. The advantage of the new approach is two fold:

(i) The difficulties associated with the Sugar–Blankenbecler procedure are avoided.

(ii) No quadratic matrix elements as those appearing in Temple-like formulae are required.

Following Biedenharn and Blatt,[15] the Lippmann–Schwinger equation for a potential scattering problem with the total Hamiltonian $H = H^{(0)} + H'$ and total energy E, i.e.,

$$|\psi\rangle = |\phi\rangle + G_0 H'|\psi\rangle, \tag{1}$$

can be rewritten as

$$\lambda H' q H'|\psi\rangle = (H' - H' G_0 H')|\psi\rangle, \tag{2}$$

where G_0 is the integral operator $(E - H^{(0)})^{-1}$ and $\lambda = \cot(\delta_{ex}) = 1/\langle\phi|H'|\psi\rangle$. The vector $|\phi\rangle$ involved in Eq. (1) satisfies the equation $(E - H^{(0)})|\phi\rangle = |0\rangle$ and q, Eq. (2), is the associated projection operator, i.e.,

$$q = |\phi\rangle\langle\phi|. \tag{3}$$

Consider now the two vectors $|\eta\rangle$ and $|\zeta\rangle$, such that

$$|\eta\rangle = H'^{1/2}|\psi\rangle \quad \text{and} \quad |\zeta\rangle = H'^{1/2}|\phi\rangle.$$

Therefore, Eq. (2) can be reduced into the compact form

$$H_1|\eta\rangle = \lambda H_2|\eta\rangle \tag{4}$$

with

$$H_1 = 1 - K \tag{5a}$$

and

$$H_2 = |\zeta\rangle\langle\zeta|, \tag{5b}$$

where K is an integral operator the kernel of which k is defined in terms of the formal Green function $g(r, r')$ by

$$k = H'^{1/2}(r) g(r, r') H'^{1/2}(r').$$

If the potential operator H' is chosen such that

$$\int\int \frac{H'(r)H'(r')}{|r - r'|^2}\, dr\, dr' < \infty \quad \text{and} \quad \int |H'(r)|\, dr < \infty, \tag{6}$$

then, by definition, the vectors $|\eta\rangle$ and $|\zeta\rangle$ belong to a separable Hilbert space χ, and, consequently, H_2 is a Hermitian positive-definite operator with a domain in χ. Also (6) implies that both K and H_1 are Hermitian Hilbert–Schmidt operators.

In the Bubnov–Galerkin variational treatment of Eq. (4), $|\eta\rangle$ is replaced by a trial expansion space $|\eta_t^{(n)}\rangle$, expressed by

$$|\eta_t^{(n)}\rangle = \sum_{i=1}^{n} a_i |\chi_i\rangle \tag{7}$$

and one demands that

$$\lambda^{(n)} = \langle\eta_t^{(n)}|H_1|\eta_t^{(n)}\rangle / \langle\eta_t^{(n)}|H_2|\eta_t^{(n)}\rangle = \text{minimum}. \tag{8}$$

(7) and (8) lead to the system of secular equations

$$\sum_{i}^{n} a_i (H_{ij}^{(1)} - \lambda^{(n)} H_{ij}^{(2)}) = 0, \; j = 1, 2, \ldots, n, \tag{9}$$

where $H_{ij}^{(1)} = \langle\chi_i|H_1|\chi_j\rangle$ and $H_{ij}^{(2)} = \langle\chi_i|H_2|\chi_j\rangle$.

The system (9) is meaningful if and only if

$$\det(H_{ij}^{(1)} - \lambda^{(n)} H_{ij}^{(2)}) = 0 \tag{10}$$

and the uniqueness of the resultant $\lambda^{(n)}$ is guaranteed by the fact that the matrix $H^{(2)}$ is of unit rank. Therefore, we have the strong inequality

$$\lambda \leqslant \lambda^{(n)} \quad \text{for all } n \geqslant 1. \tag{11}$$

The previously mentioned variational method coincides identically with SVM (Singh and Stauffer[6]), i.e., $\cot(\delta_S) = \lambda^{(n)}$ and

$$\cot(\delta_{ex}) \leqslant \cot(\delta_S), \tag{12}$$

or $\delta_S < \delta_{ex}$ as long as $\delta_{ex} < \Pi$. It is also obvious from the characteristics of the operators H_1 and H_2 that Kato's condition (Ref. 4, p. 298) for obtaining upper bounds from SVM is

a) Present address: Physics Department, York University, Toronto, Ontario, Canada M3J 7P3.

 0022-2488/83/030588-02$02.50

uniquely satisfied. On the other hand, these characteristics emphasize that Eq. (4) can be similarly treated as (I.2) with H_1 and H_2 defined by (5a) and (5b), respectively, and λ plays the role of E in I. Thus, the representations given in Sec. I.2–I.5 are valid for (4) and, consequently, partitioning lower bounds for λ become available. The technique (Sec. I.5) is as follows: Let $|\xi\rangle$ be a reference space in χ for which $\langle\xi|\xi\rangle = 1$. Defining the projection operator Q and its orthogonal complement P by $Q = |\xi\rangle\langle\xi|$, $P = 1 - Q$, where Q and P satisfy the relations $Q^+ = Q$, $Q^2 = Q$, $P^+ = P$, $P^2 = P$, and $PQ = QP = 0$, the generalized resolvent operator T associated with Eq. (4) is expressed in terms of Q and P as

$$T = P(\alpha\cdot Q + P(\epsilon H_2 - H_1)P)^{-1}P, \tag{13}$$

where α is a constant different than zero and $\partial T/\partial\alpha = 0$. The operators T, Q, and P satisfy Eqs. (I.64a, b). The first of these equations states that

$$P(\epsilon H_2 - H_1)T = P.$$

Splitting P into $1 - Q$, we obtain after rearrangement

$$(\epsilon H_2 - H_1)T = 1 - Q[1 - (\epsilon H_2 - H_1)T]. \tag{14}$$

Assuming that the reference space is composed of one vector, we may define a vector $|\eta_1\rangle$ such that

$$|\eta_1\rangle = |\xi\rangle + T(H_1 - \epsilon H_2)|\xi\rangle. \tag{15}$$

Operating on (15) from the left with $(\epsilon H_2 - H_1)$ and using Eq. (14), we find

$$(\epsilon H_2 - H_1)|\eta_1\rangle = b\,|\xi\rangle, \tag{16a}$$

where

$$b = \langle\xi|H_1 - \epsilon H_2 + (H_1 - \epsilon H_2)T(H_1 - \epsilon H_2)|\xi\rangle. \tag{16b}$$

Therefore, $|\eta_1\rangle$ is the exact solution $|\eta\rangle$ of Eq. (4) if and only if the functional

$$\epsilon_1 = f(\epsilon) = \frac{\langle\xi|H_1 + (H_1 - \epsilon H_2)T(H_1 - \epsilon H_2)|\xi\rangle}{\langle\xi|H_2|\xi\rangle} \tag{17}$$

coincides with ϵ, which is necessarily equal to λ.

An important step in the PT is to reduce the functional (17) into the form (I.69), i.e., to construct the reference space such that $QH_2P = PH_2Q = 0$ and, consequently $H_2T = TH_2 = 0$. This can be carried out by selecting $|\xi\rangle$ to be equal to $H'^{1/2}|\omega\rangle$ where $\omega\in\chi$, and demanding that the operators Q and H_2 commute with each other. The simplest compromise, however, is to choose $|\xi\rangle$ equivalent to the Hilbert space vector $|\zeta\rangle$ and require that

$$\langle\zeta|\zeta\rangle = 1. \tag{18}$$

Since Q and H_2 are now identical, therefore, (17) leads to the simple form

$$\epsilon_1 = f(\epsilon) = \langle\zeta|H_1 + H_1TH_1|\zeta\rangle. \tag{19}$$

In other words, the solution of Eq. (4) is reduced to the solution of the equation

$$F(\epsilon) = \epsilon - \epsilon_1 = 0, \tag{20}$$

which possesses, according to the argument mentioned after Eq. (10), one and only one root.

The derivatives of $f(\epsilon)$ and $F(\epsilon)$ are subjected to the following characteristics, respectively:

$$f'(\epsilon) = -\langle\zeta|H_1T^2H_1|\zeta\rangle < 0 \tag{21a}$$

and

$$F'(\epsilon) = 1 - f'(\epsilon) > 0. \tag{21b}$$

The essential point of the PT lies in the fact that if $\delta\lambda = \lambda - \epsilon$ and $\delta\lambda_1 = \lambda - \epsilon_1$ are the deviations of ϵ and ϵ_1, respectively, from λ [$= \cot(\delta_{ex})$], then, because of (21a), $\delta\lambda$ and $\delta\lambda_1$ must possess different signs, i.e., ϵ and ϵ_1 satisfy the "bracketing theory." Thus, choosing ϵ to be $\lambda^{(n)}$ [or $\cot(\delta_S)$] (which means that $\delta\lambda$ is negative), ϵ_1 becomes a lower bound for λ, i.e., we have the strong inequality

$$\lambda > \lambda_{PT} = \langle\zeta|H_1 + H_1P(\alpha\cdot Q + P[H_2\cot(\delta_S) - H_1]P)^{-1}PH_1|\zeta\rangle, \tag{22}$$

or $\delta_{PT} > \delta_{ex}$ as long as $\delta_{ex} < \Pi$.

Finally, we remark that with straightforward generalizations both inequalities given in (11) and (22) can be extended for delivering upper and lower bounds, respectively, on the elements of the inverse reactance matrix of the same type of scattering problems. Moreover, if the potential H' is negative-definite and relations (6) are valid, (11) and (22) yield lower and upper bounds, respectively.

[1] D. K. Watson and V. McKoy, Phys. Rev. A **20**, 1474 (1979).
[2] K. Takatsuka, R. R. Lucchase and V. McKoy, Phys. Rev. A **24**, 1812 (1981).
[3] R. K. Nesbet, *Variational Methods in Electron-Atom Scattering Theory* (Plenum, New York, 1980), §2.7.
[4] T. Kato, Progr. Theor. Phys. (Jpn.) **6**, 295 (1951).
[5] C. S. Sharma and I. Rebelo, Phys. Lett. **44**, 29 (1973).
[6] S. R. Singh and A. D. Stauffer, Nuovo Cimento **22**, 139 (1974).
[7] R. Sugar and R. Blankenbecler, Phys. Rev. **136**, B372 (1964).
[8] N. W. Basley and D. V. Fox, Phys. Rev. **124**, 483 (1961).
[9] M. A. Abdel-Raouf, Phys. Rep. **84**, 163 (1982).
[10] Y. Hahn, Phys. Rev. **139**, B212 (1965).
[11] M. A. Abdel-Raouf, Physica B + C **97**, 103 (1979).
[12] M. Cohen, T. Feldmann, and R. P. McEachran, J. Phys. B **15**, 677 (1982).
[13] E. A. G. Armour and D. M. Schrader, Can. J. Phys. **60**, 581 (1982).
[14] M. A. Abdel-Raouf, J. Math. Phys. **22**, 2849 (1981).
[15] L. C. Biedenharn and J. M. Blatt, Phys. Rev. **93**, 1387 (1954).

Null strings and Bianchi identities

G. F. Torres del Castillo
Departamento de Física Matemática, Instituto de Ciencias de la Universidad Autónoma de Puebla, M. A. C. #219, 72000 Puebla, Pue. México and
Departamento de Física, Centro de Investigación y de Estudios Avanzados del IPN, Apartado Postal 14-740, 07000 México, D. F., Mexico

(Received 13 April 1982; accepted for publication 6 October 1982)

The results of the theory of complexified V_4's which admit a null string are reexamined by using systematically the Bianchi identities. The formal reasons which permit the integration of the $G\mu\nu = 0$ equations to one differential constraint in the case of the $\mathscr{H}\mathscr{H}$ structures are exhibited. Some new results concerning complexified Einstein–Weyl equations in the presence of a null string are given as well.

PACS numbers: 04.20. – q, 02.30.Jr

1. INTRODUCTION

In 1976 Plebański and Robinson[1] found that if the conformal curvature of a complex space-time is algebraically degenerate, then the Einstein's vacuum equations can be reduced to a single partial differential equation of second order involving only quadratic nonlinearities. Their result was based on the fact that the algebraic degeneracy of the conformal curvature implies the existence of a congruence of totally null bidimensional surfaces.[2] This congruence ("null string"), and the algebraic degeneracy of the conformal curvature permits one to introduce coordinate systems in which the Einstein's vacuum field equations divide into three triple (triples) and one (singlet) which, when integrated in the appropriate order, led to the partial differential equation referred to above; however, this integration process can still be considerably simplified by means of the spinorial structure induced by the null string.[3] Moreover, García, Plebański, and Robinson[4] reduced all the Einstein–Maxwell equations with cosmological constant to two differential conditions on two functions assuming that the complex space-time admits a null string and correlating the electromagnetic field to that null string.

In both cases it was found that the first triple of the Einstein's equations is of the form $\partial_{\dot A}\partial_{\dot B}\Lambda_+ = 0$ where $\partial_{\dot A}$ denotes partial derivatives tangent to the null string. Solving the first triple and the scalar equation, the second triple can be written in the form $\partial_{\dot A}\partial_{\dot B}\Lambda_0 = 0$. Once this second triple is integrated, the last triple can be rearranged to read $\partial_{\dot A}\partial_{\dot B}\Lambda_- = 0$. In this paper we show that assuming the existence of a null string and the algebraic degeneracy of the conformal curvature, the Bianchi identities imply this simple structure of the Einstein's equations. We also show how this approach can be applied to other sources of the gravitational field. In Sec. 2 we deduce, using the spinorial formalism, the form of the metric for a space-time which admits a null string and is algebraically degenerate. In Sec. 4 we show that in this kind of space-time, the solutions for (massless) $D\,(0, s)$ spinor fields can be written in terms of a potential function. The formalism and notation used here follows Ref. 5. All the spinorial indices are manipulated according to the convention $\psi_{\dot A} = \epsilon_{\dot A\dot B}\psi^{\dot B}$, $\psi^{\dot A} = \psi_{\dot B}\epsilon^{\dot A\dot B}$, and similarly for undotted indices.

2. THE FORM OF THE METRIC

Let $g^{A\dot B}$ be a null (cotangent) tetrad (with $g^{A\dot B}\cdot g^{C\dot D} = -2\epsilon^{AC}\epsilon^{\dot B\dot D}$), a spinor l_A defines a system of differential equations given by

$$l_A g^{A\dot B} = 0. \tag{2.1}$$

When the system (2.1) is completely integrable it gives a family, or congruence, of bidimensional surfaces—a null string. The tangent plane at each point of these surfaces is spanned by the tangent vectors $l^A\partial_{A\dot B}$ where $\partial_{A\dot B}$ are tangent vectors such that $g^{A\dot B}(\partial_{C\dot D}) = -2\delta^A_C\delta^{\dot B}_{\dot D}$. Clearly, the vectors $l^A\partial_{A\dot B}$ are null and orthogonal. The integrability condition of the system (2.1) is

$$d(l_A g^{A\dot B}) \wedge l_C g^{C\dot 1} \wedge l_D g^{D\dot 2} = 0. \tag{2.2}$$

If l_A satisfies Eq. (2.2) it follows from the second structural equations that $l^A l^B l^C l^D C_{ABCD} = 0$, which means that l_A is a Debever–Penrose (DP) spinor. As shown in Ref. 2, when l_A defines a null string and $l^A l^B C_{AB\dot C\dot D} = 0$, then l_A is a multiple DP spinor (i.e., $l^A l^B l^C C_{ABCD} = 0$). Conversely, if l_A is a multiple DP spinor and $C_{AB\dot C\dot D} = 0$, but $C_{ABCD} \neq 0$, then l_A defines a null string [i.e., l_A satisfies Eq. (2.2)]. Now one has the following

Lemma: If the spinor l_A satisfies Eq. (2.2) it is a multiple DP spinor iff there exists a function ϕ such that

$$l^A\partial_{A\dot B}\ln\phi = m^A l^C\nabla_{A\dot B}l_C, \tag{2.3}$$

where m^A is a spinor such that $m^A l_A = 1$.

Proof: Taking l_A and m_A as a basis in such a way that $l_A = \delta^2_A$ and $m_A = \delta^1_A$, Eq. (2.3) is seen to be equivalent to

$$\partial_{1\dot B}\ln\phi = -\Gamma_{11}(\partial_{2\dot B}), \tag{2.4}$$

while Eq. (2.2) is just

$$\Gamma_{11}(\partial_{1\dot A}) = 0. \tag{2.5}$$

The integrability condition of Eq. (2.4) is $\partial_1{}^{\dot A}\partial_{1\dot A}\ln\phi = -\partial_1{}^{\dot A}(\Gamma_{11}(\partial_{2\dot A}))$. The left-hand side is just the commutator $|\partial_{1\dot 1}, \partial_{1\dot 2}|\ln\phi$, which, in view of (2.5) and (2.4), is equal to $[\Gamma_{12}(\partial_1{}^{\dot A}) + \Gamma^{\dot A\dot B}(\partial_{1\dot B})]\Gamma_{11}(\partial_{2\dot A})$. From the second structural equations and $\Gamma_{11}(\partial_{1\dot A}) = 0$, one finds that $C_{1111} = 0$ and $[\Gamma_{12}(\partial_1{}^{\dot A}) + \Gamma^{\dot A\dot B}(\partial_{1\dot B})]\Gamma_{11}(\partial_{2\dot A}) = -\partial_1{}^{\dot A}(\Gamma_{11}(\partial_{2\dot A})) - 4C_{1112}$. Thus, assuming (2.5), the integrability of Eq. (2.4)

 0022-2488/83/030590-07$02.50

and the fact that l_A is a multiple DP spinor are equivalent statements.

In the forthcoming it will be assumed that the complex space-time admits a null string defined by a spinor l_A and that l_A is a multiple DP spinor.

The complete integrability of (2.1) means that there exists a pair of independent functions q^1, q^2 such that

$$l_A g^{A\dot{B}} = \sqrt{2} l^{\dot{B}}{}_{\dot{C}} dq^{\dot{C}}, \tag{2.6}$$

where $(l^{\dot{B}}{}_{\dot{C}})$ is a nonsingular matrix. Let ϕ be a solution to Eq. (2.3). A new null tetrad $g'^{A\dot{B}}$, is introduced by

$$g'^{2\dot{A}} = -\sqrt{2}\phi^{-2} dq^{\dot{A}}. \tag{2.7}$$

Thus, from Eq. (2.6),

$$g'^{2\dot{A}} = -\phi^{-2} l^{-1\dot{A}}{}_{\dot{B}} l_C g^{C\dot{B}}, \tag{2.8}$$

where $(l^{-1\dot{A}}{}_{\dot{B}})$ denotes the inverse of $(l^{\dot{A}}{}_{\dot{B}})$ (i.e., $l^{\dot{A}}{}_{\dot{B}} l^{-1\dot{B}}{}_{\dot{C}} = \delta^{\dot{A}}_{\dot{C}}$). The remaining members of the tetrad $g'^{A\dot{B}}$ must be given by

$$g'^{1\dot{A}} = \phi^2 l_{\dot{C}}{}^{A} m_D g^{D\dot{C}} + \eta g'^{2\dot{A}}, \tag{2.9}$$

where η is an arbitrary function. Therefore, the tangent tetrad $\partial'_{A\dot{B}}$ is determined by

$$\begin{aligned} \partial'_{1\dot{A}} &= -\phi^{-2} l_{\dot{A}}{}^{-1\dot{B}} l^{C} \partial_{C\dot{B}}, \\ \partial'_{2\dot{A}} &= -\phi^2 l^{\dot{C}}{}_{\dot{A}} m^D \partial_{D\dot{C}} - \eta \partial'_{1\dot{A}}. \end{aligned} \tag{2.10}$$

Using the expression for $\partial'_{1\dot{A}}$ given in (2.10) and the relation, $0 = \sqrt{2} d\, dq^{\dot{C}} = d(l^{-1\dot{C}}{}_{\dot{B}} l_A g^{A\dot{B}})$, a direct computation shows that,

$$\begin{aligned} [\partial'_{1\dot{1}}, \partial'_{1\dot{2}}] &= 2\phi^{-4} (\det l^{-1\dot{R}}{}_{\dot{S}}) \\ &\quad \times [l^A \partial_{A\dot{C}} \ln \phi - m^A l^B \nabla_{A\dot{C}} l_B] l^D \partial_D{}^{\dot{C}}, \end{aligned} \tag{2.11}$$

which is zero by virtue of Eq. (2.3). Let $p^{\dot{A}}$ be a pair of functions such that

$$\partial'_{1\dot{A}} p^{\dot{B}} = \sqrt{2} \delta^{\dot{B}}_{\dot{A}}. \tag{2.12}$$

The integrability conditions of Eq. (2.12) are trivially satisfied since $[\partial'_{1\dot{1}}, \partial'_{1\dot{2}}] = 0$. Thus one has

$$d p^{\dot{A}} = (-\tfrac{1}{2}\partial'_{C\dot{B}} p^{\dot{A}}) g'^{C\dot{B}} = -\sqrt{2} g'^{1\dot{A}} - \tfrac{1}{2}(\partial'_{2\dot{B}} p^{\dot{A}}) g'^{2\dot{B}}; \tag{2.13}$$

hence, by Eq. (2.7),

$$g'^{1\dot{A}} = -\sqrt{2}(dp^{\dot{A}} + (1/\sqrt{2})\phi^{-2} (\partial'^{\dot{B}}_2 p^{\dot{A}}) dq_{\dot{B}}). \tag{2.14}$$

From Eqs. (2.10) and (2.12) one finds that

$$\partial'_{2\dot{A}} p^{\dot{A}} = -\phi^2 l^{\dot{C}}{}_{\dot{A}} m^D (\partial_{D\dot{C}} p^{\dot{A}}) - 2\sqrt{2}\eta, \tag{2.15}$$

therefore, choosing $\eta = -(1/2\sqrt{2})\phi^2 l^{\dot{C}}{}_{\dot{A}} m^D (\partial_{D\dot{C}} p^{\dot{A}})$ one obtains that

$$Q^{A\dot{B}} \equiv -(1/\sqrt{2})\phi^{-2} (\partial^{\dot{A}}_2 p^{\dot{B}}) \tag{2.16}$$

is a symmetric object. Thus,

$$g'^{1\dot{A}} = -\sqrt{2}(d p^{\dot{A}} - Q^{A\dot{B}} dq_{\dot{B}}) \equiv \sqrt{2} E^{\dot{A}}. \tag{2.17}$$

From Eqs. (2.7) and (2.17) it follows that the function $q^{\dot{A}}$ and $p^{\dot{A}}$ can be used as local coordinates, and that the tangent tetrad $\partial'_{A\dot{B}}$ is given by

$$\begin{aligned} \partial'_{1\dot{A}} &= \sqrt{2} \frac{\partial}{\partial p^{\dot{A}}} \equiv \sqrt{2} \partial_{\dot{A}}, \\ \partial'_{2\dot{A}} &= \sqrt{2}\phi^2 \left(\frac{\partial}{\partial q^{\dot{A}}} - Q_{\dot{A}}{}^{\dot{B}} \frac{\partial}{\partial p^{\dot{B}}} \right) \equiv \sqrt{2}\phi^2 D_{\dot{A}}. \end{aligned} \tag{2.18}$$

Thus, the metric of a space-time which admits a null string and is algebraically degenerate has the form[6]

$$\begin{aligned} ds^2 &= -\tfrac{1}{2} g'_{A\dot{B}} \otimes g'^{A\dot{B}} \\ &= 2\phi^{-2} dq^{\dot{A}} \underset{s}{\otimes} (dp_{\dot{A}} - Q_{\dot{A}}{}^{\dot{B}} dq_{\dot{B}}). \end{aligned} \tag{2.19}$$

Using the first structural equations one finds that the connection one-forms for the tetrad $g'^{A\dot{B}}$ are

$$\begin{aligned} \Gamma_{11} &= -\phi^{-3} J_{\dot{A}} dq^{\dot{A}}, \\ \Gamma_{12} &= \tfrac{1}{2}[\partial^{\dot{A}} Q_{\dot{A}\dot{B}} + \phi^{-1} D_{\dot{B}} \phi] dq^{\dot{B}} - \tfrac{3}{2}\phi^{-1} J_{\dot{A}} E^{\dot{A}}, \\ \Gamma_{22} &= \phi^2 D^{\dot{A}} Q_{\dot{A}\dot{B}} dq^{\dot{B}} - \phi D_{\dot{A}} \phi E^{\dot{A}}, \\ \Gamma_{\dot{A}\dot{B}} &= \partial_{(\dot{A}} Q_{\dot{B})\dot{C}} dq^{\dot{C}} + \phi^{-1} J_{(\dot{A}} E_{\dot{B})} - \phi^{-1} D_{(\dot{A}} \phi\, dq_{\dot{B})}, \end{aligned} \tag{2.20}$$

where $J_{\dot{A}} \equiv \partial_{\dot{A}} \phi$. Similarly, from the second structural equations one obtains the components of the conformal curvature:

$$\begin{aligned} C^{(5)} &= 2C_{1111} = 0, \\ C^{(4)} &= 2C_{1112} = 0, \\ C^{(3)} &= 2C_{1122} = -\tfrac{1}{3}\phi^2 \partial_{\dot{A}} \partial_{\dot{B}} Q^{\dot{A}\dot{B}}, \\ C^{(2)} &= 2C_{1222} = -\phi^4 \partial_{\dot{A}} D_{\dot{B}} Q^{\dot{A}\dot{B}}, \\ C^{(1)} &= 2C_{2222} = 2\phi^6 [D_{\dot{A}} D_{\dot{B}} Q^{\dot{A}\dot{B}} \\ &\qquad - (D_{\dot{B}} Q^{\dot{A}\dot{B}}) \partial^{\dot{C}} Q_{\dot{A}\dot{C}}], \end{aligned} \tag{2.21a}$$

$$C_{\dot{A}\dot{B}\dot{C}\dot{D}} = -\phi^2 \partial_{(\dot{A}} \partial_{\dot{B}} Q_{\dot{C}\dot{D})}. \tag{2.21b}$$

The scalar curvature and the spinorial components of the traceless part of the Ricci tensor are

$$R = -2\phi^2 \partial_{\dot{A}} \partial_{\dot{B}} Q^{\dot{A}\dot{B}} - 12\phi^3 \partial_{\dot{A}} (\phi^{-2} D^{\dot{A}} \phi), \tag{2.22a}$$

$$\begin{aligned} C_{11\dot{A}\dot{B}} &= -\phi^{-1} \partial_{\dot{A}} \partial_{\dot{B}} \phi, \\ C_{12\dot{A}\dot{B}} &= \tfrac{1}{2}\phi^2 \partial^{\dot{C}} \partial_{(\dot{A}} Q_{\dot{B})\dot{C}} - \phi \partial_{(\dot{A}} D_{\dot{B})} \phi, \\ C_{22\dot{A}\dot{B}} &= \phi^5 \partial_{(\dot{A}} (\phi^{-1} D^{\dot{C}} Q_{\dot{B})\dot{C}}) \\ &\quad - \phi^3 (D^{\dot{C}} \phi) \partial_{\dot{C}} Q_{\dot{A}\dot{B}} - \phi^3 D_{(\dot{A}} D_{\dot{B})} \phi. \end{aligned} \tag{2.22b}$$

3. SOLUTION OF EINSTEIN'S EQUATIONS

The general expressions for the connection and the curvature given in Sec. 2 [Eqs. (2.20)–(2.22)], will be used to determine the functions ϕ and $Q_{\dot{A}\dot{B}}$ [Eq. (2.19)] when the Einstein's equations for the gravitational field with sources are fulfilled.

The Bianchi identities expressed in spinorial form are

$$\nabla^S{}_{\dot{D}} C_{SABC} + \nabla_{(A}{}^{\dot{S}} C_{BC)\dot{D}\dot{S}} = 0, \tag{3.1a}$$

$$\nabla_D{}^{\dot{S}} C_{\dot{S}\dot{A}\dot{B}\dot{C}} + \nabla^S{}_{(\dot{A}} C_{|DS|\dot{B}\dot{C})} = 0, \tag{3.1b}$$

$$\nabla^{R\dot{S}} C_{AR\dot{B}\dot{S}} + \tfrac{1}{8}\partial_{A\dot{B}} R = 0. \tag{3.1c}$$

Referring all the quantities in Eqs. (3.1) to the tetrad $\partial'_{A\dot{B}}$, setting $A = B = C = 1$ in Eq. (3.1a) and using Eqs. (2.18), (2.20), and (2.21a) one finds that $\partial^{\dot{S}}(\phi C_{11\dot{D}\dot{S}}) = 0$ which, since $C_{11\dot{D}\dot{S}}$ is symmetric (see Appendix of Ref. 3), implies the existence of a function Λ_+ such that

$$\phi C_{11\dot{D}\dot{S}} = \partial_{\dot{D}} \partial_{\dot{S}} \Lambda_+. \tag{3.2}$$

In fact, from Eq. (2.22b) one sees that $\Lambda_+ = -\phi$; therefore, from $C_{11\dot{A}\dot{B}} = 0$ one has

$$\phi = J_{\dot{A}}\, p^{\dot{A}} + \kappa, \tag{3.3}$$

where $J_{\dot{A}}$ and κ are functions of $q^{\dot{A}}$ only and can be made constant using the freedom in the choice of the coordinates $q^{\dot{A}}$. The case when $J_{\dot{A}} = 0$ is called case I, the other is designated as case II.

Taking $A = 1$ in Eq. (3.1c) and assuming that $C_{11\dot{A}\dot{B}} = 0$ [that is, taking ϕ as given in Eq. (3.3)] it follows that

$$\phi^4 \partial^{\dot{S}}(\phi^{-4} C_{12\dot{B}\dot{S}}) + \tfrac{1}{8}\partial_{\dot{B}} R = 0. \tag{3.4}$$

Thus, when $C_{11\dot{A}\dot{B}} = 0$ and if $\partial_{\dot{B}} R = 0$ there exists a function Λ_0 such that

$$\phi^{-4} C_{12\dot{B}\dot{S}} = \partial_{\dot{B}}\partial_{\dot{S}}\Lambda_0. \tag{3.5}$$

Taking now $A = 2$ in Eq. (3.1c), $C_{11\dot{A}\dot{B}} = 0$, $\partial_{\dot{B}} R = 0$, and using Eq. (3.5) one finds that

$$\partial^{\dot{S}}(\phi^{-6} C_{22\dot{B}\dot{S}}) - \partial_{\dot{B}} D^{\dot{S}}\partial_{\dot{S}}\Lambda_0 + \tfrac{1}{8}\phi^{-4} D_{\dot{B}} R = 0. \tag{3.6}$$

Therefore, if one further assumes $R = \text{const}$, it follows that $\partial^{\dot{B}}\partial^{\dot{S}}(\phi^{-6} C_{22\dot{B}\dot{S}}) = 0$ which implies (see Appendix of Ref. 3) the existence of a spinor $X_{\dot{S}}$ such that

$$\phi^{-6} C_{22\dot{B}\dot{S}} = -\partial_{(\dot{B}} X_{\dot{S})}. \tag{3.7}$$

Note that when $C_{11\dot{A}\dot{B}} = 0$ and if $C_{12\dot{A}\dot{B}} = 0$ and R is a constant, Eq. (3.6) implies then the existence as a function Λ_- such that

$$\phi^{-6} C_{22\dot{B}\dot{S}} = \partial_{\dot{B}}\partial_{\dot{S}}\Lambda_-. \tag{3.8}$$

Assuming $J_{\dot{A}} \neq 0$ and using Eqs. (2.22a) and (3.3) one finds that the condition $\partial_{\dot{B}} R = 0$ implies that the $Q^{\dot{A}\dot{B}}$ must have the form

$$\text{II: } Q^{\dot{A}\dot{B}} = -\phi^3 \partial^{(\dot{A}} A^{\dot{B})} + \left[\phi K^{\dot{C}} \frac{\partial J^{\dot{D}}}{\partial q^{\dot{C}}} J_{\dot{D}} - J^{\dot{C}} \frac{\partial \phi}{\partial q^{\dot{C}}} - \frac{R}{24}\right] K^{\dot{A}} K^{\dot{B}}, \tag{3.9a}$$

where $K^{\dot{A}}$ is a spinor such that $K^{\dot{A}} J_{\dot{A}} = 1$ and $A^{\dot{B}}$ is an arbitrary spinor. If $J_{\dot{A}} = 0$, the condition $\partial_{\dot{B}} R = 0$ implies then

$$\text{I: } Q^{\dot{A}\dot{B}} = -\partial^{(\dot{A}} A^{\dot{B})} - (R/12)\kappa^{-2} p^{\dot{A}} p^{\dot{B}}. \tag{3.9b}$$

From Eqs. (2.22b), (3.3), and (3.9) one obtains that the function Λ_0 appearing in Eq. (3.5) can be taken as

$$\text{II: } \Lambda_0 = \tfrac{1}{4}\phi^{-1}\partial_{\dot{C}}\left(\phi^2 A^{\dot{C}} - 2K^{\dot{D}} \frac{\partial \phi}{\partial q^{\dot{D}}} K^{\dot{C}}\right), \tag{3.10a}$$

$$\text{I: } \Lambda_0 = \tfrac{1}{4}\kappa^{-2}\partial_{\dot{C}} A^{\dot{C}}. \tag{3.10b}$$

Similarly, if R is a constant, from Eqs. (2.22b), (3.3), and (3.9) it follows that the spinor $X_{\dot{A}}$ given in Eq. (3.7), can be taken as

$$\begin{aligned}\text{II: } X_{\dot{A}} = {} & -\phi^{-2} D^{\dot{C}} Q_{\dot{C}\dot{A}} - J^{\dot{C}} \frac{\partial A_{\dot{A}}}{\partial q^{\dot{C}}} \\ & - A_{\dot{A}} \frac{\partial J^{\dot{C}}}{\partial q^{\dot{C}}} + \frac{\partial \phi}{\partial q^{\dot{C}}} \partial^{\dot{C}} A_{\dot{A}} \\ & - \frac{\partial J^{\dot{C}}}{\partial q^{\dot{A}}} A_{\dot{C}} + \phi^{-2} K^{\dot{C}} J^{\dot{D}} \frac{\partial^2 \phi}{\partial q^{\dot{C}} \partial q^{\dot{D}}} K_{\dot{A}} \\ & + \phi^{-1} K^{\dot{C}} J^{\dot{D}} \frac{\partial^2 J_{\dot{B}}}{\partial q^{\dot{C}} \partial q^{\dot{D}}} K^{\dot{B}} K_{\dot{A}} \\ & - \frac{1}{2}\phi^{-2} K^{\dot{C}} K^{\dot{D}} \frac{\partial^2 \phi}{\partial q^{\dot{C}} \partial q^{\dot{D}}} J_{\dot{A}} \\ & - \frac{1}{2}\phi^{-1} K^{\dot{C}} K^{\dot{D}} \frac{\partial^2 J_{\dot{A}}}{\partial q^{\dot{C}} \partial q^{\dot{D}}} \\ & + \phi^{-2} \frac{\partial K^{\dot{S}}}{\partial q^{\dot{S}}} \left\{ J^{\dot{C}} \frac{\partial \phi}{\partial q^{\dot{C}}} - 2\phi K^{\dot{C}} \frac{\partial J^{\dot{D}}}{\partial q^{\dot{C}}} J_{\dot{D}} \right. \\ & \left. - \phi J^{\dot{C}} \frac{\partial J^{\dot{D}}}{\partial q^{\dot{C}}} K_{\dot{D}} + \frac{R}{24}\right\} K_{\dot{A}}, \end{aligned} \tag{3.11a}$$

while if $J_{\dot{A}} = 0$ one obtains

$$\begin{aligned}\text{I: } X_{\dot{A}} = {} & -\kappa^{-2} D^{\dot{C}} Q_{\dot{C}\dot{A}} + \kappa^{-3} \frac{\partial \kappa}{\partial q^{\dot{C}}} \partial^{\dot{C}} A_{\dot{A}} \\ & + \kappa^{-3} \frac{\partial^2 \kappa}{\partial q^{\dot{A}} \partial q^{\dot{C}}} p^{\dot{C}} - \frac{R}{6}\kappa^{-5} \frac{\partial \kappa}{\partial q^{\dot{C}}} p^{\dot{C}} p_{\dot{A}}. \end{aligned} \tag{3.11b}$$

Assuming $R = \text{const}$, Eq. (3.6) can also be written as

$$\partial^{\dot{S}}(\phi^{-6} C_{22\dot{B}\dot{S}} + D_{\dot{S}}\partial_{\dot{B}}\Lambda_0 + (\partial_{\dot{S}} Q_{\dot{B}\dot{C}})\partial^{\dot{C}}\Lambda_0) = 0, \tag{3.12}$$

which implies the existence of a spinor $\chi_{\dot{B}}$ such that

$$\partial_{\dot{S}}\chi_{\dot{B}} = \gamma^{-6} C_{22\dot{B}\dot{S}} + D_{\dot{S}}\partial_{\dot{B}}\Lambda_0 + (\partial_{\dot{S}} Q_{\dot{B}\dot{C}})\partial^{\dot{C}}\Lambda_0. \tag{3.13}$$

Therefore, one has

$$\begin{aligned}\partial^{\dot{S}}\chi_{\dot{S}} &= D^{\dot{S}}\partial_{\dot{S}}\Lambda_0 + (\partial^{\dot{S}} Q_{\dot{S}\dot{C}})\partial^{\dot{C}}\Lambda_0 \\ &= \partial^{\dot{S}}(2(\partial^{\dot{C}} Q_{\dot{C}\dot{S}})\Lambda_0 - D_{\dot{S}}\Lambda_0) - 2\Lambda_0 \partial^{\dot{C}}\partial^{\dot{S}} Q_{\dot{C}\dot{S}}. \end{aligned} \tag{3.14}$$

From Eqs. (3.9) and (3.10) it follows that

$$\text{II: } \partial^{\dot{A}}\partial^{\dot{B}} Q_{\dot{A}\dot{B}} = 3\phi J^{\dot{A}}\partial_{\dot{A}}(\phi^{-1}\partial_{\dot{C}}(\phi^2 A^{\dot{C}})) = 12\phi J^{\dot{A}}\partial_{\dot{A}}\Lambda_0, \tag{3.15a}$$

$$\text{I: } \partial^{\dot{A}}\partial^{\dot{B}} Q_{\dot{A}\dot{B}} = -(R/2)\kappa^{-2}. \tag{3.15b}$$

Thus, the spinor $\chi_{\dot{B}}$ is given by

$$\text{II: } \chi_{\dot{B}} = 2(\partial^{\dot{C}} Q_{\dot{C}\dot{B}})\Lambda_0 - D_{\dot{B}}\Lambda_0 + 12\phi\Lambda_0^2 J_{\dot{B}} - \partial_{\dot{B}}\sigma, \tag{3.16a}$$

$$\text{I: } \chi_{\dot{B}} = 2(\partial^{\dot{C}} Q_{\dot{C}\dot{B}})\Lambda_0 - D_{\dot{B}}\Lambda_0 - (R/4)\kappa^{-4} A_{\dot{B}} - \partial_{\dot{B}}\sigma, \tag{3.16b}$$

where σ is any arbitrary function. Inserting this result into Eq. (3.13) one finds

$$\begin{aligned}\text{II: } & \phi^{-6} C_{22\dot{B}\dot{S}} + 2D_{(\dot{B}}\partial_{\dot{S})}\Lambda_0 + 2(\partial_{\dot{C}} Q_{\dot{B}\dot{S}})\partial^{\dot{C}}\Lambda_0 \\ & - 2\Lambda_0 \partial^{\dot{C}}\partial_{\dot{S}} Q_{\dot{B}\dot{C}} - 12\partial_{\dot{S}}(\phi\Lambda_0^2 J_{\dot{B}}) = -\partial_{\dot{B}}\partial_{\dot{S}}\sigma, \end{aligned} \tag{3.17a}$$

$$\begin{aligned}\text{I: } & \phi^{-6} C_{22\dot{B}\dot{S}} + 2D_{(\dot{B}}\partial_{\dot{S})}\Lambda_0 + 2(\partial_{\dot{C}} Q_{\dot{B}\dot{S}})\partial^{\dot{C}}\Lambda_0 \\ & - 2\Lambda_0 \partial^{\dot{C}}\partial_{\dot{S}} Q_{\dot{B}\dot{C}} + (R/4)\kappa^{-4}\partial_{\dot{S}} A_{\dot{B}} = -\partial_{\dot{B}}\partial_{\dot{S}}\sigma. \end{aligned} \tag{3.17b}$$

Equivalently, using Eq. (3.7), the last two equations can be written as

$$\text{II: } \partial_{(\dot{B}}\{X_{\dot{S})} - 2D_{\dot{S})}\Lambda_0 + 2\Lambda_0(\partial^{\dot{C}}Q_{\dot{S})\dot{C}}) + 12\phi\Lambda\,{}^{2}_{0}J_{\dot{S})}\} = \partial_{\dot{B}}\partial_{\dot{S}}\sigma \tag{3.18a}$$

and

$$\text{I: } \partial_{(\dot{B}}\{\chi_{\dot{S})} - 2D_{\dot{S})}\Lambda_0 + 2\Lambda_0(\partial^{\dot{C}}Q_{\dot{S})\dot{C}}) - (R/4)\kappa^{-4}A_{\dot{S})}\} = \partial_{\dot{B}}\partial_{\dot{S}}\sigma. \tag{3.18b}$$

Since the energy-momentum tensor of any physical field must satisfy a relation analogous to the contracted Bianchi identities [Eq. (3.1c)], denoting by $T_{AB\dot{C}\dot{D}}$ the spinorial components of the traceless part of the energy-momentum tensor and assuming that $T_{11\dot{C}\dot{D}} = 0$ and that the trace of the energy-momentum tensor is constant, it follows that

$$T_{12\dot{C}\dot{D}} = (1/8\pi)\phi^{4}\partial_{\dot{C}}\partial_{\dot{D}}T_0, \tag{3.19}$$

where T_0 is some function made out of the matter field. Therefore Einstein's equations and Eqs. (3.5) and (3.19) imply

$$\Lambda_0 = T_0 + \tfrac{1}{2}(L_{\dot{A}}\,p^{\dot{A}} + \xi), \tag{3.20}$$

where $L_{\dot{A}}$ and ξ are functions of $q^{\dot{R}}$ only.

In the case with $J_{\dot{A}} \neq 0$, Eq. (3.15a) gives

$$\partial^{\dot{A}}\partial^{\dot{B}}Q_{\dot{A}\dot{B}} = 6\phi J^{\dot{A}}(L_{\dot{A}} + 2\partial_{\dot{A}}T_0) \equiv 6\phi(\mu + 2J^{\dot{A}}\partial_{\dot{A}}T_0). \tag{3.21}$$

The transformation $A^{\dot{A}} \to A^{\dot{A}} + \tfrac{1}{2}L_{\dot{B}}K^{\dot{B}}(\kappa K^{\dot{A}} - p^{\dot{A}}) - \xi K^{\dot{A}}$ leaves the value of $Q_{\dot{A}\dot{B}}$ unchanged [see Eq. (3.9a)] and eliminates the irrelevant terms on the right-hand side of Eq. (3.20) giving

$$\Lambda_0 = T_0 + \tfrac{1}{2}\mu K^{\dot{A}}p_{\dot{A}}, \tag{3.22}$$

which has as the most general solution

$$A^{\dot{A}} = G^{\dot{A}} + \phi^{-2}\partial^{\dot{A}}W + 2\phi^{-2}K^{\dot{A}}K^{\dot{B}}\frac{\partial\phi}{\partial q^{\dot{B}}} + \mu K^{\dot{A}}K^{\dot{B}}p_{\dot{B}}, \tag{3.23}$$

where W is an arbitrary function and $G^{\dot{A}}$ is such that

$$\phi^{-1}\partial_{\dot{A}}(\phi^{2}G^{\dot{A}}) = 4T_0. \tag{3.24}$$

In the case $J_{\dot{A}} = 0$, the transformation $A^{\dot{A}} \to A^{\dot{A}} - \xi\kappa^{2}p^{\dot{A}}$ leaves $Q_{\dot{A}\dot{B}}$ invariant, giving

$$\Lambda_0 = T_0 + \tfrac{1}{2}L_{\dot{A}}\,p^{\dot{A}}, \tag{3.25}$$

which has the solution

$$A^{\dot{A}} = \kappa^{3}G^{\dot{A}} + \partial^{\dot{A}}\Theta + \tfrac{2}{3}\kappa^{2}L_{\dot{B}}\,p^{\dot{B}}p^{\dot{A}}, \tag{3.26}$$

where Θ is any arbitrary function and $G^{\dot{A}}$ satisfies Eq. (3.24)

Using the freedom involved in the definition of the coordinates $q^{\dot{A}}$ one makes $J_{\dot{A}}$ and κ constant; moreover, in the case with $J_{\dot{A}} = 0$, κ will be taken as 1. This choice of coordinates simplifies greatly the remaining computations. However, even when $J_{\dot{A}}$ and κ are constants, there is still considerable freedom in the choice of the coordinates $q^{\dot{A}}$, $p^{\dot{A}}$ (see Refs. 3 and 7).

Inserting the respective expression for $A^{\dot{A}}$ into Eqs. (3.9a) and (3.9b) one finds that

$$\text{II: } Q_{\dot{A}\dot{B}} = -\phi^{3}\partial_{(\dot{A}}B_{\dot{B})} + (\mu\phi^{3} - R/24)K_{\dot{A}}K_{\dot{B}}, \quad B_{\dot{A}} \equiv G_{\dot{A}} + \phi^{-2}\partial_{\dot{A}}W, \tag{3.27a}$$

and

$$\text{I: } Q_{\dot{A}\dot{B}} = -\partial_{(\dot{A}}B_{\dot{B})} - \tfrac{2}{3}L_{(\dot{A}}\,p_{\dot{B})} - (R/12)p_{\dot{A}}\,p_{\dot{B}}, \quad B_{\dot{A}} \equiv G_{\dot{A}} + \partial_{\dot{A}}\Theta. \tag{3.27b}$$

One may now evaluate the left-hand sides of Eqs. (3.18) which after some rearrangement read

$$\text{II: } X_{\dot{A}} - 2D_{\dot{A}}\Lambda_0 + 2\Lambda_0(\partial^{\dot{C}}Q_{\dot{A}\dot{C}}) + 12\phi\Lambda\,{}^{2}_{0}J_{\dot{A}} = \partial_{\dot{A}}\left\{\frac{1}{2}\phi^{-2}Q^{\dot{B}\dot{C}}Q_{\dot{B}\dot{C}} - \phi\frac{\partial A^{\dot{C}}}{\partial q^{\dot{C}}} + (\phi J_{\dot{C}}B^{\dot{C}})^{2} + 2\mu\phi^{2}K_{\dot{B}}\,p^{\dot{B}}J^{\dot{C}}B_{\dot{C}} + 2(T_0\phi)^{2} + \frac{R}{12}K^{\dot{B}}B_{\dot{B}} + \frac{3}{2}(\mu\phi K^{\dot{C}}p_{\dot{C}})^{2} + 2\mu T_0\phi^{2}K^{\dot{C}}p_{\dot{C}}\right\}, \tag{3.28a}$$

$$\text{I: } X_{\dot{A}} - 2D_{\dot{A}}\Lambda_0 + 2\Lambda_0(\partial^{\dot{C}}Q_{\dot{A}\dot{C}}) - \frac{R}{4}A_{\dot{A}} = \partial_{\dot{A}}\left\{\frac{1}{2}Q^{\dot{B}\dot{C}}Q_{\dot{B}\dot{C}} - \frac{\partial B^{\dot{C}}}{\partial q^{\dot{C}}} - \frac{R}{4}B_{\dot{C}}\,p^{\dot{C}} + \frac{1}{2}(2T_0 + L_{\dot{C}}\,p^{\dot{C}})^{2} - \frac{2}{3}\frac{\partial L_{\dot{C}}}{\partial q^{\dot{B}}}p^{\dot{B}}p^{\dot{C}}\right\}. \tag{3.28b}$$

Substituting these results into Eqs. (3.18) and considering the energy-momentum tensor as known one obtains a differential constraint on W or Θ, the solutions to which determine the metric. However, we can improve our result by noting that

$$\text{II: } -2D_{\dot{A}}\left(\frac{1}{2}\mu K^{\dot{B}}p_{\dot{B}}\right) + \mu K^{\dot{B}}p_{\dot{B}}(\partial^{\dot{C}}Q_{\dot{A}\dot{C}}) + 12\phi\left(\mu T_0K^{\dot{B}}p_{\dot{B}} + \frac{1}{4}\mu^{2}(K^{\dot{B}}p_{\dot{B}})^{2}\right)J_{\dot{A}} = 3\mu\phi^{2}G_{\dot{A}} - \frac{1}{2}K^{\dot{B}}\frac{\partial\mu}{\partial q^{\dot{B}}}p_{\dot{A}} + \partial_{\dot{A}}\left\{2\mu\phi^{2}K_{\dot{B}}\,p^{\dot{B}}J^{\dot{C}}B_{\dot{C}} + \frac{3}{2}(\mu\phi K^{\dot{C}}p_{\dot{C}})^{2} + 2\mu T_0\phi^{2}K^{\dot{C}}p_{\dot{C}} - \mu\phi^{3}K^{\dot{B}}B_{\dot{B}} + 3\mu W + \frac{1}{2}K_{\dot{B}}\,p^{\dot{B}}\frac{\partial\mu}{\partial q^{\dot{C}}}p^{\dot{C}}\right\}, \tag{3.29}$$

and that

$$\text{I: } -2D_{\dot A}\left(\frac{1}{2}L_{\dot B}\,p^{\dot B}\right)+L_{\dot B}\,p^{\dot B}(\partial^{\dot C}Q_{\dot A\dot C})-\frac{R}{4}A_{\dot A}=-\frac{R}{4}G_{\dot A}+\frac{1}{2}\frac{\partial L^{\dot B}}{\partial q^{\dot B}}p_{\dot A}$$
$$+\partial_{\dot A}\left\{L^{\dot C}B_{\dot C}-\frac{R}{4}\Theta+\frac{1}{3}(L_{\dot C}\,p^{\dot C})^2+2T_0L_{\dot C}\,p^{\dot C}-\frac{1}{2}\frac{\partial L_{\dot C}}{\partial q^{\dot B}}p^{\dot B}p^{\dot C}\right\}.$$

Therefore,

$$\text{II: } X_{\dot A}-2D_{\dot A}T_0+2T_0(\partial^{\dot C}Q_{\dot A\dot C})+12\phi T_0^2J_{\dot A}+3\mu\phi^{\,2}G_{\dot A}=\frac{1}{2}K^{\dot B}\frac{\partial\mu}{\partial q^{\dot B}}p_{\dot A}$$
$$+\partial_{\dot A}\left\{\frac{1}{2}\phi^{\,-2}Q^{\dot B\dot C}Q_{\dot B\dot C}-\phi\frac{\partial A^{\dot C}}{\partial q^{\dot C}}+(\phi J_{\dot C}B^{\dot C})^2+2(T_0\phi)^2+\frac{R}{12}K^{\dot B}B_{\dot B}\right.$$
$$\left.+\mu\phi^{\,3}K^{\dot B}B_{\dot B}-3\mu W+\frac{1}{2}K^{\dot B}p_{\dot B}\frac{\partial\mu}{\partial q^{\dot C}}p^{\dot C}\right\} \tag{3.30a}$$

and

$$\text{I: } X_{\dot A}-2D_{\dot A}T_0+2T_0(\partial^{\dot C}Q_{\dot A\dot C})-\frac{R}{4}G_{\dot A}=-\frac{1}{2}\frac{\partial L^{\dot B}}{\partial q^{\dot B}}p_{\dot A}+\partial_{\dot A}\left\{\frac{1}{2}Q^{\dot B\dot C}Q_{\dot B\dot C}\right.$$
$$\left.-\frac{\partial B^{\dot C}}{\partial q^{\dot C}}+2T_0^2-\frac{R}{4}B_{\dot C}\,p^{\dot C}+\frac{R}{4}\Theta-L^{\dot C}B_{\dot C}+\frac{1}{6}(L_{\dot C}\,p^{\dot C})^2-\frac{1}{6}\frac{\partial L_{\dot C}}{\partial q^{\dot B}}p^{\dot B}p^{\dot C}\right\}. \tag{3.30b}$$

On the other hand, under the assumptions already made regarding the energy-momentum tensor (i.e., $T_\mu{}^\mu=\text{const}$; $T_{11\dot A\dot B}=0$), it follows that the components $T_{22\dot C\dot D}$ are of the form

$$T_{22\dot C\dot D}=-(1/8\pi)\phi^{\,6}\partial_{(\dot C}T_{\dot D)}; \tag{3.31}$$

hence, from Einstein's equations and Eq. (3.7) we conclude that

$$X_{\dot A}=T_{\dot A}+\alpha\,p_{\dot A}+N_{\dot A}, \tag{3.32}$$

where α and $N_{\dot A}$ are functions of $q^{\dot R}$ only.

Finally, expressing $Q^{\dot B\dot C}Q_{\dot B\dot C}$ in terms of $B_{\dot B}$ and from Eqs. (3.30) and (3.32) we find that

$$\text{II: } \partial_{\dot A}\left\{\phi D^{\dot C}B_{\dot C}-\frac{1}{2}\phi^{\,4}\partial^{\dot B}B^{\dot C}\partial_{(\dot B}B_{\dot C)}+(\phi J^{\dot C}B_{\dot C})^2+\left(\frac{R}{12}+\mu\phi^{\,3}\right)K^{\dot B}B_{\dot B}\right.$$
$$\left.+\frac{1}{2}K^{\dot B}p_{\dot B}[K^{\dot D}p_{\dot D}J^{\dot C}-(\phi+\kappa)K^{\dot C}]\frac{\partial\mu}{\partial q^{\dot C}}+2(T_0\phi)^2-3\mu W-N_{\dot B}\,p^{\dot B}\right\}$$
$$=T_{\dot A}-2D_{\dot A}T_0+2T_0(\partial^{\dot C}Q_{\dot A\dot C})+12\phi T_0^2J_{\dot A}+3\mu\phi^{\,2}G_{\dot A}+\left(\alpha-\frac{1}{2}K^{\dot B}\frac{\partial\mu}{\partial q^{\dot B}}\right)p_{\dot A} \tag{3.33a}$$

and

$$\text{I: } \partial_{\dot A}\left\{D^{\dot C}B_{\dot C}-\frac{1}{2}\partial^{\dot B}B^{\dot C}\partial_{(\dot B}B_{\dot C)}-\left(L^{\dot C}+\frac{R}{4}p^{\dot C}\right)B_{\dot C}+\frac{R}{4}\Theta\right.$$
$$\left.+\frac{1}{18}(L_{\dot B}\,p^{\dot B})^2-\frac{1}{6}\frac{\partial L_{\dot C}}{\partial q^{\dot B}}p^{\dot B}p^{\dot C}+2T_0^2-N_{\dot B}\,p^{\dot B}\right\}$$
$$=T_{\dot A}-2D_{\dot A}T_0+2T_0(\partial^{\dot C}Q_{\dot A\dot C})-\frac{R}{4}G_{\dot A}+\left(\alpha+\frac{1}{2}\frac{\partial L^{\dot B}}{\partial q^{\dot B}}\right)p_{\dot A}. \tag{3.33b}$$

The function $\alpha(q^{\dot R})$ is determined by the integrability condition of Eqs. (3.33) which gives

$$\text{II: } \alpha=\frac{1}{2}K^{\dot A}\frac{\partial\mu}{\partial q^{\dot A}}+\frac{1}{2}\partial^{\dot A}T_{\dot A}+D^{\dot A}\partial_{\dot A}T_0 \tag{3.34a}$$

and

$$\text{I: } \alpha=-\frac{1}{2}\frac{\partial L^{\dot A}}{\partial q^{\dot A}}+\frac{1}{2}\partial^{\dot A}T_{\dot A}+D^{\dot A}\partial_{\dot A}T_0. \tag{3.34b}$$

From Eqs. (2.21a), (3.11), (3.32), and (3.34) one obtains the components of the left conformal curvature as

II: $C^{(5)} = C^{(4)} = 0$,

$$C^{(3)} = -2\phi^{3}(\mu + 2J^{\dot{A}}\partial_{\dot{A}}T_0),$$

$$C^{(2)} = 2\phi^{5}\Big\{J^{\dot{A}}N_{\dot{A}} - (p^{\dot{A}} + \frac{1}{2}\kappa K^{\dot{A}})\frac{\partial\mu}{\partial q^{\dot{A}}} - 2J^{\dot{A}}\frac{\partial T_0}{\partial q^{\dot{A}}} + J^{\dot{A}}T_{\dot{A}}$$
$$- \phi D^{\dot{A}}\partial_{\dot{A}}T_0 + J^{\dot{A}}p_{\dot{A}}\Big(\frac{1}{2}\partial^{\dot{B}}T_{\dot{B}} + D^{\dot{B}}\partial_{\dot{B}}T_0\Big)\Big\},$$

$$C^{(1)} = 2\phi^{6}[2\phi^{2}J^{\dot{C}}(\partial_{\dot{C}}B^{\dot{A}}) - 2\phi^{2}(\partial^{\dot{A}}T_0) + 3\mu\phi^{2}K^{\dot{A}} + D^{\dot{A}}]\Big\{\phi^{2}\Big[T_{\dot{A}} + N_{\dot{A}}$$
$$+ \Big(\frac{1}{2}K^{\dot{B}}\frac{\partial\mu}{\partial q^{\dot{B}}} + \frac{1}{2}\partial^{\dot{B}}T_{\dot{B}} + D^{\dot{B}}\partial_{\dot{B}}T_0\Big)p_{\dot{A}}\Big] + J^{\dot{B}}\frac{\partial B_{\dot{A}}}{\partial q^{\dot{B}}} + \phi^{2}K_{\dot{A}}K^{\dot{B}}p_{\dot{B}}J^{\dot{C}}\frac{\partial\mu}{\partial q^{\dot{C}}}\Big\},$$

and

I: $C^{(5)} = C^{(4)} = 0$,

$$C^{(3)} = \frac{R}{6},$$

$$C^{(2)} = \frac{\partial L^{\dot{A}}}{\partial q^{\dot{A}}} + 2D_{\dot{A}}\partial^{\dot{A}}T_0,$$

$$C^{(1)} = -2\Big[L^{\dot{A}} + \frac{R}{4}p^{\dot{A}} + 2(\partial^{\dot{A}}T_0) - D^{\dot{A}}\Big]\Big\{T_{\dot{A}} + N_{\dot{A}}$$
$$+ \Big(-\frac{1}{2}\frac{\partial L^{\dot{B}}}{\partial q^{\dot{B}}} + \frac{1}{2}\partial^{\dot{B}}T_{\dot{B}} + D^{\dot{B}}\partial_{\dot{B}}T_0\Big)p_{\dot{A}}\Big\}.$$

From Eqs. (2.21b) and (3.27) one finds that the spinorial components of the anti-self-dual part of the conformal curvature are given by

II: $C_{\dot{A}\dot{B}\dot{C}\dot{D}} = \phi^{3}\partial_{(\dot{A}}\partial_{\dot{B}}\partial_{\dot{C}}\phi^{2}[B_{\dot{D})} - \mu K_{\dot{D})}K_{\dot{R}}p^{\dot{R}}]$,

I: $C_{\dot{A}\dot{B}\dot{C}\dot{D}} = \partial_{(\dot{A}}\partial_{\dot{B}}\partial_{\dot{C}}B_{\dot{D})}$.

The results obtained up to now can be summarized as follows. If in a space-time which admits a null string defined by a spinor l_A there is a distribution of matter such that the trace of its energy-momentum tensor is constant and $l^{A}l^{B}T_{A\dot{B}C\dot{D}} = 0$, then Einstein's field equations can be reduced to one second-order partial differential equation with quadratic nonlinearities which together with the dynamical equations for the matter field gives a set of coupled differential equations.

In the case of vacuum, the existence of a null string is equivalent to the algebraic degeneracy of the self-dual or anti-self-dual part of the conformal curvature.[2] In this case $T_0 = T_{\dot{A}} = 0$ and one can take $G_{\dot{A}} = 0$ (any other choice amounts to a redefinition of W or Θ). Setting $R = -4\lambda$, where λ is possibly a nonzero cosmological constant, and inserting these values in Eqs. (3.33) one obtains the hyperheavenly equation (cf. Refs. 1 and 3).

As was shown in Ref. 4, this approach allows one to reduce the Einstein–Maxwell equations to two differential conditions provided that the space-time admits a null string and that the spinorial components of the self-dual part of the electromagnetic field satisfy the condition $l^{A}l^{B}f_{AB} = 0$ where l^{A} is a spinor which defines the null string. Since the energy-momentum tensor of the electromagnetic field is traceless and $T_{A\dot{B}C\dot{D}} = -(1/\pi)f_{AB}f_{\dot{C}\dot{D}}$ where f_{AB} and $f_{\dot{A}\dot{B}}$ denote the spinorial components of the self-dual and anti-self-dual part of the electromagnetic field, respectively, the condition $l^{A}l^{B}f_{AB} = 0$ implies $l^{A}l^{B}T_{A\dot{B}C\dot{D}} = 0$.

4. MASSLESS SPINOR FIELDS

In an algebraically degenerate complex space-time which admits a null string the solution to the usual spinor field equations for a massless field can be expressed in terms of one potential which has to satisfy a wavelike equation.

Let $\Psi_{A_1\cdots A_{2s}}$ be a symmetric spinor in its $2s$ indices which satisfies the field equations

$$\nabla_{B}{}^{\dot{R}}\Psi_{\dot{R}\dot{A}_1\cdots\dot{A}_{2s-1}} = 0. \tag{4.1}$$

With $B = 1$ and using Eq. (2.20) one obtains $\phi^{s+1}\partial^{\dot{R}}(\phi^{-s-1}\Psi_{\dot{R}\dot{A}_1\cdots\dot{A}_{2s-1}}) = 0$ which implies the existence of a potential H such that

$$\Psi_{A_1\cdots A_{2s}} = \phi^{s+1}\partial_{A_1}\cdots\partial_{A_{2s}}H. \tag{4.2}$$

Using Eqs. (4.2) and (2.20), Eq. (4.1) with $B = 2$ implies that H must satisfy the wavelike condition

$$D^{\dot{B}}\partial_{\dot{B}}\partial_{\dot{A}_1}\cdots\partial_{\dot{A}_{2s-1}}H + (s-1)(\partial_{\dot{B}}Q^{\dot{B}\dot{C}})\partial_{\dot{C}}\partial_{\dot{A}_1}\cdots\partial_{\dot{A}_{2s-1}}H$$
$$- (2s-1)(\partial_{(\dot{A}_1}Q^{\dot{B}\dot{C}})\partial_{|\dot{B}}\partial_{\dot{C}|}\partial_{\dot{A}_2}\cdots\partial_{\dot{A}_{2s-1})}H = 0. \tag{4.3a}$$

For $s \geqslant 1$ one has the equivalent equation

$$\partial_{(\dot{A}_1}\big[D^{\dot{B}}\partial_{|\dot{B}|}\partial_{\dot{A}_2}\cdots\partial_{\dot{A}_{2s-1})}H\big] + (s-1)(\partial_{\dot{B}}Q^{\dot{B}\dot{C}})\partial_{\dot{C}}\partial_{\dot{A}_1}\cdots\partial_{\dot{A}_{2s-1}}H - 2(s-1)(\partial_{(\dot{A}_1}Q^{\dot{B}\dot{C}})\partial_{|\dot{B}}\partial_{\dot{C}|}\partial_{\dot{A}_2}\cdots\partial_{\dot{A}_{2s-1})}H = 0. \tag{4.3b}$$

This result generalizes the one of Finley and Plebański[8] for the case of Heavens and it also contains as a special case the one for flat space-time obtained by Penrose.[9]

Similarly, if $C_{12\dot{A}\dot{B}} = 0 = C_{11\dot{A}\dot{B}}$, from Eqs. (3.1b) and (2.20) one finds that $\phi^{3}\partial^{S}(\phi^{-3}C_{S\dot{A}\dot{B}\dot{C}}) = 0$ which implies that $C_{A\dot{B}C\dot{D}} = \phi^{3}\partial_{A}\partial_{\dot{B}}\partial_{C}\partial_{\dot{D}}\tilde{W}$ where $\tilde{W}$ is a function. If, additionally, $C_{22\dot{A}\dot{B}} = 0$ when W satisfies Eqs. (4.3) with $s = 2$.

5. EINSTEIN–WEYL EQUATIONS

The previous results are now applied to reduce the Einstein–Weyl equations when the space-time admits a null string and by suitably restricting the Weyl field. The Weyl's equations for the two-component neutrino are

$$\nabla^{A}{}_{\dot{B}}\Psi_{A}=0, \tag{5.1a}$$

$$\nabla_{A}{}^{\dot{B}}\Psi_{\dot{B}}=0. \tag{5.1b}$$

The energy-momentum tensor of the neutrino field is traceless and its spinorial components are

$$T_{AB\dot{C}\dot{D}}=(i\hbar/16)[\Psi_{A}\nabla_{B\dot{D}}\Psi_{\dot{C}}+\Psi_{B}\nabla_{A\dot{C}}\Psi_{\dot{D}}-\Psi_{\dot{C}}\nabla_{B\dot{D}}\Psi_{A}-\Psi_{\dot{D}}\nabla_{A\dot{C}}\Psi_{B}]. \tag{5.2}$$

We can easily include a cosmological constant writing $R=-4\lambda$.

Using Eqs. (5.2) and (2.20) one obtains

$$T_{11\dot{C}\dot{D}}=(i\hbar/8)[\Psi_{1}\nabla_{1(\dot{C}}\Psi_{\dot{D})}-\sqrt{2}\phi^{-3/2}\Psi_{(\dot{C}}\partial_{\dot{D})}(\phi^{3/2}\Psi_{1})];$$

therefore, in order to have $T_{11\dot{C}\dot{D}}=0$, we take $\Psi_{1}=0$ (though this is not the only way to make $T_{11\dot{C}\dot{D}}=0$). With this assumption Eq. (5.1a) gives $\partial_{\dot{B}}(\phi^{-5/2}\Psi_{2})=0$ which implies the existence of a function $\zeta=\zeta(q^{\dot{R}})$ such that

$$\Psi_{2}=(\sqrt{2}/i\hbar\pi)\phi^{5/2}\zeta. \tag{5.3}$$

On the other hand, Eq. (5.1b) implies that Ψ_{A} is of the form

$$\Psi_{A}=\tfrac{1}{4}\phi^{3/2}\partial_{A}H, \tag{5.4}$$

where H is a solution of the equation [see Eq. (4.3a)]

$$D^{\dot{B}}\partial_{\dot{B}}H-\tfrac{1}{2}(\partial_{B}Q^{\dot{B}\dot{C}})\partial_{\dot{C}}H=0. \tag{5.5}$$

From Eqs. (5.2), (5.3), and (5.4) one finds that $T_{12\dot{C}\dot{D}}=(1/32\pi)\phi^{4}\zeta\partial_{\dot{C}}\partial_{\dot{D}}H$ and $T_{22\dot{C}\dot{D}}=(1/16\pi)\phi^{6}\partial_{(\dot{C}}[\zeta D_{\dot{D})}H-HD_{\dot{D})}\zeta]$. Comparing with Eqs. (3.19) and (3.31), one obtains

$$T_{0}=\tfrac{1}{4}\zeta H,\quad T_{A}=\tfrac{1}{2}(HD_{A}\zeta-\zeta D_{A}H). \tag{5.6}$$

Commuting $\partial_{\dot{B}}$ and $D^{\dot{B}}$ Eq. (5.5) gives

$$0=\partial_{\dot{B}}(D^{\dot{B}}H+\tfrac{1}{2}H\partial_{\dot{C}}Q^{\dot{B}\dot{C}})-\tfrac{1}{2}H\partial_{\dot{B}}\partial_{\dot{C}}Q^{\dot{B}\dot{C}},$$

which, due to Eqs. (3.21), (3.15b), and (5.6), is equivalent to

$$\text{II:}\quad 0=\partial_{\dot{B}}(D^{\dot{B}}H+\tfrac{1}{2}H\partial_{\dot{C}}Q^{\dot{B}\dot{C}}-3(\mu/\zeta)\phi^{2}G^{\dot{B}}-\tfrac{3}{4}\zeta\phi H^{2}J^{\dot{B}}),$$

$$\text{I:}\quad 0=\partial_{\dot{B}}(D^{\dot{B}}H+\tfrac{1}{2}H\partial_{\dot{C}}Q^{\dot{B}\dot{C}}-(\lambda/\zeta)G^{\dot{B}}),$$

where $G^{\dot{B}}$ satisfies $\partial_{\dot{B}}(\phi^{2}G^{\dot{B}})=\phi\zeta H$. This implies that there exists a function χ such that

$$\text{II:}\quad \partial_{\dot{B}}\chi=D_{\dot{B}}H-\tfrac{1}{2}H\partial^{\dot{C}}Q_{\dot{B}\dot{C}}-3(\mu/\zeta)\phi^{2}G_{\dot{B}}-\tfrac{3}{4}\zeta\phi H^{2}J_{\dot{B}},$$

$$\text{I:}\quad \partial_{\dot{B}}\chi=D_{\dot{B}}H-\tfrac{1}{2}H\partial^{\dot{C}}Q_{\dot{B}\dot{C}}-(\lambda/\zeta)G_{\dot{B}}.$$

Turning the attention to the (reduced) Einstein's equations (3.33) one notes, firstly, that $\tfrac{1}{2}\partial^{A}T_{A}+D^{A}\partial_{A}T_{0}=\tfrac{1}{4}\zeta(D^{A}\partial_{A}H+\partial_{A}D^{A}H)=\tfrac{1}{2}\zeta[D^{A}\partial_{A}H-\tfrac{1}{2}(\partial_{A}Q^{A\dot{B}})\partial_{\dot{B}}H]$ which is zero by Eq. (5.5). Thus, in case II $\alpha=\tfrac{1}{2}K^{A}\partial\mu/\partial q^{A}$, and in I $\alpha=-\tfrac{1}{2}\partial L^{A}/\partial q^{A}$. Therefore, by Eq. (5.6), the right-hand sides of Eqs. (3.33) are

$$\text{II:}\quad T_{A}-2D_{A}T_{0}+2T_{0}(\partial^{\dot{C}}Q_{A\dot{C}})+12\phi T_{0}^{2}J_{A}+3\mu\phi^{2}G_{A}=-\zeta(D_{A}H-\tfrac{1}{2}H\partial^{\dot{C}}Q_{A\dot{C}}-\tfrac{3}{4}\zeta\phi H^{2}J_{A}-(3\mu/\zeta)\phi^{2}G_{A})=-\zeta\partial_{A}\chi,$$

$$\text{I:}\quad T_{A}-2D_{A}T_{0}+2T_{0}(\partial^{\dot{C}}Q_{A\dot{C}})+\lambda G_{A}=-\zeta(D_{A}H-\tfrac{1}{2}H\partial^{\dot{C}}Q_{A\dot{C}}-(\lambda/\zeta)G_{A})=-\zeta\partial_{A}\chi.$$

Inserting these expressions into Eqs. (3.33) and defining $\chi'\equiv\chi+\tfrac{1}{8}\zeta\phi^{2}H^{2}-3(\mu/\zeta)W$ for case II and $\chi'\equiv\chi+\tfrac{1}{8}\zeta H^{2}-(\lambda/\zeta)\Theta$ for case I, one obtains the pair of equations to which the Einstein–Weyl equations have been reduced; these equations are

$$\text{II:}\quad \phi D^{\dot{C}}B_{\dot{C}}-\frac{1}{2}\phi^{4}\partial^{\dot{B}}B^{\dot{C}}\partial_{(\dot{B}}B_{\dot{C})}+(\phi J^{\dot{C}}B_{\dot{C}})^{2}+(\mu\phi^{3}-\lambda/3)K^{\dot{B}}B_{\dot{B}}+\frac{1}{2}K^{\dot{B}}p_{\dot{B}}[K^{\dot{D}}p_{\dot{D}}J^{\dot{C}}-(\phi+\kappa)K^{\dot{C}}]\times\frac{\partial\mu}{\partial q^{\dot{C}}}-N_{\dot{B}}p^{\dot{B}}=-\zeta\chi',$$

$$\partial_{\dot{B}}\chi'=D_{\dot{B}}H+3\mu\phi^{2}\left(\frac{1}{2}HK_{\dot{B}}-\frac{B_{\dot{B}}}{\zeta}\right)+HJ^{\dot{C}}\partial_{(\dot{B}}\phi^{2}B_{\dot{C})}$$

$$\text{I:}\quad D^{\dot{C}}B_{\dot{C}}-\frac{1}{2}\partial^{\dot{B}}B^{\dot{C}}\partial_{(\dot{B}}B_{\dot{C})}-(L^{\dot{C}}-\lambda p^{\dot{C}})B_{\dot{C}}+\frac{1}{18}(L_{\dot{B}}p^{\dot{B}})^{2}-\frac{1}{6}\frac{\partial L_{\dot{C}}}{\partial q^{\dot{B}}}p^{\dot{B}}p^{\dot{C}}-N_{\dot{B}}p^{\dot{B}}=-\zeta\chi',$$

$$\partial_{\dot{B}}\chi'=D_{\dot{B}}H-\frac{1}{2}H(L_{\dot{B}}-\lambda p_{\dot{B}})-\frac{\lambda}{\zeta}B_{\dot{B}}.$$

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor J. F. Plebański for many helpful suggestions and comments. He also acknowledges the Consejo Nacional de Ciencia y Tecnología for financial assitance through a doctoral fellowship.

[1] J. F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976). The details can be found in *Asymptotic Structure of Space-Time*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).

[2] J. F. Plebański and S. Hacyan, J. Math. Phys. **16**, 2403 (1975).

[3] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **17**, 2207 (1976).

[4] 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)

[5] J. F. Plebański, J. Math. Phys. **16**, 2395 (1975).

[6] This result, originally presented in Ref. 1, was independently obtained by J. F. Plebański and K. Rózga using a different approach.

[7] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **22**, 667 (1981).

[8] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **17**, 585 (1976).

[9] R. Penrose, Proc. R. Soc. London. Ser. A **284**, 159 (1965). See also J. F. Plebański and I. Robinson, J. Math. Phys. **19**, 2350 (1978).

Timelike curves of limited acceleration in general relativity[a)]

Sandip K. Chakrabarti, Robert Geroch, and Can-bin Liang
Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637

(Received 20 July 1982; accepted for publication 6 October 1982)

A simple result, restricting the behavior of a timelike curve under limitations on its integrated acceleration, is obtained and discussed.

PACS numbers: 04.20. − q, 02.40. + m

There are numerous examples of space-times in general relativity in which various feats can, at least in principle, be performed by certain observers (arbitrary timelike curves), but not by those freely falling (timelike geodesics). For example, in the Reissner–Nordström solution,[1] timelike curves, but no timelike geodesics, can reach arbitrarily small r-values; in the Gödel solution,[2] there are closed timelike curves, but no closed timelike geodesics; in Minkowski space-time, timelike curves, but no timelike geodesics, can reach points of future null infinity. But what is possible in practice in such examples? A real observer, after all, is presumably incapable of following an arbitrary timelike curve, and yet certainly has more maneuverability than afforded by a geodesic. Thus, in order to analyze physical effects in such examples, it is natural to consider the behavior of timelike curves intermediate between these extremes.

A class of physically reasonable observers consists of those described by timelike curves γ with finite integrated acceleration, i.e., with $\int_\gamma a\, ds < \infty$, where s is length along γ and a is the norm of the acceleration, $a^b = \xi^c \nabla_c \xi^b$, where ξ^b is the unit tangent to γ. Indeed, this condition is motivated by the following remark: Timelike curve γ has a finite integrated acceleration if and only if there exists on γ a positive function m, bounded away from zero, such that

$$J^b = -\xi^c \nabla_c (m\xi^b) \tag{1}$$

is future-directed timelike or null everywhere on γ. Physically, m represents the mass of some rocket ship describing the curve, whence the J^a given by (1) is the rate of discharge of energy–momentum by the ship's exhaust. That m be bounded away from zero requires that the ship carry finite payload, while that J^a be future-directed requires that the exhaust satisfy the energy condition. In this sense, then, timelike curves of finite integrated acceleration are just those which can be traversed by physically realistic rocket ships. To prove the remark, let $m > 0$ be given along γ, and expand (1) to obtain: J^a is future-directed timelike or null if and only if

$$a \leqslant -\xi^b \nabla_b (\ln m). \tag{2}$$

Now integrate.

We here derive and apply a simple relation between the integrated acceleration of a timelike curve and the explorations available to it.

Let M, g_{ab} be a space-time, γ a timelike curve and t^a a Killing field, timelike or null everywhere on γ. Introduce the effective energy, $E = -\xi^b t_b$, so, e.g., E would be constant along γ were it a geodesic. The basic inequality is the following:

$$|\xi^b \nabla_b E| \leqslant aE. \tag{3}$$

To prove this, expand the left side:

$$\begin{aligned} |\xi^b \nabla_b E| &= |-a_b t^b| = |a_b t_c h^{bc}| \\ &\leqslant (a_b a_c h^{bc})^{1/2} (t_m t_n h^{mn})^{1/2} \\ &= a(t^b t_b + E^2)^{1/2} \leqslant aE, \end{aligned}$$

where we introduced $h^{ab} = g^{ab} + \xi^a \xi^b$, the (positive and semi-definite) projection orthogonal to ξ^a, in the second step, and used $t^b t_b \leqslant 0$ in the last. Some examples of the application of (3) follow.

Integrating (3) along the curve, it follows that, if γ has finite integrated acceleration, then E remains finite along the curve. But $E = -\xi^b t_b \geqslant (-t^b t_b)^{1/2}$, and so this right side must also remain finite. In the Reissner–Nordström solution, $(-t^b t_b) = (1 - 2m/r + q^2/r^2)$. We conclude that no timelike curve of finite integrated acceleration in this solution can reach arbitrarily small r-values.[3] In this sense, then, the singular region of Reissner–Nordström is physically inaccessible. A similar remark holds for certain negative-mass Weyl solutions,[4] including the Schwarzschild solution (with $m < 0$) and its plane-symmetric limit.[5]

Recall that the Gödel space-time[2] is given by the metric product $S \times \mathbb{R}$, where $\mathbb{R}$ is the real line with its usual positive-definite metric, and $S = \mathbb{R}^3$ has metric

$$h^{ab} = \bar{h}^{ab} + t^a t^b. \tag{4}$$

Here,[6] $\bar{h}^{ab}$ is a complete metric on S of signature $(-, +, +)$ and constant negative curvature, and t^a is an $\bar{h}$-Killing field of constant squared norm $-1/2$. Thus, h^{ab} results from $\bar{h}^{ab}$ by "opening out the light cones about t^a." This t^a—a unit timelike Killing field for h^{ab}—gives the four velocities of the galaxies in the Gödel universe. Now let γ be a closed timelike curve in this space-time, with initial tangent $\xi^a = t^a$. Then $E = 1$ intially. Since the metric $\bar{h}^{ab}$ has no closed timelike curves, the projection of the tangent ξ^a into S must somewhere be $\bar{h}$-space-like, i.e., from (4), we must somewhere on γ have $E > \sqrt{2}$. Integrating (3) along γ, we now conclude that the integrated acceleration of γ must be at least $\ln \sqrt{2}$. Thus, no timelike curve with integrated acceleration less than this value and with initial tangent t^a can return to its initial point. This means physically that certain rocket ships (including, by Eq. (2), all those with less than about 29% of their initial mass as fuel), launched initially at rest from a galaxy in the Gödel universe, cannot return to the event of their launch. It is apparently not known whether the condition of zero initial speed can be dropped.

Finally, let, in Minkowksi space-time, γ be a timelike curve which reaches a point of future null infinity $\mathscr{I}^+$.

[a)] Supported in part by the NSF.

 0022-2488/83/030597-02$02.50

Choosing for t^a a boost Killing field such that $-t^b t_b$ approaches infinity along γ, and again integrating (3), we conclude that this curve must have infinite integrated acceleration. As expected, physical rocket ships cannot reach null infinity in flat space-time.

It is curious that this same result holds also for asymptotically flat space-times. In this case, the result is not so obvious, for it is conceivable that the rocket ship could work its way out to $\mathscr{I}^+$ by riding gravitational waves, using its limited ability to accelerate to reach a succession of waves. It is a question of whether the ship can thereby compensate for the fact that, as required by asymptotic flatness, each wave dies out asymptotically.

Let M, g_{ab} be an asymptotically flat space-time,[7] with future null infinity $\mathscr{I}^+$, conformal factor Ω vanishing at $\mathscr{I}^+$, and unphysical metric $\Omega^2 g_{ab}$ smooth and invertible at $\mathscr{I}^+$. Let, for contradiction, γ be a timelike curve with finite integrated acceleration which reaches a point p of $\mathscr{I}^+$. Let t^a be a BMS generator[8] corresponding to the boost Killing field used in the flat case. Then[9] (i) $X_{ab} = \Omega \nabla_{(a} t_{b)}$ and (ii) $\Omega^2 t_a$ must remain smooth at $\mathscr{I}^+$ (by definition of a BMS generator), (iii) $\Omega t^b t_b$ must have a nonzero limit along γ (by our choice of t^a), and (iv) $X_{ab} t^b$ must vanish at p (since t^a, being timelike along γ, must be a multiple of the null generator of $\mathscr{I}^+$ at p, while X_{ab} is transverse to this generator). We now have

$$\begin{aligned} |\xi^b \nabla_b E| &\leqslant aE + \Omega^{-1} |\xi^a \xi^b X_{ab}| \\ &\leqslant aE + \Omega^{-1} \left(\xi^b \xi^c \overset{+}{g}_{bc} \right) \left(X_{mn} X_{pq} \overset{+}{g}^{mp} \overset{+}{g}^{nq} \right)^{1/2} \\ &\leqslant aE + c\Omega^2 E^2, \end{aligned} \tag{5}$$

where c is a constant. In the first step, we have modified (3) to take into account that t^a is no longer a Killing field; in the second, we have introduced positive-definite metric

$$\overset{+}{g}^{ab} = g^{ab} - 2(t^m t_m)^{-1} t^a t^b, \tag{6}$$

with inverse $\overset{+}{g}_{ab}$, along γ; and in the third, we have used the fact that $(\xi^b \xi^c \overset{+}{g}_{bc})$ is bounded by a multiple of ΩE^2 [from (6) and (iii)] and that $(X_{mn} X_{pq} \overset{+}{g}^{mp} \overset{+}{g}^{nq})$ is bounded by a multiple of Ω^4 [from (i), (iv), and (iii), and the fact that the physical metric is smooth at $\mathscr{I}^+$]. Now divide (5) by E and integrate along γ. Then the first term on the right, the integrated acceleration of γ, is finite by assumption; while the second term, a multiple of the line integral on γ of $\Omega^2 t_b$, is finite[10] by (ii). We conclude that E must remain bounded along γ, which, since $E \geqslant (-t^b t_b)^{1/2}$, contradicts (iii).

A similar argument [including, and bounding by $c\Omega^2 E^2$, the extra term $F_{ab} \xi^a t^b$ on the right in (5)] shows: In an asymptotically flat space-time with asymptotically regular Maxwell field, no charged rocket ship, with finite integrated nonelectromagnetic acceleration, can reach $\mathscr{I}^+$.

ACKNOWLEDGMENT

We wish to thank Robert Wald for stimulating discussions.

[1] See, for example, S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973), p. 159.

[2] K. Gödel, Rev. Mod. Phys. **21**, 447 (1949); S. Chandrasekhar and J. P. Wright, Proc. Natl. Acad. Sci. **47**, 341 (1961).

[3] The same conclusion holds for charged-particle trajectories with finite integrated nonelectromagnetic acceleration, provided $q\kappa > -|q|$ where κ is the charge/mass ratio of the charged particle, and q is the Reissner–Nordström charge. To see this, modify (3) for this case to read $|\xi^b \nabla_b (E - \kappa\phi)| \leqslant aE$, where ϕ, the electric potential, is determined by $\nabla_a \phi = F_{ab} t^b$. For Reissner–Nordström, $\phi = -q/r$, and $E \geqslant (1 - 2m/r + q^2/r^2)^{1/2} \sim |q|/r$. So, under the above condition on the charge, the aE on the right is bounded by a multiple of a $(E - \kappa\phi)$. The argument now goes through as before.

[4] See, for example, J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1964), p. 309.

[5] See, for example, A. Das, J. Math. Phys. **12**, 1136 (1971).

[6] These objects may be constructed as follows. Let V be a four-dimensional vector space with metric G_{ab} of signature $(-,-,+,+)$. Consider the locus of vectors x^a in V with $x^a x^b G_{ab} = -1$. Then S is the covering space of this three-submanifold of V, while $\bar{h}^{ab}$ is the metric induced from G^{ab}. The Killing fields of $(S, \bar{h}^{ab})$ are the generators of G-preserving linear maps on V, and so are described by antisymmetric tensors over V. In particular, t^b is described by $2^{-1/2} F_a{}^b$, with $F_a{}^m G_{bm}$ antisymmetric, and (by norm-condition on t^b) *and* $F_a{}^m F_m{}^b = -\delta_a{}^b$. (Such F's exist in this signature.) So, F defines a complex structure on V, and so we may regard V as a complex two-dimensional vector space, on which G_{ab} becomes a Hermitian metric of signature $(-,+)$. The Killing fields of (S, h^{ab}) are the generators of linear maps on V which preserve G and F; i.e., the generators of complex-linear maps on the complex vector space which preserve its Hermitian metric. So, for example, the Killing fields of the full Gödel space-time form the Lie algebra of $U(1,1) \times R$.

[7] R. Penrose, in *Relativity, Groups and Topology*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964); R. Geroch, in *Asymptotic Structure of Space-Time*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).

[8] H. Bondi, M. G. J. Van der Burg, and A. W. K. Metzner, Proc. R. Soc. London, Ser. A **270**, 103 (1962); R. Sachs, Phys. Rev. **128**, 2851 (1962).

[9] R. Geroch and J. Winicour, J. Math. Phys. **22**, 803 (1981).

[10] To see this, let γ' be a smooth curve reaching p. Then the line integral of $\Omega^2 t_a$ on γ equals the line integral on γ' (which is finite), plus the integral of the curl of $\Omega^2 t_a$ on a two-surface connecting γ and γ' (which is also finite).

Orthogonal 2 + 2 decomposition of space-time

J. Smallwood[a)]
Department of Engineering Mathematics, University of Newcastle upon Tyne, NE1 7RU, England

(Received 4 December 1981; accepted for publication 8 January 1982)

A formalism is developed describing the most general 2 + 2 decomposition of space-time into orthogonal spacelike and timelike 2-surface elements $\{S\}$ and $\{T\}$. No restriction is made on the holonomicity of $\{S\}$ or $\{T\}$, and the formalism is 2-covariant in both $\{S\}$ and $\{T\}$. As an example of the utility of this formalism, the anholonomic 2 + 2 formulation of the initial value problems in relativity is discussed and the standard holonomic approach is briefly reviewed. Further applications of the formalism are indicated.

PACS numbers: 04.20. − q, 02.40.Ky

I. INTRODUCTION

In some recent work,[1,2] a 2 + 2 formalism was developed in which space-time is broken up into a foliation of spacelike 2-surfaces $\{S\}$ and the orthogonal timelike 2-surface elements $\{T\}$. Using this formalism, several initial value problems (henceforth referred to as IVP's), namely, the Cauchy, characteristic, and mixed (timelike–null) problems were analyzed. The main result of this work is that it shows that in each case, the gravitational degrees of freedom may be cast into the so-called conformal 2-structure[1-3]—essentially the conformal metric of $\{S\}$.

The formalism developed in Refs. 1 and 2 suffers from two limitations. Firstly, it is restricted to a breakup where $\{S\}$ consists necessarily of holonomic surfaces. Secondly, the formalism is covariant only in so far as arbitrary transformations are allowed among the dual bases of $\{S\}$. A particular dyad basis is introduced in $\{T\}$ from the beginning, and this destroys the covariance of objects defined in $\{T\}$. In the analysis of the IVP's, the introduction of a dyad basis into $\{T\}$ at some stage is almost undoubtedly a necessity, since this is equivalent to specifying the directions in which the initial data are propagated. However, it is of great advantage to do this only after the formalism is developed. It is mathematically more elegant, it avoids certain *ad hoc* definitions within $\{T\}$ (in particular, of some derivatives) previously necessary, since all operations in $\{T\}$ are now covariant, and it simplifies manipulation of the formalism.

In Sec. II, the most general covariant 2 + 2 formalism possible is developed. It removes the aforementioned limitations, in that, first, no restriction is placed on the holonomicity of $\{S\}$ and $\{T\}$, the only requirement being that they be mutually orthogonal. Second, all quantities are defined with respect to arbitrary dual bases on $\{S\}$ and $\{T\}$, hence their geometrical properties are immediately clear. If the condition of holonomicity is imposed on $\{S\}$, and a dyad basis introduced into $\{T\}$, then the present formalism reduces (aside from some minor changes in notation) to that developed in Refs. 1 and 2.

In Sec. III, the formalism is used to discuss the anholonimic 2 + 2 formulation of the IVP's (compare, for example, the work of Kulhanek,[4] Ó. Murchadha,[5] and Stachel[6] on the anholonomic 3 + 1 formulation of the Cauchy problem). In this formulation, the holonomicity condition is imposed on $\{T\}$. This allows one to pick as a basis in $\{T\}$ a pair of commuting vector fields. These vectors are the natural ones with which to propagate the field variables through space-time. The price one pays is that $\{S\}$ must now be allowed to be anholonomic, with the resulting increase in complexity of the definitions of quantities on $\{S\}$. The gravitational degrees of freedom in this case are taken to be embodied in the conformal metric of the 2-surface *elements* $\{S\}$.

The signature of space-time is taken to be $+\ -\ -\ -$. Greek indices run through 0 to 3. Early lower case Latin indices a to f run through 0, 1. Lower case Latin indices i to l run through the early lower case, then the upper case Latin indices, e.g., $i = a$; $A = 0, 1; 2, 3$.

II. THE FORMALISM[7]

Let V be a space-time with metric ${}^4\mathbf{g}$ and let $\mathbf{E}_\alpha$ be a general basis of vectors on V, with reciprocal basis $\mathbf{E}^\alpha$. Let $(\mathbf{E}_i, \mathbf{E}^i)$ be reciprocal bases on V, related to the general basis by

$$\mathbf{E}_i = \Phi_i{}^\alpha \mathbf{E}_\alpha, \quad \mathbf{E}^i = \Phi^i{}_\alpha \mathbf{E}^\alpha,$$

(where $\Phi_i{}^\alpha$, $\Phi^i{}_\alpha$ are the usual transformation matrices) and required to satisfy

$$\begin{aligned} &{}^4g(\mathbf{E}_A, \mathbf{E}_a) = 0 \qquad (2.1)\\ &|{}^4g(\mathbf{E}_A, \mathbf{E}_B)| > 0, \quad |{}^4g(\mathbf{E}_a, \mathbf{E}_b)| < 0. \end{aligned}$$

The above conditions imply that $(\mathbf{E}_A, \mathbf{E}^A)$ and $(\mathbf{E}_a, \mathbf{E}^a)$ are, respectively, reciprocal bases of sets of mutually orthogonal spacelike 2-surface elements $\{S\}$, and timelike 2-surface elements $\{T\}$. The bases of $\{S\}$ and $\{T\}$ are allowed to transform among themselves, but not among each other. That is, if $(\mathbf{E}_{A'}, \mathbf{E}^{A'})$ are new reciprocal bases of $\{S\}$ and $\{T\}$, then

$$\Phi_A{}^{\cdot a} = \Phi_{a'}{}^{\cdot A} = \Phi^{a'}{}_A = \Phi^{A'}{}_a = 0. \qquad (2.2)$$

In particular, transformations obeying (2.2) preserve (2.1). Equations (2.1) and (2.2) together define the most general orthogonal 2 + 2 (spacelike + timelike) breakup possible. The tensors

$$\mathbf{B} = \mathbf{E}_A \otimes \mathbf{E}^A, \quad {}^*\mathbf{B} = \mathbf{E}_a \otimes \mathbf{E}^a$$

are the unit tensors on $\{S\}$ and $\{T\}$, respectively, and act as projection operators into $\{S\}$ and $\{T\}$. Clearly,

a) Present address: Department of Applied Mathematics, Queen Mary College, London E1 4NS, England.

 0022-2488/83/030599-07$02.50

$$\mathbf{I} = \mathbf{B} + {}^*\mathbf{B},$$

where $\mathbf{I}$ is the unit tensor on V.

The projections of ${}^4\mathbf{g}$ onto $\{S\}$ and $\{T\}$ are their respective induced metrics and are denoted by $\mathbf{g}$ and ${}^*\mathbf{g}$. Their components with respect to $\mathbf{E}^i$ are given by

$$g_{AB} = {}^4g_{AB}, \tag{2.3a}$$

$${}^*g_{ab} = {}^4g_{ab}, \tag{2.3b}$$

and, since $\{S\}$ and $\{T\}$ are mutually orthogonal,

$${}^4\mathbf{g} = \mathbf{g} + {}^*\mathbf{g}. \tag{2.4}$$

The metrical covariant derivatives in V, $\{S\}$, and $\{T\}$ are denoted by ${}^4\nabla$, ∇, and ${}^*\nabla$, and the components of the metric connections in the bases $(\mathbf{E}_i, \mathbf{E}^i)$ are denoted by ${}^4\Gamma^i_{jk}$, Γ^A_{BC}, and ${}^*\Gamma^a_{bc}$, respectively.

A tensor ${}^4\nabla\mathbf{B} = -{}^4\nabla\,{}^*\mathbf{B}$ defines the curvature tensors of valence three, $\mathbf{H}$ and ${}^*\mathbf{H}$ of $\{S\}$ and $\{T\}$, via the equation

$${}^4\nabla_\gamma B^\alpha_\beta = H_{\gamma\beta}{}^\alpha + H_\gamma{}^\alpha{}_\beta - {}^*H_{\gamma\beta}{}^\alpha - {}^*H_\gamma{}^\alpha{}_\beta. \tag{2.5}$$

In terms of the bases $(\mathbf{E}_i, \mathbf{E}^i)$, it follows from (2.5) that

$$\mathbf{H} = H_{CB}{}^a\,\mathbf{E}^C \otimes \mathbf{E}^B \otimes \mathbf{E}_a, \quad {}^*\mathbf{H} = {}^*H_{cb}{}^A\,\mathbf{E}^c \otimes \mathbf{E}^b \otimes \mathbf{E}_A.$$

The vectors $\mathbf{E}_i$ do not in general commute, and the geometric object ${}^4\Omega$ is defined by

$$[\mathbf{E}_k, \mathbf{E}_j] = 2\,{}^4\Omega^i_{jk}\,\mathbf{E}_i,$$

which implies that

$${}^4\Gamma^i_{[jk]} = -{}^4\Omega^i_{jk}. \tag{2.6}$$

Similarly, ${}''\Omega$ and ${}^*{}''\Omega$ are defined by the equations

$${}''\Omega^A_{BC} = -\Gamma^A_{[BC]}, \tag{2.7a}$$

$${}^*{}''\Omega^a_{bc} = -{}^*\Gamma^a_{[bc]}. \tag{2.7b}$$

Now, ${}^4\Gamma^i_{jk}$ splits up into a number of separate geometrical objects with differing transformation properties under the allowable transformations (2.2). The following results can easily be proved.

$${}^4\Gamma^A_{BC} = \Gamma^A_{BC}, \tag{2.8a}$$

$${}^4\Gamma^a_{bc} = {}^*\Gamma^a_{bc}, \tag{2.8b}$$

as follows from (2.3) and (2.4), and

$${}^4\Gamma^a_{BC} = H_{BC}{}^a, \tag{2.9a}$$

$${}^4\Gamma^A_{bc} = {}^*H_{bc}{}^A, \tag{2.9b}$$

$${}^4\Gamma^A_{Bc} = -H_B{}^A{}_c, \tag{2.10a}$$

$${}^4\Gamma^a_{bC} = -{}^*H_b{}^a{}_C, \tag{2.10b}$$

which follow from (2.5). Equations (2.6) and (2.9) lead to the definitions of the tensors Ω and ${}^*\Omega$ by

$$\Omega_{BC}{}^a \equiv -H_{[BC]}{}^a, \tag{2.11a}$$

$${}^*\Omega_{bc}{}^A \equiv -{}^*H_{[bc]}{}^A. \tag{2.11b}$$

Ω and ${}^*\Omega$ are the objects of anholonomicity of $\{S\}$ and $\{T\}$; $\Omega = 0$ if and only if $\{S\}$ is holonomic, and similarly for ${}^*\Omega$. Two further geometric objects, ${}'\Omega$ and ${}^*{}'\Omega$ exhaust the contents of ${}^4\Gamma^i_{jk}$. They are defined by

$${}'\Omega^A_{Cb} = {}^4\Omega^A_{Cb}, \tag{2.12a}$$

$${}^*{}'\Omega^a_{cB} = {}^4\Omega^a_{cB}. \tag{2.12b}$$

${}'\Omega$ and ${}^*{}'\Omega$ transform as 1-forms of the indicated type in $\{T\}$ and $\{S\}$, respectively, but as more complicated objects in $\{S\}$ and $\{T\}$.

The above breakup of ${}^4\Gamma^i_{jk}$ leads to the following definitions of certain invariant differential operators. Let

$$\mathbf{v} = v^a{}_B\,\mathbf{E}_a \otimes \mathbf{E}^B$$

be an aribitrary tensor of the indicated type. The operator $\mathbf{D}$ is defined by

$$D_C\,v^a{}_B \equiv \delta^{iak}_{CjB}\,{}^4\nabla_i\,v^j{}_k = {}^4\nabla_C\,v^a{}_B. \tag{2.13}$$

Hence

$$D_C\,v^a{}_B = \partial_C\,v^a{}_B - {}^4\Gamma^E_{CB}\,v^a{}_E + {}^4\Gamma^a_{Ce}\,v^e{}_B,$$

where ∂_C denotes the derivative in the direction of $\mathbf{E}_C$. The above may be rewritten, using (2.8a), (2.10b), and (2.12b), as

$$D_C\,v^a{}_B = \mathscr{D}_C\,v^a{}_B - {}^*H_e{}^a{}_C\,v^e{}_B, \tag{2.14}$$

where

$$\mathscr{D}_C\,v^a{}_B \equiv \partial_C\,v^a{}_B - \Gamma^E_{CB}\,v^a{}_E + 2\,{}^*{}'\Omega^a_{eC}\,v^e{}_B \tag{2.15}$$

defines the invariant operator $\mathscr{D}$. Note that for any tensor $\mathbf{w}$ lying completely in $\{S\}$,

$$\mathbf{D}\mathbf{w} = \mathscr{D}\mathbf{w} = \nabla\mathbf{w}. \tag{2.16a}$$

Similarly, the operators ${}^*\mathbf{D}$ and ${}^*\mathscr{D}$ are defined by

$${}^*D_c\,v^a{}_B \equiv {}^4\nabla_c\,v^a{}_B, \tag{2.17}$$

which leads to

$${}^*D_c\,v^a{}_B = {}^*\mathscr{D}_c\,v^a{}_B + H_B{}^E{}_c\,v^a{}_E, \tag{2.18}$$

$${}^*\mathscr{D}_c\,v^a{}_B \equiv \partial_c\,v^a{}_B + {}^*\Gamma^a_{ce}\,v^e{}_B - 2'\Omega^E_{Bc}\,v^a{}_E. \tag{2.19}$$

Hence

$${}^*\mathbf{D}\,{}^*\mathbf{w} = {}^*\mathscr{D}\,{}^*\mathbf{w} = {}^*\nabla\,{}^*\mathbf{w} \tag{2.16b}$$

for any tensor ${}^*\mathbf{w}$ tangent to $\{T\}$.

The operators $\mathbf{D}$ and ${}^*\mathbf{D}$ are useful for calculational purposes since from their definitions (2.13) and (2.17), it follows that

$$\mathbf{D}\mathbf{g} = {}^*\mathbf{D}\mathbf{g} = \mathbf{D}\,{}^*\mathbf{g} = {}^*\mathbf{D}\,{}^*\mathbf{g} = 0, \tag{2.20}$$

and hence raising and lowering of indices a, A, commutes with both $\mathbf{D}$ and ${}^*\mathbf{D}$. However, $\mathscr{D}$ and ${}^*\mathscr{D}$ have the important property that they are dependent on the connection of V only via the induced connections in $\{S\}$ and $\{T\}$, respectively, a property not shared by $\mathbf{D}$ and ${}^*\mathbf{D}$. In fact, $\mathscr{D}$ and ${}^*\mathscr{D}$ are closely related to the Lie derivative when acting on quantities in $\{T\}$ and $\{S\}$, respectively. It can be shown, for example, that for any vectors $\mathbf{p}$ in S, and $\mathbf{q}$ in $\{T\}$,

$$q^e\,{}^*\mathscr{D}_e\,p^A = (\pounds_{\mathbf{q}}\,\mathbf{p})^A, \tag{2.21}$$

where $(\pounds_{\mathbf{q}}\mathbf{p})^A$ denotes the components, in the basis $\mathbf{E}_A$, of the projection of $\pounds_{\mathbf{q}}\mathbf{p}$ into $\{S\}$. This property of ${}^*\mathscr{D}$ illustrated in (2.21) is vital to the analysis of the IVP in the next section. The operators ${}^*\mathscr{D}$ and $\mathscr{D}$ are also used to define the extrinsic curvatures of $\{S\}$ and $\{T\}$. These are given, respectively, by

$$h_{CBa} \equiv H_{(CB)a} = -\tfrac{1}{2}\,{}^*\mathscr{D}_a\,g_{CB} \tag{2.22a}$$

$${}^*h_{cbA} \equiv {}^*H_{(cb)A} = -\tfrac{1}{2}\,\mathscr{D}_A\,{}^*g_{cb}. \tag{2.22b}$$

Equations (2.22) follow directly from (2.14), (2.18), and (2.20). Defining

$$h_a = h_E{}^E{}_a, \tag{2.23a}$$

$$^*h_A = {}^*h_e{}^e{}_A, \tag{2.23b}$$

and

$$\mathscr{g} \equiv \gamma^2 \equiv |g_{AB}|, \tag{2.24a}$$

$$^*\mathscr{g} \equiv |{}^*g_{ab}|, \tag{2.24b}$$

it follows that

$$^*\mathscr{D}_e \mathscr{g} = 2\mathscr{g} h_e, \tag{2.25a}$$

$$\mathscr{D}_E {}^*\mathscr{g} = 2\,{}^*\mathscr{g}\,{}^*h_E. \tag{2.25b}$$

The conformal 2-structure of $\{S\}$ is defined by

$$\tilde{g}_{AB} \equiv \mathscr{g}^{-1/2} g_{AB} \equiv \gamma^{-1} g_{AB}. \tag{2.26}$$

The conformal extrinsic curvature of $\{S\}$

$$\tilde{h}_{CBa} \equiv -\tfrac{1}{2}\mathscr{D}_a \tilde{g}_{CB} \tag{2.27}$$

is thus related to the trace-free part of the extrinsic curvature of $\{S\}$, $\bar{h}_{CBa}$ [in general, $\bar{(\ )}$ denotes the trace-free part with respect to **g**] by

$$\bar{h}_{CBa} \equiv \mathscr{g}^{1/2} \tilde{h}_{CBa}. \tag{2.28}$$

The gravitational degrees of freedom may be considered as being embodied in $\tilde{g}_{AB}$, $\tilde{h}_{CBa}$, or $\bar{h}_{CBa}$. One advantage of using $\bar{h}_{CBa}$ is that it is a tensor in $\{S\}$, as opposed to the other two, which are tensor densities in $\{S\}$.

The curvature tensors Riem (**g**) and Riem (*g) of $\{S\}$ and $\{T\}$ are defined, respectively, by

$$2D_{[D}\,D_{C]}p^A = R_{DCB}{}^A p^B - 2\Omega_{DC}{}^e\,{}^*\mathscr{D}_e p^A, \tag{2.29a}$$

$$2\,{}^*D_{[d}\,{}^*D_{c]}q^a = {}^*R_{dcb}{}^a q^b - 2{}^*\Omega_{dc}{}^E \mathscr{D}_E q^a. \tag{2.29b}$$

Note that, on account of (2.16), the operator **D** on the lhs of (2.29a) may be replaced by $\mathscr{D}$ or ∇, and similarly for (2.29b). Equations (2.29) lead to the following explicit expressions for $R_{DCB}{}^A$ and ${}^*R_{dcb}{}^a$:

$$R_{DCB}{}^A = 2\partial_{[D}\,\Gamma^A_{C]B} + 2\Gamma^A_{[D|E|}\,\Gamma^E_{C]B} + 4\Omega_{DC}{}^e\,{}'\Omega_B{}^A{}_e + 2\,{}''\Omega^E_{DC}\,\Gamma^A_{EB}, \tag{2.30a}$$

$${}^*R_{dcb}{}^a = 2\partial_{[d}\,{}^*\Gamma^a_{c]b} + 2\,{}^*\Gamma^a_{[d|e|}\,{}^*\Gamma^e_{c]b} + 4\,{}^*\Omega_{dc}{}^E\,{}^*{}'\Omega_b{}^a{}_E + 2\,{}^*{}''\Omega^e_{dc}\,{}^*\Gamma^a_{eb}. \tag{2.30b}$$

Thus Riem (**g**) depends on ${}^4\Gamma^i_{jk}$ only via its projection into $\{S\}$, and reduces to the usual expression for the Riemann tensor if $\{S\}$ is holonomic. Similarly for Riem (*g). Ric (**g**) and Ric (*g) are defined by contraction on the first and fourth indices of Riem (**g**), and Riem (*g), respectively. Ein (**g**) and Ein (*g), which of course vanish identically if $\{S\}$ and $\{T\}$ are holonomic, are, in general, given by

$$G_{CB} \equiv R_{CB} - \tfrac{1}{2}g_{CB} R = 2\Omega_{ECe} h_B{}^{Ee}, \tag{2.31a}$$

$${}^*G_{cb} \equiv {}^*R_{cb} - \tfrac{1}{2}\,{}^*g_{cb}\,{}^*R = 2\,{}^*\Omega_{ecE}\,{}^*h_b{}^{eE}. \tag{2.31b}$$

The independent projections of Riem (^{4}**g**) into $\{S\}$ and $\{T\}$ can now be calculated with little further effort. Note first that in the basis ($\mathbf{E}_i$, $\mathbf{E}^i$), Riem (^{4}**g**) is given by

$${}^4R_{kji}{}^h = 2\partial_{[k}\,{}^4\Gamma^h_{j]i} + 2{}^4\Gamma^h_{[k|l|}\,{}^4\Gamma^l_{j]i} + 2{}^4\Omega^l_{kj}\,{}^4\Gamma^h_{li}. \tag{2.32}$$

The projections are obtained merely by appropriately restricting the indices on either side of (2.32) and using (2.6), (2.8)–(2.13), (2.17), and (2.30). Four of the projections are just the well-known equations of Gauss and Codazzi:

$${}^4R_{DCB}{}^A = R_{DCB}{}^A + 2H_{[D|B|}{}^e H_{C]}{}^A{}_e + 2H_{[DC]}{}^e H_B{}^A{}_e \text{ (Gauss)}, \tag{2.33a}$$

$${}^4R_{dcb}{}^a = {}^*R_{dcb}{}^a + 2\,{}^*H_{[d|b|}{}^E\,{}^*H_{c]}{}^a{}_E + 2\,{}^*H_{[dc]}{}^E\,{}^*H_b{}^a{}_E \text{ (Gauss)}, \tag{2.33b}$$

$${}^4R_{DCB}{}^a = 2D_{[D}\,H_{C]B}{}^a + 2H_{[DC]}{}^e\,{}^*H_e{}^a{}_B \text{ (Codazzi)}, \tag{2.34a}$$

$${}^4R_{dcb}{}^A = 2\,{}^*D_{[d}\,{}^*H_{c]b}{}^A + 2\,{}^*H_{[dc]}{}^E H_E{}^A{}_b \text{ (Codazzi)}. \tag{2.34b}$$

The remaining independent projection is

$${}^4R_{dCB}{}^a = {}^*D_d\,H_{CB}{}^a + D_C\,{}^*H_d{}^a{}_B - H_C{}^E{}_d\,H_{EB}{}^a - {}^*H_d{}^e{}_C\,{}^*H_e{}^a{}_B. \tag{2.35}$$

The derivatives **D** and *D may be replaced in the above by $\mathscr{D}$ and $^*\mathscr{D}$, by using the identities (2.14) and (2.18). **H** and *H may be further broken up using (2.11), (2.22), and (2.23). Carrying out these operations, and taking suitable contractions and combinations of (2.33)–(2.35) and using (2.31), the following expressions for the projections of Ric (^{4}**g**) may be obtained:

$${}^4\bar{R}_{CB} = {}^*\mathscr{D}_e \bar{h}_{CB}{}^e + 2\bar{h}_{BEe}\,\bar{h}_C{}^{Ee} - 2\bar{h}_{(B}{}^{Ee}\Omega_{C)Ee} + \bar{T}\,[\mathscr{D}_{(C}\,{}^*h_{B)} - {}^*h_{efC}\,{}^*h^{ef}{}_B + {}^*\Omega_{efC}\,{}^*\Omega^{ef}{}_B], \tag{2.36}$$

$${}^4R_E{}^E = {}^*\mathscr{D}_e h^e - h_e h^e + \mathscr{D}_E\,{}^*h^E - {}^*h_{efE}\,{}^*h^{efE} + {}^*\Omega_{efE}\,{}^*\Omega^{efE} - 2\Omega_{EFe}\,\Omega^{EFe} + R, \tag{2.37}$$

$${}^4R_{cB} = {}^*\mathscr{D}_c\,{}^*h_B - {}^*\mathscr{D}_e\,{}^*h_c{}^e{}_B + {}^*\mathscr{D}_e\,{}^*\Omega_c{}^e{}_B + \tfrac{1}{2}\mathscr{D}_B h_c - \mathscr{D}_E \bar{h}_B{}^E{}_c + \mathscr{D}_E \Omega_B{}^E{}_c + \bar{h}_{BEc}\,{}^*h^E + \tfrac{1}{2}h_c\,{}^*h_B - \Omega_{BEc}\,{}^*h^E - 2\bar{h}_B{}^{Ee}\,{}^*h_{ceE} - 2\Omega_{BEe}\,{}^*\Omega_c{}^{eE} - {}^*\Omega_{ceB}\,h^e, \tag{2.38}$$

$${}^4R_{cb} = {}^*\mathscr{D}_{(c}\,h_{b)} - \bar{h}_{EFc}\,\bar{h}^{EF}{}_b + \tfrac{1}{2}h_c h_b + \Omega_{EFc}\Omega^{EF}{}_b + \mathscr{D}_E\,{}^*h_{cb}{}^E + 2{}^*h_{ceE}\,{}^*h_b{}^{eE} - 2\,{}^*h_{(c}{}^{eE}\,{}^*\Omega_{b)eE} - {}^*h_{ebE}\,{}^*h^E - {}^*g_{cb}({}^*\Omega_{efE}\,{}^*\Omega^{efE} - \tfrac{1}{2}{}^*R), \tag{2.39}$$

where $\bar{T}[.]$ denotes the trace-free part (with respect to **g**). From (2.39), one obtains the useful result that

$${}^4R_e{}^e = {}^*\mathscr{D}_e h^e - \tfrac{1}{2}h_e h^e - \bar{h}_{EFe}\,\bar{h}^{EFe} + \mathscr{D}_E\,{}^*h^E - {}^*h_E\,{}^*h^E + \Omega_{EFe}\,\Omega^{EFe} - 2{}^*\Omega_{efE}\,{}^*\Omega^{efE} + {}^*R. \tag{2.40}$$

The contracted Bianchi identities ${}^4\nabla_i\,{}^4G^{ij} = 0$ can be written in the form

$${}^4\nabla_i\,{}^4G^{ib} \equiv {}^*\mathscr{D}_e\,{}^4G^{eb} - h_e\,{}^4G^{eb} + \mathscr{D}_E\,{}^4G^{Eb} - {}^*h_E\,{}^4G^{Eb} - 2{}^*H_e{}^b{}_E\,{}^4G^{Ee} + h_{EF}{}^b\,{}^4G^{EF} = 0, \tag{2.41}$$

$${}^4\nabla_i\,{}^4G^{iB} \equiv {}^*\mathscr{D}_e\,{}^4G^{Be} - h_e\,{}^4G^{Be} - 2H_E{}^B{}_e\,{}^4G^{Ee} + \mathscr{D}_E\,{}^4G^{EB} - {}^*h_E\,{}^4G^{EB} + {}^*h_{ef}{}^B\,{}^4G^{ef} = 0. \tag{2.42}$$

Finally, note that a symmetry operation (*) may be defined by requiring that

$$*(\mathbf{E}_a) = \mathbf{E}_A,\ *(\mathbf{E}_A) = \mathbf{E}_a,\ *(\mathbf{E}^a) = \mathbf{E}^A,\ *(\mathbf{E}^A) = \mathbf{E}^a \tag{2.43}$$

and

$$*(\mathbf{g}) = *\mathbf{g},\quad *(*\mathbf{g}) = \mathbf{g}. \tag{2.44}$$

Hence $*(\mathbf{H}) = *\mathbf{H}$, $*(\mathbf{D}) = *\mathbf{D}$, etc. This essentially halves the number of calculations that need be performed. For example, all those equations labeled b are the (*) versions of those labeled a, and vice versa. This operation essentially inverts $\{S\}$ and $\{T\}$, and bears a close relationship to the Sachs symmetry operation defined in the Geroch, Held, and Penrose (GHP) modification of the spin-coefficient formalism.[8]

III. APPLICATION OF THE FORMALISM TO THE INITIAL VALUE PROBLEM

A. Foliations and fibrations

Suppose in some region of space-time V, a pair of commuting vector fields $\mathbf{e}_a$, which necessarily span a family of 2-surfaces $\{T_H\}$, are chosen together with a two-dimensional spacelike cross-section, ${}^oS_H$. The integral curves $\mathscr{C}_a$ of $\mathbf{e}_a$ form a fibration of space-time. A foliation $\{S_H\}$ of space-time is induced by Lie dragging ${}^oS_H$ by $\mathbf{e}_a$. Since $\mathbf{e}_a$ commute, a unique one-one correspondence between points on ${}^oS_H$ and any other $S_H \in \{S_H\}$ is set up by the curves $\mathscr{C}_a$; starting at points on ${}^oS_H$, and traveling fixed-parameter distances along $\mathscr{C}_0$, $\mathscr{C}_1$, respectively, will always lead to the same points on the same $S_H \in \{S_H\}$, independent of the order of traveling. The families of 3-surfaces obtained by dragging with $\mathbf{e}_0(\mathbf{e}_1)$ alone are denoted by $\{\Sigma_0\}(\{\Sigma_1\})$. In particular, the two 3-surfaces intersecting in ${}^oS_H$ are denoted by ${}^o\Sigma_0$ and ${}^o\Sigma_1$. Since $\mathbf{e}_a$ commute, there exist (locally) two scalar fields ϕ^a such that

$$\mathbf{e}_a = \partial/\partial\phi^a \tag{3.1}$$

and such that $\{\Sigma_0\}(\{\Sigma_1\})$ arise as the level surfaces of $\phi^1(\phi^0)$. It can be assumed without loss of generality that ${}^oS_H$ is given by $\phi^0 = \phi^1 = 0$.

Two IVP's will be considered in the sequel. First, the Cauchy problem, where the initial data surface $\mathscr{H} \equiv {}^o\Sigma_1$ is taken to be spacelike, and $\mathbf{e}_a$ to be nonpast pointing. Second, the characteristic problem, where ${}^o\Sigma_a$ are assumed to be null surfaces with $\mathbf{e}_a$ nonpast pointing, and $\mathscr{H} \equiv {}^o_+\Sigma_0 \cup {}^o_+\Sigma_1$ to consist of those portions of ${}^o\Sigma_a$ to the future of ${}^oS_H$, together with ${}^oS_H$ itself. (A third case which may be considered, but is not discussed in this paper, is the mixed problem in which, $\mathscr{H} \equiv {}^o_+\Sigma_0 \cup {}^o_+\Sigma_1$, with ${}^o_+\Sigma_1$ null and ${}^o_+\Sigma_0$ timelike.) In each case the problem is to determine what data on $\mathscr{H}$ are necessary and sufficient to determine a solution in some region in the future of $\mathscr{H}$.

Two ways in which the formalism of the preceding section may be used to study the IVP's considered above are as follows. First, it is possible to require that $\{S\}$ and $\{S_H\}$ coincide. A basis of 1-forms in $\{T\}$, $\mathbf{n}^a$, can then be chosen such that

$$\mathbf{E}^a \overset{\circ}{=} \mathbf{n}^a = d\phi^a. \tag{3.2}$$

In a general space-time, in which no gauge (coordinate) conditions have been imposed, $\{T_H\}$ and $\{T\}$ will not coincide. In fact, $\{T\}$ will be anholonomic. That is, the reciprocal basis of vectors in $\{T\}$, $\mathbf{n}_a$, is given by

$$\mathbf{E}_a \overset{\circ}{=} \mathbf{n}_a = \mathbf{e}_a - \mathbf{b}_a, \tag{3.3}$$

where $\mathbf{b}_a$ are "shift" vectors tangent to $\{S\}$. Imposition of conditions (3.2) and (3.3) leads to (what will be termed) the holonomic 2 + 2 formulation of the IVP. This has been extensively analyzed in Refs. 1 and 2 and will not be considered in any detail here.

A second possibility is to demand that $\{T\}$ and $\{T_H\}$ coincide. A basis of vectors in $\{T\}$, $\mathbf{n}_a$, can be chosen so that

$$\mathbf{E}_a \overset{\circ}{=} \mathbf{n}_a = \mathbf{e}_a = \partial/\partial\phi^a. \tag{3.4}$$

In this case it is not possible, in general, to demand that $\{S\}$ and $\{S_H\}$ coincide, and $\{S\}$ will now be anholonomic. It then follows that the reciprocal basis of 1-forms in $\{T\}$, $\mathbf{n}^a$, are given by

$$\mathbf{E}^a \overset{\circ}{=} \mathbf{n}^a = d\phi^a - \mathbf{c}^a, \tag{3.5}$$

where $\mathbf{c}^a$ are 1-forms in $\{S\}$. The conditions (3.4) and (3.5) lead to the anholonomic 2 + 2 formulation of the IVP.

The feature common to both formulations described above is the introduction of a particular dyad $(\mathbf{n}_a, \mathbf{n}^a)$ into $\{T\}$. One then works, not with the components of geometrical objects with respect to the basis $(\mathbf{E}_a, \mathbf{E}^a)$, but with the (scalar) dyad components of objects in $\{T\}$. The process of going from "live" tensorial indices to "dead" dyad indices (called strangling the indices) is denoted by $\overset{\circ}{=}$. [Cf. Eqs. (3.2)–(3.5).] For example, suppose that $\mathbf{w} = w_{aA}\,\mathbf{E}^a \otimes \mathbf{E}^A$ is a tensor of type (0,2). Writing

$$w_{aA} \overset{\circ}{=} {}_ow_{aA}$$

implies that

$${}_ow_{aA} = {}_ow_a(\mathbf{E}_A) \tag{3.6}$$

are the components of two 1-forms ${}_o\mathbf{w}_a$, lying in $\{S\}$, and labeled by a. Before going on to analyze the IVP's it is instructive to investigate the effect of introducing a dyad into $\{T\}$ on the formalism of Sec. II.

B. Introduction of a dyad basis in $\{T\}$

The dyad is defined by

$$\mathbf{E}_a \overset{\circ}{=} \mathbf{n}_a,\ \mathbf{E}^a \overset{\circ}{=} \mathbf{n}^a. \tag{3.7}$$

The definition of ${}^4\mathbf{\Omega}$ in Eq. (2.6), and its breakup in (2.7), (2.11), and (2.12) imply that

$$*''\Omega^a_{cb} \overset{\circ}{=} -\tfrac{1}{2}\langle \mathbf{n}^a,\ \pounds_{\mathbf{n}_c}\mathbf{n}_b\rangle = \tfrac{1}{2}\langle \pounds_{\mathbf{n}_c}\mathbf{n}^a,\ \mathbf{n}_b\rangle, \tag{3.8}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product,

$$*\Omega_{cb}{}^A \overset{\circ}{=} -\tfrac{1}{2}(\pounds_{\mathbf{n}_c}\mathbf{n}_b)^A, \tag{3.9}$$

$$*'\Omega_c{}^a{}_B \overset{\circ}{=} \tfrac{1}{2}(\pounds_{\mathbf{n}_c}\mathbf{n}^a)_B. \tag{3.10}$$

If $*\{{}^a_{cb}\}$ denotes the usual Christoffel symbols built out of

${}^*g_{ab}$, then

$$ {}^*\left\{ {a \atop c\,b} \right\} \mathrel{\mathring{=}} \tfrac{1}{2}\, {}^*g^{ad} \left[\pounds_{\mathbf{n}_c}\, {}^*_o g_{db} + \pounds_{\mathbf{n}_b}\, {}^*_o g_{dc} - \pounds_{\mathbf{n}_d}\, {}^*_o g_{cb} \right]. \quad (3.11) $$

For any $\mathbf{w} = w_a{}^B E^a \otimes E_B$, it follows from (2.21) that

$$ {}^*\mathscr{D}_c\, w_b{}^A \mathrel{\mathring{=}} (\pounds_{\mathbf{n}_c}\, {}_o\mathbf{w}_b)^A - {}^*_o\Gamma^e_{cb}\, {}_o w_e{}^A, \quad (3.12) $$

and from (2.15) and (3.10) that

$$ \mathscr{D}_C\, w_b{}^A \mathrel{\mathring{=}} \nabla_C\, {}_o w_b{}^A - (\pounds_{\mathbf{n}_b}\, \mathbf{n}^e)_C\, {}_o w_e{}^A. \quad (3.13) $$

Equations (3.10)–(3.13) applied to (2.22) lead to

$$ h_{CBa} \mathrel{\mathring{=}} -\tfrac{1}{2}(\pounds_{\mathbf{n}_a}\, \mathbf{g})_{CB} \equiv -\tfrac{1}{2}\pounds_{\mathbf{n}_a}\, g_{CB}, \quad (3.14) $$

$$ {}^*h_{cbA} \mathrel{\mathring{=}} -\tfrac{1}{2}\nabla_A\, {}^*_o g_{cb} + {}^*_o g_{e(c}\, (\pounds_{\mathbf{n}_{b)}}\, \mathbf{n}^e)_A. \quad (3.15) $$

From (2.6) it follows that ${}^4\boldsymbol{\Omega}$ obeys the following identity:

$$ \partial_{[k}\, {}^4\Omega^h_{ji]} - 2\Omega^l_{[kj}\, \Omega^h_{i]l} = 0, $$

and taking a suitable projection leads to the propagation equation

$$ \pounds_{\mathbf{n}_d}\, {}_o\Omega_{CB}{}^a = \nabla_{[C}\, (\pounds_{\mathbf{n}_d}\, \mathbf{n}^a)_{B\,]} \quad (3.16) $$

for ${}_o\boldsymbol{\Omega}^a$.

In either the holonomic or the anholonomic formulation of the IVP described in IIIA, that is, under either (3.2) and (3.3), or (3.4) and (3.5), it follows that

$$ {}^*_o\Omega^a_{cb} = 0 \Leftrightarrow {}^*_o\Gamma^a_{cb} = {}^*_o\left\{ {a \atop c\,b} \right\}. \quad (3.17) $$

Further, in the holonomic case,

$$ {}_o\boldsymbol{\Omega}^a = 0, \quad (3.18) $$

and (3.2) implies that (3.9) and (3.10) reduce to

$$ \pounds_{\mathbf{n}_c}\, \mathbf{n}_b = -2\, {}^*_o\Omega_{cb}{}^A\, \mathbf{E}_A, \quad \pounds_{\mathbf{n}_b}\, \mathbf{n}^a = 0, \quad (3.19) $$

respectively. Hence (3.15) becomes

$$ h_{cbA} \mathrel{\mathring{=}} -\tfrac{1}{2}\nabla_A\, {}^*_o g_{cb}. \quad (3.20) $$

Another consequence of (3.2) is that the Lie derivative with respect to $\mathbf{n}_a$ of any contravariant tensor tangent to $\{S\}$ remains tangent to $\{S\}$. For example,

$$ \mathbf{w} = w^{AB}\, \mathbf{E}_A \otimes \mathbf{E}_B \Rightarrow \pounds_{\mathbf{n}_c}\, \mathbf{w} = (\pounds_{\mathbf{n}_c}\, \mathbf{w})^{AB}\, \mathbf{E}_A \otimes \mathbf{E}_B. \quad (3.21) $$

Under the imposition of (3.2), (3.3), and (3.17)–(3.20), the formalism derived in Sec. II reduces to that of Refs. 1 and 2, and the analysis of the holonomic formulation of the IVP goes through as described therein.

In the anholonomic case, besides (3.17), the following specializations occur on account of (3.4) and (3.5):

$$ {}^*_o\boldsymbol{\Omega}_{ab} = 0, \quad (\pounds_{\mathbf{n}_c}\, \mathbf{n}^a) = {}^*{}'_o\Omega_c{}^a{}_B \mathbf{E}^B, \quad (3.22) $$

and a formula analogous to (3.21) holds, this time for covariant tensors, i.e.,

$$ \mathbf{w} = w_{AB}\, \mathbf{E}^A \otimes \mathbf{E}^B \Rightarrow \pounds_{\mathbf{n}_c}\, \mathbf{w} = (\pounds_{\mathbf{n}_c}\, \mathbf{w})_{AB}\, \mathbf{E}^A \otimes \mathbf{E}^B. \quad (3.23) $$

On each $S \in \{S\}$, an arbitrary basis of 1-forms $\mathbf{E}^A$ can be introduced. Having specified $\mathbf{E}^A$ on ${}^oS_H$, it is then possible to require that $\mathbf{E}^A$ be Lie propagated through V by

$$ \pounds_{\mathbf{n}_b}\, \mathbf{E}^A = 0 \Leftrightarrow {}'_o\boldsymbol{\Omega}_b = 0. \quad (3.24) $$

The above condition implies that, locally,

$$ \pounds_{\mathbf{n}_a} \equiv \partial/\partial\phi^a. \quad (3.25) $$

The metric ${}^4\mathbf{g}$ can be written as

$$ {}^4\mathbf{g} = \gamma \tilde{g}_{AB} \mathbf{E}^A \otimes \mathbf{E}^B + {}^*_o g_{ab}\, \mathbf{n}^a \otimes \mathbf{n}^b. \quad (3.26) $$

The ten unknown functions are then γ, $\tilde{g}_{AB}$, ${}^*_o g_{ab}$, and $\mathbf{n}^a$. [That $\mathbf{n}^a$ contain only four independent functions follows from the restriction that $\langle \mathbf{n}^a, \mathbf{n}_b \rangle = \delta^a_b$, or that, locally, Eq. (3.5) holds.] Four of these functions can be specified freely as gauge (coordinate) conditions, the remaining six being determined by the field equations. It is assumed, in order that the integration schemes in the sequel be valid, that all the field variables, Φ_A, say, are expandable in a power series in ϕ^a about ${}^oS_H$. Then Φ_A is given by

$$ \Phi_A(\phi^a) = \exp\left\{ \phi^e\, \pounds_{\mathbf{n}_e} \right\} \Phi_A(0). \quad (3.27) $$

Hence $\Phi_A(\phi^a)$ is determined in some neighborhood of ${}^oS_H$, provided its Lie derivatives with respect to $\mathbf{n}_a$ up to arbitrary order on ${}^oS_H$ are known.

C. The Cauchy problem

It is assumed in this subsection that ${}^*g_{00} > 0$. (The left subscript $({}_o)$ on dyad components will be omitted for the remainder of this section.) This ensures that $\mathbf{n}_0$ is tangent to a timelike congruence. It is also assumed that on $\mathscr{H}$ (and hence, by continuity, in some neighborhood of $\mathscr{H}$) ${}^*g_{11} < 0$. This ensures that $\mathbf{n}_1$ is spacelike. Since an analysis of the projections of ${}^4R_{ij}$ given by (2.36)–(2.40) reveals that they contain no second Lie derivatives with respect to $\mathbf{n}_0$ ("timelike" derivatives) of the set

$$ \{ {}^*g_{00},\ {}^*g_{01},\ \mathbf{n}^0 \}, \quad (3.28) $$

one possible choice of gauge variable is just the set (3.28). For simplicity one may put

$$ {}^*g_{00} = 1, \quad {}^*g_{01} = 0, \quad \mathbf{n}^0 = d\phi^0; \quad (3.29) $$

however, it is not necessary to choose these values. The six field variables to be determined by the field equations are then the set

$$ \{ {}^*g_{11},\ \mathbf{n}^1,\ \gamma,\ \tilde{g}_{AB} \}. \quad (3.30) $$

The analysis of the integration scheme for this version of the Cauchy problem is very similar to that of its holonomic counterpart given in Refs. 1 and 2. Hence it will only be briefly sketched here. The second (and by iteration, higher) time derivatives of (3.30) are determined on $\mathscr{H}$ by

$$ \left.\begin{aligned} &{}^4R_E{}^E - {}^4R_e{}^e = 0 \Rightarrow \pounds^2_{\mathbf{n}_0}\, {}^*g_{11} + J_{11} = 0, \\ &{}^4R^1{}_A = 0 \Rightarrow (\pounds^2_{\mathbf{n}_0}\, \mathbf{n}^1)_A + J^1{}_A = 0, \quad \text{constraint} \\ &{}^4R_E{}^E = 0 \Rightarrow \pounds^2_{\mathbf{n}_0}\, \gamma + J_E{}^E = 0, \\ &{}^4\bar{R}_{CB} = 0 \Rightarrow \pounds^2_{\mathbf{n}_0}\, \tilde{g}_{CB} + \tilde{J}_{CB} = 0, \quad \text{dynamical} \end{aligned} \right\} \text{main equations}, \quad (3.31) $$

where in each case J refers to "junk" terms not depending upon second-time derivatives. Of the initial data required on $\mathscr{H}$ to solve the main equations, the set

$$ \{ {}^*g_{11},\ \pounds_{\mathbf{n}_0}\, {}^*g_{11}, \mathbf{n}^1 \} \quad (3.32) $$

may be regarded as embodying the lower-dimensional gauge freedom and again, for simplicity, the values

$$ {}^*g_{11} = -1,\ \pounds_{\mathbf{n}_0}\, {}^*g_{11} = 0, \mathbf{n}^1 = d\phi^1 \text{ on } \mathscr{H} \quad (3.33) $$

may be chosen. The set

$$\{(\pounds_{\mathbf{n}_0}\,\mathbf{n}^1)_A,\ \gamma,\ \pounds_{\mathbf{n}_0}\gamma\} \tag{3.34}$$

is propagated off ${}^{\circ}S_H$ onto $\mathscr{H}$ by the four remaining independent field equations

$$\left.\begin{aligned} &{}^4G^{00}=0\Rightarrow \pounds^2_{\mathbf{n}_1}\,\gamma + J^{00}=0\\ &{}^4G^{01}=0\Rightarrow \pounds_{\mathbf{n}_0}\pounds_{\mathbf{n}_1}\gamma + J^{01}=0\\ &{}^4R^{\ 0}{}_A=0\Rightarrow (\pounds_{\mathbf{n}_1}\,\pounds_{\mathbf{n}_0}\,\mathbf{n}^1)_A + J^{0}{}_A=0 \end{aligned}\right\}\text{subsidiary equations.} \tag{3.35}$$

The subsidiary equations must be solved iteratively on ${}^{\circ}S_H$ for the successive Lie derivatives of (3.34) with respect to $\mathbf{n}_1$. The junk terms are known at each stage of the iteration.

The exact form of (3.31) and (3.35) may be obtained from (2.36)–(2.40). An analysis of the Bianchi identities (2.41) and (2.42) shows that the usual lemma can be proved, namely that if (3.31) hold everywhere, and (3.35) hold on $\mathscr{H}$, then the latter hold everywhere. It follows from the foregoing analysis that the freely specifiable, physically meaningful initial data are

$$\begin{aligned} &\{\tilde{g}_{AB},\ \pounds_{\mathbf{n}_0}\,\tilde{g}_{AB}\}\ \text{on}\ \mathscr{H},\\ &\{\gamma,\ \pounds_{\mathbf{n}_0}\,\gamma,\ \pounds_{\mathbf{n}_1}\,\gamma,(\pounds_{\mathbf{n}_0}\,\mathbf{n}^1)_A\}\ \text{on}\ {}^{\circ}S_H. \end{aligned} \tag{3.36}$$

Actually, $\tilde{g}_{AB}$ may be set to any value on ${}^{\circ}S_H$ by judicious choice of $\mathbf{E}^4$ on ${}^{\circ}S_H$.

D. The characteristic initial value problem

It is assumed in this subsection that ${}^*g_{11}=0$, which ensures that $\mathbf{n}_1$ is tangent to a null congruence. The anholonomic version of the Sachs double-null problem[9] will be the only one considered here, and the integration scheme, as for the Cauchy problem discussed in Sec. IIIC, is very similar to its holonomic counterpart, so that again only a brief sketch of the integration scheme will be given.

The remaining gauge freedom is embodied in the set

$$\{{}^*g_{00},\ \mathbf{n}^0\}, \tag{3.37}$$

which may conveniently specified to be

$${}^*g_{00}=0,\quad \mathbf{n}^0=d\phi^{\,0}. \tag{3.38}$$

The lower dimensional gauge freedom is

$$\{{}^*g_{01}\}\ \text{on}\ \mathscr{H},\quad \{\mathbf{n}^1\}\ \text{on}\ {}^{\circ}_{+}\Sigma_0. \tag{3.39}$$

It is required in any case that ${}^*g_{00}=0$ on ${}^{\circ}_{+}\Sigma_0$, which ensures that $\mathbf{n}_1$ is null on ${}^{\circ}_{+}\Sigma_0$ and for simplicity one can set

$${}^*g_{01}=1\ \text{on}\ \mathscr{H},\quad \mathbf{n}^1=d\phi^{\,1}\ \text{on}\ {}^{\circ}_{+}\Sigma_0. \tag{3.40}$$

Note that unless $\mathbf{n}^0=d\phi^{\,0}$ on ${}^{\circ}_{+}\Sigma_1$ and $\mathbf{n}^1=d\phi^{\,1}$ on ${}^{\circ}_{+}\Sigma_0$, then ${}^{\circ}_{+}\Sigma_a$ will not be null surfaces. The six field variables

$$\{{}^*g_{01},\ \mathbf{n}^1,\ \gamma,\ \tilde{g}_{AB}\} \tag{3.41}$$

are propagated via the equations

$$\left.\begin{aligned} &{}^4R_{11}=0\Rightarrow \pounds^2_{\mathbf{n}_1}\,\gamma - \pounds_{\mathbf{n}_1}\,\gamma\cdot\pounds_{\mathbf{n}_1}\\ &\quad \times\ln {}^*g_{01} + J_{11}=0,\\ &{}^4R_{1B}=0\Rightarrow(\pounds^2_{\mathbf{n}_1}\,\mathbf{n}^1)_B + J_{1B}=0,\\ &{}^4R_{E}{}^{E}=0\Rightarrow \pounds_{\mathbf{n}_1}\ \pounds_{\mathbf{n}_0}\,\gamma + J_E{}^E=0,\\ &{}^4\bar{R}_{CB}=0\Rightarrow \pounds_{\mathbf{n}_1}\ \pounds_{\mathbf{n}_0}\,\tilde{g}_{CB} + \tilde{J}_{CB}=0,\ \text{dynamical.} \end{aligned}\right\}\begin{array}{l}\text{constraint}\\ \\ \\ \\ \text{main}\\ \text{equations.}\end{array} \tag{3.42}$$

These equations must be solved iteratively on each successive portion of $\Sigma_1\in\{\Sigma_1\}$ to the future of ${}^{\circ}_{+}\Sigma_0$; the junk terms are known at each stage of iteration. Note that ${}^4R_{11}=0$ is solved for γ on ${}^{\circ}_{+}\Sigma_1$, but for ${}^*g_{01}$ thereafter. Of the initial data on ${}^{\circ}_{+}\Sigma_0$ required to solve the main equations, the set

$$\{\gamma,\ (\pounds_{\mathbf{n}_1}\,\mathbf{n}^1)_B\} \tag{3.43}$$

is determined by

$$\begin{aligned} &{}^4R_{00}=0\Rightarrow \pounds^2_{\mathbf{n}_0}\,\gamma + J_{00}=0,\quad \text{subsidiary equations}\\ &{}^4R_{0B}=0\Rightarrow(\pounds_{\mathbf{n}_0}\ \pounds_{\mathbf{n}_1}\,\mathbf{n}^1)_B + J_{0B}=0, \end{aligned} \tag{3.44}$$

which are solved iteratively on ${}^{\circ}_{+}\Sigma_0$. The Bianchi identities lead to the usual lemma, namely, that providing the main equations hold everywhere and the subsidiary equations hold on ${}^{\circ}_{+}\Sigma_0$, the latter hold everywhere else and the trivial equation ${}^4R_{01}=0$ is an algebraic consequence of the main equations. As usual, this lemma requries that h_1, the expansion of the null curves generated by $\mathbf{n}_1$, be nonzero. The physically meaningful initial data, which is freely specifiable, is seen to be

$$\{\tilde{g}_{AB}\}\ \text{on}\ \mathscr{H}/{}^{\circ}S_H,\{\gamma,\ \pounds_{\mathbf{n}_0}\,\gamma,\ \pounds_{\mathbf{n}_1}\,\gamma,(\pounds_{\mathbf{n}_1}\,\mathbf{n}^1)_A\}\ \text{on}\ {}^{\circ}S_H. \tag{3.45}$$

Anholonomic versions of the Robinson–Trautmann gauge[10] and the mixed null-timelike[11] (Bondi gauge) IVP can also be constructed, but again the analysis is similar to that of the holonomic analogues of Refs. 1 and 2.

IV. CONCLUSION

Comparison of the anholonomic formulation of the IVP given in Sec. III with that of the more standard holonomic version,[1,2] shows that in many ways the two approaches are complementary. It is yet an open question as to whether one has any particular advantage over the other, either from the point of view of proving "hard" theorems about the well-posedness of the IVP (in the nonanalytic case) for the particular choice of initial data considered here (i.e., the conformal 2-structure), or from purely calculational considerations of actually generating solutions from given intial data.

In this paper only one possible application of the formalism is given, and that is to the IVP under a certain class of gauge conditions. However, the formalism is sufficiently general to allow a careful analysis of the IVP using virtually any (any?) conceivable choice of gauge which casts the gravitational degrees of freedom into the conformal 2-structure. Two other applications which will be considered in forthcoming papers are to the asymptotic (characteristic) IVP and to the IVP for electrodynamcis on a curved space-time background.

Finally, it is to be noted that, under certain circumstances, the present formalism must be closely related to GHP. Indeed GHP takes as its starting point the singling out of two conjugate null directions at each point (necessarily orthogonal to a spacelike, and spanning a timelike, 2-surface element), realized by a pair of normalized null vectors. This can be achieved in the present formalism by the simple expedient of setting ${}^*g_{00}={}^*g_{11}\mathrel{\overset{\circ}{=}}0$, ${}^*g_{01}\mathrel{\overset{\circ}{=}}1$. A fuller discussion of this point will be given elsewhere.

ACKNOWLEDGMENT

I should like to thank R. A. d'Inverno for his helpful comments during the preparation of this manuscript.

[1]R. A. d'Inverno and J. Smallwood, Phys. Rev. D **22**, 1233 (1980).
[2]J. Smallwood, "On a Covariant 2 + 2 Formulation of the Initial Value Problem in Relativity," Ph.D. thesis, University of Southampton, 1980.
[3]R. A. d'Inverno and J. Stachel, J. Math. Phys. **19**, 2447 (1978).
[4]J. D. Kulhanek, Lett. Nuovo Cimento **10**, 389 (1974).
[5]N. Ó. Murchadha, "Existence and Uniqueness of Solutions to the Hamiltonian Constraint of General Relativity," Ph.D. thesis, Princeton University, 1973.
[6]J. Stachel, J. Math. Phys. **21**, 1776 (1980).
[7]J. A. Schouten, *Ricci Calculus* (Springer, Berlin, 1954). Sec. II draws heavily on Chap. V, Sec. 7.
[8]R. Geroch, A. Held, and R. Penrose, J. Math. Phys. **14**, 874 (1973).
[9]R. K. Sachs, J. Math. Phys. **3**, 908 (1962).
[10]R. Gambini and A. Restuccia, Phys. Rev. D **17**, 3150 (1978).
[11]L. A. Tamburino and J. H. Winicour, Phys. Rev. **150**, 1039 (1966).

Symmetries of stationary axially symmetric vacuum Einstein equations and the new family of exact solutions

Yoshimasa Nakamura
Department of Applied Mathematics and Physics, Faculty of Engineering, Kyoto University, Kyoto 606, Japan

(Received 1 July 1982; accepted for publication 1 October 1982)

A new family of exact solutions of the stationary axially symmetric vacuum Einstein equations is presented. The internal symmetries, SL(2,ℝ) rotation, and duality of parametrization, are combined to construct a Bäcklund transformation. For the special ansatz of the field equations, the Bäcklund transformation can be integrated and the hierarchy of ansatz is generated, recursively. The Riemann–Hilbert problem is also discussed for the inverse scattering formulas generating nonlocal symmetries.

PACS numbers: 04.20.Jb

1. INTRODUCTION

Soliton theoretic treatment and related results have played an important role in the systematic generation of exact solutions of the stationary axially symmetric vacuum Einstein equations. The first example is a discovery of several types of Bäcklund transformations.[1–7] Note especially the works of Kinnersley *et al.*[1,2] and Neugebauer.[3,4] Kinnersley *et al.* have exponentiated the infinitesimal tranformations due to the internal symmetries of the field equations, and have derived the Kerr–Tomimatsu–Sato family of exact solutions from the static ones. Neugebauer has developed the method of finding "multisoliton" solutions which are characterized by determinants.

On the other hand, in gauge theory, Corrigan, Fairlie, Yates, and Goddard[8] have succeeded in integrating the ansatz of Atiyah and Ward.[9] This ansatz yields the instanton solutions of SU(2) self-dual Yang–Mills equations. Consequently, the gauge potentials are described by solutions of the four-dimensional Laplace equation.

The purpose of this paper is twofold. First, we discuss the internal symmetries of the stationary axially symmetric vacuum gravitational fields and present a Bäcklund transformation. Following the idea of Corrigan *et al.*, we show that there exists a new family of exact solutions being outside the works of Kinnersley *et al.* and Neugebauer.

The second topic of soliton theory in general relativity is an application of the inverse scattering method[10–12] and the Riemann–Hilbert problem.[5,13] However, this framework is not available except for the construction of special solutions. To avoid the difficulties due to complicated spectral parameters, we propose an alternative approach for the Riemann–Hilbert problem. This brings us to the second purpose of the paper.

In the next section, we prepare a dual parametrization for the stationary axially symmetric vacuum Einstein equations. In Sec. 3, we explain how a Bäcklund transformation is constructed by the internal symmetries of the field equations. This nontrivial transformation is indispensable for our theory. Having defined a special ansatz which relates the field equations to a linear differential equation, we derive a family of exact solutions in Sec. 4. In Sec. 5, we discuss the Einstein equations and their linear problem over complexified space. The infinite number of nonlocal symmetries derives inverse scattering formulas. It is showed that the Riemann–Hilbert transformation generates new solutions. Finally in the Appendix, by solving linear differential equations we give concrete solutions which correspond to the ansatz.

2. FIELD EQUATIONS

The metric of a stationary axially symmetric space-time can be written as

$$-ds^2 = e^{\Gamma}(d\rho^2 + dz^2) + q_{\mu\nu}dx^{\mu}dx^{\nu}, \tag{2.1}$$

where Γ and $q_{\mu\nu}$ ($\mu, \nu = 1, 2$) are functions of ρ, z only. Here we use the notation $(x^0, x^1, x^2, x^3) = (\rho, t, \phi, z)$. Let us impose the supplementary condition $\det Q = -\rho^2$ on 2×2 matrix $Q = (q_{\mu\nu})$. It is known that the function Γ is determined up to a constant by quadratures once $q_{\mu\nu}$ are known (see Refs. 5 and 14). Thus Γ may be ignored in our discussion. The remaining Einstein field equations reduce to

$$\partial_\rho(\rho\partial_\rho Q\cdot Q^{-1}) + \partial_z(\rho\partial_z Q\cdot Q^{-1}) = 0, \tag{2.2}$$

where $\partial_\rho = \partial/\partial\rho$ and so on.

If we focus on a particular parametrization (f, ω) for Q,

$$Q = \begin{pmatrix} f & f\omega \\ f\omega & f^2\omega^2 - \rho^2 f^{-1} \end{pmatrix}, \tag{2.3}$$

we have a system of field equations

$$\begin{aligned} &f(\partial_\rho^2 + \rho^{-1}\partial_\rho + \partial_z^2)f - (\partial_\rho f)^2 - (\partial_z f)^2 \\ &\quad + (\rho^{-1}f^2\partial_\rho\omega)^2 + (\rho^{-1}f^2\partial_z\omega)^2 = 0, \\ &\partial_\rho(\rho^{-1}f^2\partial_\rho\omega) + \partial_z(\rho^{-1}f^2\partial_z\omega) = 0. \end{aligned} \tag{2.4}$$

The second equation implies that there exists a function $\psi = \psi(\rho, z)$ defined by

$$\partial_\rho\psi = \rho^{-1}f^2\partial_z\omega, \quad \partial_z\psi = -\rho^{-1}f^2\partial_\rho\omega. \tag{2.5}$$

The function ψ is called a twist potential. Eliminating ω in (2.4), we obtain

$$\begin{aligned} &f(\partial_\rho^2 + \rho^{-1}\partial_\rho + \partial_z^2)f - (\partial_\rho f)^2 - (\partial_z f)^2 \\ &\quad + (\partial_\rho\psi)^2 + (\partial_z\psi)^2 = 0, \\ &\partial_\rho(\rho f^{-2}\partial_\rho\psi) + \partial_z(\rho f^{-2}\partial_z\psi) = 0. \end{aligned} \tag{2.6}$$

There is a discrete mapping I which directly transforms (2.4) to (2.6) as follows:

$$I:\begin{pmatrix} f \\ \omega \end{pmatrix} \to \begin{pmatrix} \rho f^{-1} \\ i\psi \end{pmatrix}. \tag{2.7}$$

 0022-2488/83/030606-04$02.50

This operation was introduced originally by Neugebauer and Kramer.[15]

Now we define a matrix P by

$$P=\frac{1}{f}\begin{pmatrix}1 & \psi\\ \psi & f^2+\psi^2\end{pmatrix}. \tag{2.8}$$

Here note $\det P=1$. It can be proved that the field equations (2.6) are equivalent to

$$\partial_\rho(\rho\partial_\rho P\cdot P^{-1})+\partial_z(\rho\partial_z P\cdot P^{-1})=0. \tag{2.9}$$

We call P a dual parametrization of Q in the sense that Eq. (2.9) is the same as Eq. (2.2).

3. INTERNAL SYMMETRIES AND BÄCKLUND TRANSFORMATION

Let H be an SL(2,$\mathbb{R}$) constant matrix. The field equation (2.9) is invariant under the SL(2,$\mathbb{R}$) rotation

$$H{:}P\to HPH^{\mathrm{tr}}. \tag{3.1}$$

What this means is that (2.9) possesses an invariance group. There are three independent generators of this group which correspond to scale, gauge, and Ehlers transformations of (2.6), respectively.[16] Particularly, let H be

$$H_\gamma=\begin{pmatrix}0 & \gamma\\ -\gamma^{-1} & 0\end{pmatrix}, \tag{3.2}$$

which is an appropriate linear combination of generators for gauge and Ehlers transformations.

Providing that $f^2+\psi^2\neq 0$, we obtain the following lemma.

Lemma 3.1: If (f,ψ) with $f^2+\psi^2\neq 0$ satisfies (2.6), so does (f',ψ') defined by

$$f'=\gamma^2 f(f^2+\psi^2)^{-1},\qquad \psi'=-\gamma^2\psi(f^2+\psi^2)^{-1}. \tag{3.3}$$

We call (3.3) transformation γ after the work of Corrigan *et al.*[8] They have found a Bäcklund transformation for SU(2) self-dual gauge fields. It must be remarked that the transformation γ is discrete, since it gives an identity transformation when operated twice.

Next, we consider another internal symmetry of (2.9). Let us recall the parametrizations (2.3) and (2.8). The metric coefficients f and ω and the twist potential ψ are real functions of ρ and z. The mapping I given in (2.7) transforms real potentials to complex ones. Hence, it is natural to treat ψ analytically continued into complex space. Combining (2.5) with (2.7), we can prove

Lemma 3.2: Let (f,ψ) be a solution of (2.6). Then (f',ψ') defined by

$$f'=\rho f^{-1},\qquad \partial_\rho\psi'=-i\rho f^{-2}\partial_z\psi,\quad \partial_z\psi'=i\rho f^{-2}\partial_\rho\psi \tag{3.4}$$

is also a solution.

We call (3.4) transformation β. It is obvious that β is a discrete transformation.

We now propose a product transformation α: $\alpha=\beta\circ\gamma$ with the parameter γ equal to 1. Since $\beta\circ\gamma\neq\gamma\circ\beta$, α has a nontrivial effect. The operation of α, a Bäcklund tranformation, is given by the proposition which we now state.

Proposition 3.3: The transformation α acts on an initial solution (f,ψ) of (2.6) according to

$$\begin{aligned}&f'=\rho f^{-1}(f^2+\psi^2),\\ &\partial_\rho\psi'=i\rho f^{-2}(f^2+\psi^2)^2\cdot\partial_z\{\psi(f^2+\psi^2)^{-1}\},\\ &\partial_z\psi'=-i\rho f^{-2}(f^2+\psi^2)^2\cdot\partial_\rho\{\psi(f^2+\psi^2)^{-1}\},\end{aligned} \tag{3.5}$$

where (f',ψ') is another solution of (2.6).

It must be noted that by operating the transformation α even times, we can derive real potentials from real ones. The Bäcklund transformation α is a special case of the Kinnersely–Chitre transformation of the Geroch group.[1,16] While Cosgrove[17] has pointed out that Bäcklund tranformations given in Ref. 3 lie outside the Geroch group. It is also pointed out[18] that the transformation proposed in Ref. 6 is a member of the Geroch group.

4. FAMILY OF EXACT SOLUTIONS

We shall begin by introducing a new ansatz Δ which transforms the field equations (2.6) to a single linear differential equation. We have the following.

Proposition 4.1: A solution of (2.6) is given by

$$f=\partial_\rho\Delta,\quad \psi=\partial_z\Delta, \tag{4.1}$$

where $\Delta=\Delta(\rho,z)$ is a solution of the linear equation

$$(\partial_\rho^2-\rho^{-1}\partial_\rho+\partial_z^2)\Delta=0. \tag{4.2}$$

Proof: Substituting (4.1) into (2.6), we have

$$\begin{aligned}&(\partial_\rho\Delta\partial_\rho-\partial_\rho^2\Delta+\partial_z^2\Delta+\rho^{-1}\partial_\rho\Delta)\\ &\qquad\times(\partial_\rho^2-\rho^{-1}\partial_\rho+\partial_z^2)\Delta=0,\\ &(\partial_\rho\Delta\partial_z-2\partial_\rho\partial_z\Delta)(\partial_\rho^2-\rho^{-1}\partial_\rho+\partial_z^2)\Delta=0,\end{aligned}$$

respectively. This proves the proposition.

As a consequence of Proposition 4.1, the stationary axially symmetric Einstein equations are linearized via the ansatz Δ. A similar situation is known for the case of Weyl's family of exact solutions (see Ref. 14). However, Weyl solutions take the form $f=\exp(2\theta)$, $\psi=0$, where $\theta=\theta(\rho,z)$ satisfies the cylindrical Laplace equation $(\partial_\rho^2+\rho^{-1}\partial_\rho+\partial_z^2)\theta=0$. We call $(f,\psi)=(\partial_\rho\Delta,\partial_z\Delta)$ the ansatz P_2.

These heuristic considerations lead us to seek hierarchy of ansatz P_2 by means of the Bäcklund transformation α and the internal symmetries β and γ. We operate the transformation β on (f,ψ) of P_2. Let us substitute $f=\rho f'^{-1}$ and $\partial_z f=\partial_\rho\psi$ into the last equation of (3.4). We have $\partial_z f'=i\partial_z\psi'$, which is consistent with the second of (3.4). Therefore, neglecting the integration constant, we obtain

$$f'=i\psi'. \tag{4.3}$$

Writing $(f',\psi')=(\Delta_0^{-1},-i\Delta_0^{-1})$, we get the ansatz P_1'.

The variable $\Delta_0=\Delta_0(\rho,z)$ also satisfies a linear equation as follows.

Lemma 4.2: If Δ_0 satisfies

$$(\partial_\rho^2+\rho^{-1}\partial_\rho+\partial_z^2)\Delta_0=0, \tag{4.4}$$

then $(f,\psi)=(\Delta_0^{-1},-i\Delta_0^{-1})$ is a solution of (2.6).

Next, we define a new ansatz P_2' by α: $P_2'\to P_1'$. The ansatze P_2', P_1' correspond to (f,ψ), (f',ψ') in (3.5), respectively. Substituting $(f',\psi')=(\Delta_0^{-1},-i\Delta_0^{-1})$ into (3.5), we

have

$$f(f^2+\psi^2)^{-1}=\rho\Delta_0,$$
$$\partial_\rho\{\psi(f^2+\psi^2)^{-1}\}=\rho\partial_z\Delta_0, \tag{4.5}$$
$$\partial_z\{\psi(f^2+\psi^2)^{-1}\}=-\rho\partial_\rho\Delta_0.$$

Let us introduce $\Delta_1=\Delta_1(\rho,z)$ through

$$\partial_\rho\Delta_1=-\rho\partial_z\Delta_0,\quad \partial_z\Delta_1=\rho\partial_\rho\Delta_0. \tag{4.6}$$

These imply $(\partial_\rho^2-\rho^{-1}\partial_\rho+\partial_z^2)\Delta_1=0$. We can integrate (4.5) to obtain $\psi(f^2+\psi^2)^{-1}=-\Delta_1$. Then we prove the following proposition.

Proposition 4.3: The ansatz P_2' is written as

$$P_2':(f,\psi)=\left(\frac{\rho\Delta_0}{\begin{vmatrix}\rho\Delta_0 & \Delta_1\\ -\Delta_1 & \rho\Delta_0\end{vmatrix}},\frac{\Delta_1}{\begin{vmatrix}\rho\Delta_0 & \Delta_1\\ -\Delta_1 & \rho\Delta_0\end{vmatrix}}\right). \tag{4.7}$$

Corollary 4.4: The ansatz P_2 takes the form

$$P_2:(f,\psi)=(\rho\Delta_0,\Delta_1). \tag{4.8}$$

We use similar manipulation to derive P_3 and P_3' such that β: $P_3\to P_2'$, α: $P_3'\to P_2'$. Define Δ_2 by

$$\partial_\rho\Delta_2=-\rho\partial_z\Delta_1,\quad \partial_z\Delta_2=\rho\partial_\rho\Delta_1-2\Delta_1. \tag{4.9}$$

Then we have $(\partial_\rho^2-3\rho^{-1}\partial_\rho+\partial_z^2)\Delta_2=0$. After slightly complicated calculation, we prove

Proposition 4.5: The ansatze P_3 and P_3' are integrated to be

$$P_3:\ (f,\psi)=\left(\frac{\begin{vmatrix}\rho\Delta_0 & \Delta_1\\ -\Delta_1 & \rho\Delta_0\end{vmatrix}}{\Delta_0},\frac{i\begin{vmatrix}\Delta_1 & \Delta_2\\ -\Delta_0 & \Delta_1\end{vmatrix}}{\Delta_0}\right), \tag{4.10}$$

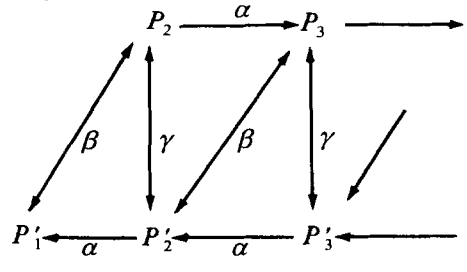
$$P_3':(f,\psi)=\left(\frac{\begin{vmatrix}\rho\Delta_0 & \Delta_1\\ -\Delta_1 & \rho\Delta_0\end{vmatrix}}{\begin{vmatrix}\rho^2\Delta_0 & \Delta_1 & -\Delta_2\\ -\Delta_1 & \Delta_0 & \Delta_1\\ -\Delta_2 & -\Delta_1 & \rho^2\Delta_0\end{vmatrix}},\frac{i\begin{vmatrix}\Delta_1 & \Delta_2\\ -\Delta_0 & \Delta_1\end{vmatrix}}{\begin{vmatrix}\rho^2\Delta_0 & \Delta_1 & -\Delta_2\\ -\Delta_1 & \Delta_0 & \Delta_1\\ -\Delta_2 & -\Delta_1 & \rho^2\Delta_0\end{vmatrix}}\right). \tag{4.11}$$

This suggests that we may express the ansatze P_l and P_l', $l\geqslant 4$, using determinants of ρ and Δ_r, $r\geqslant 0$. Here the variable Δ_r is defined by

$$\partial_\rho\Delta_r=-\rho\partial_\rho\Delta_{r-1},$$
$$\partial_z\Delta_r=\rho\partial_\rho\Delta_{r-1}+2(1-r)\Delta_{r-1}. \tag{4.12}$$

These lead to $\{\partial_\rho^2+(1-2r)\rho^{-1}\partial_\rho+\partial_z^2\}\Delta_r=0$.

The relationship between P_l and P_l' is shown in the following diagram;



Finally, we notice that P_l and P_l' (l even) are real functions of ρ and z. On the other hand, P_l and P_l' (l odd) give real metric coefficients by way of the discrete mapping I.

5. RIEMANN–HILBERT PROBLEM

Let us analytically continue $P(\rho,z)$ into complex space $(y,\bar{y},z)$, where $\bar{y}$ denotes a variable independent of complex conjugate y^* of y and $\rho^2=yy^*$. Real Euclidean space is specified by $\bar{y}=y^*$, and consequently

$$2y^{-1}\partial_{\bar{y}}=2\bar{y}^{-1}\partial_y=\rho^{-1}\partial_\rho. \tag{5.1}$$

It must be noted that a solution of (2.9) is given by a solution $P(\rho,z)$, $\rho^2=y\bar{y}$ of the equation

$$\partial_{\bar{y}}(\partial_y P\cdot P^{-1})+\partial_z(\partial_z P\cdot P^{-1})=0, \tag{5.2}$$

where $P\in SL(2,\mathbb{C})$. Equation (5.2) can be obtained as an Euler–Lagrange equation for the Lagrangian density $\mathcal{L}=\mathrm{Tr}(\partial_y P\partial_{\bar{y}}P^{-1}+\partial_z P\partial_z P^{-1})$.

We consider an infinitesimal transformation for (5.2),

$$\Lambda: P\to P(1+\Lambda), \tag{5.3}$$

where Λ is a 2×2 matrix function of y, $\bar{y}$, and z. We have

Lemma 5.1: If Λ satisfies

$$(\partial_y\partial_{\bar{y}}+\partial_z^2)\Lambda+[P^{-1}\partial_{\bar{y}}P,\partial_y\Lambda]+[P^{-1}\partial_z P,\partial_z\Lambda]=0, \tag{5.4}$$

then Λ gives a nonlocal symmetry of (5.2).

Using Lemma 5.1, we can introduce a sequence of nonlocal symmetries $\{S_0=\Lambda_0,S_1,S_2,\cdots\}$, where Λ_0 is a constant matrix sufficiently close to 0. That is, we have

Lemma 5.2: A sequence $\{S_n\}$, $n\geqslant 0$, defined by $S_0=\Lambda_0$ and

$$\partial_y S_{n+1}+\partial_z S_n+[P^{-1}\partial_z P,S_n]=0,$$
$$\partial_z S_{n+1}-\partial_{\bar{y}}S_n-[P^{-1}\partial_{\bar{y}}P,S_n]=0, \tag{5.5}$$

reads an infinite number of nonlocal symmetries of (5.2).

We set a generating function $S(\zeta)=\Sigma_{n=0}^{\infty}S_n\zeta^n$ and derive

$$D_1S(\zeta)+\zeta[P^{-1}\partial_z P,S(\zeta)]=0,$$
$$D_2S(\zeta)-\zeta[P^{-1}\partial_{\bar{y}}P,S(\zeta)]=0, \tag{5.6}$$

where $D_1=\partial_y+\zeta\partial_z$, $D_2=\partial_z-\zeta\partial_{\bar{y}}$, and $\zeta\in\mathbb{C}$. Let

$V(\zeta) = V(y, \bar{y}, z; \zeta)$ be a 2×2 matrix function satisfying $D_k V(\zeta) = 0$, $k = 1, 2$, and $V(0) = \Lambda_0$. We define the following 2×2 matrix function

$$S(\zeta) = P^{-1} Y(\zeta) V(\zeta) Y(\zeta)^{-1} P. \tag{5.7}$$

Providing that $Y(\zeta) = Y(y, \bar{y}, z; \zeta)$ be holomorphic near $\zeta = 0$, we may take $Y(0) = P$. We can prove

Proposition 5.3: Equation (5.2) is equivalent to the compatibility condition of linear differential equations, the inverse scattering formulas,

$$\begin{aligned} D_1 Y(\zeta) &= \partial_y P \cdot P^{-1} Y(\zeta), \\ D_2 Y(\zeta) &= \partial_z P \cdot P^{-1} Y(\zeta). \end{aligned} \tag{5.8}$$

As an application of this, we expand $Y(\zeta) = \Sigma_{n=0}^{\infty} Y_n \zeta^n$ to present an infinite number of nonlocal conserved currents $\{Y_n\}$, $n \geqslant 0$.

Next, we discuss the Riemann–Hilbert problem for (5.8). Let C be a closed analytic curve in ζ plane encircling the origin, and $C_+ (C_-)$ be the inside (outside) of C. Supposing the curve C be so small that $Y(\zeta)$ is analytic in $C \cup C_+$, we consider the following Riemann–Hilbert problem.

$$\begin{aligned} &X_-(\zeta') = X_+(\zeta') H(\zeta'), \quad \zeta' \in C, \\ &H(\zeta) = Y(\zeta) Y(\zeta) u(\zeta)^{-1}, \quad X_-(\infty) = 1. \end{aligned} \tag{5.9}$$

Here $u(\zeta)$ is an SL(2,C) matrix annihilated by D_1 and D_2. We assume that there exists a pair of fundamental solutions $X_{\pm}(\zeta)$, nonsingular matrices analytic on $C \cup C_{\pm}$. Following Hauser and Ernst[13] and Ueno and Nakamura,[19] we set

$$Y'(\zeta) = \begin{cases} X_+(\zeta) Y(\zeta) & \text{in } C_+, \\ X_-(\zeta) Y(\zeta) u(\zeta)^{-1} & \text{in } C_-. \end{cases} \tag{5.10}$$

This is the Riemann–Hilbert transformation induced from $u(\zeta)$. Then we obtain

Proposition 5.4: The matrix P' defined by $P' = Y'(0)$ is a solution of (5.2).

We give an outline of the proof of this proposition. Denote (5.8) as $D_k Y(\zeta) = B_k Y(\zeta)$, $k = 1, 2$, for simplicitly. Since $D_k u(\zeta) = 0$, the definition (5.10) with (5.9) gives

$$\begin{aligned} &D_k X_+ \cdot X_+^{-1} + X_+ B_k X_+^{-1} \\ &\quad = D_k X_- \cdot X_-^{-1} + X_- B_k X_-^{-1} \quad \text{on } C. \end{aligned}$$

Set $X = X_{\pm}(\zeta)$ in $C_{\pm}$. Since $D_k X \cdot X^{-1} + X B_k X^{-1}$ is independent of ζ, we may write $D_k X = B_k' X - X B_k$. These lead to $D_k Y'(\zeta) = B_k' Y'(\zeta)$, hence P' satisfies (5.2). Using (5.9) and $\det u(\zeta) = 1$, we can show $\det X = 1$, which means $\det P' = 1$. This completes the proof.

We recall that the Riemann–Hilbert transformation must satisfy the condition $D_k u(\zeta) = 0$. This implies $u(\zeta)$ is an entire function of ζ and w, the latter being defined by

$$w = \zeta y - z - \zeta^{-1} \bar{y}. \tag{5.11}$$

The quantity w also emerges in the theory of Yang–Mills–Higgs monopole solutions.[20]

ACKNOWLEDGMENTS

The author is deeply indebted to Professor Mineo Ikeda for reading the manuscript and offering valuable comments. Thanks are also due to Dr. Toshihiro Iwai and Dr. Kimio Ueno for useful discussions.

APPENDIX

The purpose of this appendix is to give concrete solutions of the field equations. For the ansatz P_1', we use Lemma 4.2 to derive real metric coefficients. A special solution $\Delta_0 = R^{-1}$, $R = \{\rho^2 + (z-a)^2\}^{1/2}$, $a \in \mathbb{R}$ of the cylindrical Laplace equation gives f and ω via the mapping I. We can easily determine the metric coefficient Γ appearing in (2.1). Then we have a metric

$$-ds^2 = \rho^{-1/2}(d\rho^2 + dz^2) - \rho R^{-1} dt^2 + 2\rho\, dt\, d\phi,$$

which includes a real parameter a.

The recursive relations in (4.12) lead to the variables Δ_1 and Δ_2,

$$\begin{aligned} \Delta_1 &= \{bR - (z-a)\} R^{-1}, \\ \Delta_2 &= \{R^2 - 2b(z-c)R + (z-a)^2\} R^{-1}, \end{aligned}$$

$a, b, c \in \mathbb{R}$. These variables give the metric coefficients f and ω of the ansatz P_3 as follows [cf. (4.10)]:

$$\begin{aligned} f &= \rho\{(b^2+1)R - 2b(z-a)\}^{-1}, \\ \omega &= \{(b^2+1)\rho^2 + (b^2-1)(z-a)^2 \\ &\quad + 2b(2z - a - c)R\} R^{-1}. \end{aligned}$$

The metric coefficients corresponding to the ansatz P_3' are also described by the above variables.

On the other hand, setting $\Delta_0 = e^{\pm kz} \chi(\rho)$, we can derive an ordinary differential equation,

$$\frac{d^2}{d\rho^2}\chi(\rho) + \rho^{-1}\frac{d}{d\rho}\chi(\rho) + k^2 \chi(\rho) = 0.$$

Therefore, it is known that a general solution of this equation is a linear combination of the Bessel function $J_0(k\rho)$ and the Neumann function $N_0(k\rho)$.

[1] W. Kinnersley and D. M. Chitre, J. Math. Phys. **19**, 2037 (1978).
[2] C. Hoenselaers, W. Kinnersley, and B. C. Xanthopoulos, J. Math. Phys. **20**, 2530 (1979).
[3] G. Neugebauer, J. Phys. A: Math. Gen. **12**, L67 (1979).
[4] G. Neugebauer, J. Phys. A: Math. Gen. **13**, L19 (1980).
[5] V. A. Belinsky and V. E. Zakharov, Sov. Phys. JETP **50**, 1 (1979).
[6] B. K. Harrison, Phys. Rev. Lett. **41**, 1197 (1978).
[7] Y. Nakamura, Ann. Mat. Pura. Appl. **129**, 153 (1981).
[8] E. F. Corrigan, D. B. Fairlie, R. G. Yates, and P. Goddard, Commun. Math. Phys. **58**, 223 (1978).
[9] M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977).
[10] D. Maison, J. Math. Phys. **20**, 871 (1979).
[11] Y. Nakamura, Math. Jpn. **24**, 469 (1979).
[12] G. Neugebauer, Phys. Lett. A **86**, 91 (1981).
[13] I. Hauser and F. J. Ernst, J. Math. Phys. **21**, 1126 (1980).
[14] W. Kinnersley, *Recent Progress in Exact Solutions, General Relativity and Gravitation*, edited by G. Shaviv and J. Rosen (Wiley, New York, 1975).
[15] G. Neugebauer and D. Kramer, Ann. Phys. **24**, 62 (1969).
[16] W. Kinnersley, J. Math. Phys. **18**, 1529 (1977).
[17] C. M. Cosgrove, J. Math. Phys. **21**, 2417 (1980).
[18] W. Kinnersley, in *Group Theoretical Methods in Physics, Proceedings of the IX International Colloquium*, edited by J. Ehlers *et al.*, *Lecture Notes in Physics*, Vol. 135 (Springer-Verlag, Berlin, 1980).
[19] K. Ueno and Y. Nakamura, Phys. Lett. B **109**, 273 (1982).
[20] R. S. Ward, Commun. Math. Phys. **79**, 317 (1981).

Plane symmetric charged dust distribution

Utpal Kumar De
Physics Department, Jadavpur University, Calcutta-700032, India
Dipankar Ray
Centre of Advanced Study in Applied Mathematics, Calcutta University, 92, A.P.C. Road, Calcutta-700009, India

(Received 28 September 1981; accepted for publication 11 December 1981)

An attempt is made to obtain the complete set of solutions of Einstein's equations for a nonstatic charged dust distribution with a comoving system of coordinates and the most general plane symmetric metric, i.e.,

$ds^2 = \exp[2u(x,t)]dt^2 - \exp[2v(x,t)]dx^2 - \exp[2w(x,t)](dy^2 + dz^2)$.

The field equations have been reduced to a single ordinary differential equation which is integrated in a particular case.

PACS numbers: 04.20. Jb, 02.30.Hq

1. INTRODUCTION

If some matter distribution flattens in the form of a disk, the likeliest symmetry it will have is plane symmetry. Since in the universe there is no dearth of such flattened cosmological objects, study of matter distribution with plane symmetry has some relevance. Charge-carrying incoherent matter in the form of dust is our concern here. Since it has already been established[1] that plane-symmetric incoherent dust distribution in equilibrium is not possible in general relativity, we seek solutions of Einstein's equations for a nonstatic charge-carrying dust distribution with the most general plane-symmetric metric given by

$$ds^2 = e^{2u(x,t)}dt^2 - e^{2v(x,t)}dx^2 - e^{2w(x,t)}(dy^2 + dz^2) \tag{1}$$

and

$(t,x,y,z) \equiv (x^0,x^1,x^2,x^3)$.

Some particular solutions of this type were previously reported by one of the present authors.[2,3] The present paper is an attempt to obtain the complete set of solutions of the above type for a comoving frame, i.e.,

$$\left.\begin{aligned} &v^0 = \frac{1}{(g_{00})^{1/2}} = e^{-u} \\ &\text{and} \\ &v^\mu = 0 \quad \text{for } \mu = 1,2,3. \end{aligned}\right\} \tag{2}$$

2. FIELD EQUATIONS

Equations (1) and (2) reduce the Einstein field equations for a charged dust to

$$\begin{aligned} R_0^0 &= e^{-2u}[\ddot{v} + 2\ddot{w} + \dot{v}^2 + 2\dot{w}^2 - \dot{u}(\dot{v} + 2\dot{w})] \\ &\quad - e^{-2v}[u'' + u'(u' - v' + 2w')] \\ &= -4\pi\rho + F^{01}F_{01}, \end{aligned} \tag{3a}$$

$$\begin{aligned} R_1^1 &= e^{-2u}[\ddot{v} + \dot{v}^2 - \dot{u}\dot{v} + 2\dot{v}\dot{w}] \\ &\quad - e^{-2v}[u'' + 2w'' + u'^2 + 2w'^2 - v'(u' + 2w')] \\ &= 4\pi\rho + F^{01}F_{01}, \end{aligned} \tag{3b}$$

$$\begin{aligned} R_2^2 = R_3^3 &= e^{-2u}[\ddot{w} + 2\dot{w}^2 + \dot{w}(\dot{v} - \dot{u})] \\ &\quad - e^{-2v}[w'' + 2w'^2 + w'(u' - v')] \\ &= 4\pi\rho - F^{01}F_{01}, \end{aligned} \tag{3c}$$

$$R_{01} = 2[\dot{w}(w' - u') + (\dot{w}' - w'\dot{v})] = 0, \tag{3d}$$

and the Maxwell equations for the same to

$$F^{01} = C(x)e^{-(u + v + 2w)}, \tag{4a}$$

$$4\pi\sigma = C'(x)e^{-(v + 2w)}, \tag{4b}$$

where ρ is the mass density, σ is the charge density and $F^{\mu\nu}$ is the electromagnetic field tensor whose only nonvanishing components are F^{01} and F^{10} due to symmetry of the problem. (A dot indicates differentiation with respect to t and a prime indicates differentiation with respect to x.)

Equations (3) and (4) provide the complete set of equations to be solved. However, to simplify the calculations we make direct use of conservation laws as follows.

From (1) and (2) $T^{\mu\nu}{}_{;\nu} = 0$ gives

$$F^{01} = -(\rho/\sigma)u'e^{-(u + 2v)} \tag{5}$$

and conservation of mass gives

$$e^{v + 2w}\rho = \text{const.} \tag{6}$$

From (4b) and (6), (ρ/σ) must be a function of x and hence

$$C^2(x) = (\rho/\sigma)^2 A^2(x), \tag{7}$$

where $A(x)$ is some function of x. From (4a), (5), and (7),

$$u'e^{(2w - v)} = -A(x). \tag{8}$$

From (3d) and (8),

$$(\dot{w}' + \dot{u}') + (w' + u')(\dot{w} - \dot{v}) = 0,$$

which on integration gives

$$w' + u' = B(x)e^{(v - w)}, \tag{9}$$

where $B(x)$ is some function. From (8) and (9)

$$w' = B(x)e^{(v - w)} + A(x)e^{(v - 2w)}. \tag{10}$$

Using (4b), Eq. (3a) − Eq. (2b) + 2×Eq. (3c) gives

$$e^{-2u}(4\ddot{w} + 6\dot{w}^2 - 4\dot{u}\dot{w}) - e^{-2v}(2w'^2 + 4u'w') = 2C^2e^{-2w}. \tag{11}$$

Putting w' and u' from (8) and (10) into (11) and integrating,

 0022-2488/83/030610-03$02.50

$$w = \pm e^{u}\{[A^2(x) - C^2(x)]e^{-4w} + D(x)e^{-3w} + B^2(x)e^{-2w}\}^{1/2} \tag{12}$$

where $D(x)$ is some function. Similarly, using (4b), (6), and (12), Eq. (3a) − Eq. (3b) − 2×Eq. (3c) gives

$$\dot{v} = \pm \frac{\{[A'(x) \pm C'(x)C(x)/A(x)]e^{-(v+2w)} + B'(x)e^{-(v+2w)} + [C^2(x) - A^2(x)]e^{-4w} - [D(x)/2]e^{-3w}\}}{e^{-u}\{[A^2(x) - C^2(x)]e^{-4w} + D(x)e^{-3w} + B^2(x)e^{-2w}\}^{1/2}}.$$

From (10), (12), and $\partial w/\partial x = \partial w'/\partial t$ one sees that the negative sign must be taken in $\pm C'(x)C(x)/A(x)$ in the above expression for $\dot{v}$, i.e.,

$$\dot{v} = \pm \frac{\{[A'(x) - C'(x)C(x)/A(x)]e^{-(v+2w)} + B'(x)e^{-(v+2w)} + [C^2(x) - A^2(x)]e^{-4w} - [D(x)/2]e^{-3w}\}}{e^{-u}\{[A^2(x) - C^2(x)]e^{-4w} + D(x)e^{-3w} + B^2(x)e^{-2w}\}^{1/2}} \tag{13}$$

and further

$$2B(x)[A(x)A'(x) - C(x)C'(x)] = A(x)[D'(x) - 2A(x)B'(x)]. \tag{14}$$

From Eq. (12) and (13)

$$\left.\frac{de^{v}}{dw}\right|_{x=\text{const}} = P(x,w)e^{v} + Q(x,w),$$

which is equivalent to

$$e^{v} = \exp\left[\int P(x,w)\,dw\right]\left\{\int Q(x,w)\exp\left[-\int P(x,w)\,dw\right]dw + R(x)\right\}, \tag{15}$$

where $R(x)$ is a function of x, and

$$P(x,w) = \frac{[C^2(x) - A^2(x)] - \frac{1}{2}D(x)e^{w}}{[A^2(x) - C^2(x)] + D(x)e^{w} + B^2(x)e^{2w}},$$

$$Q(x,w) = \frac{[A'(x) - C(x)C'(x)/A(x)] + B'(x)e^{w}}{[A^2(x) - C^2(x)]e^{-2w} + D(x)e^{-w} + B^2(x)}.$$

Now it can be checked through direct substitution that all the equations of (3) and (4) are satisfied by Eqs. (10), (12), (14), and (15) combined. Of these equations, Eq. (15) is already in the integrated form, (14) can be regarded as a definition of $D(x)$ for preassigned $A(x)$, $B(x)$, and $C(x)$, and (12) can be regarded as a definition of u. Thus we are left with Eq. (10), where v is given by (15) and $D(x)$ by (14) for arbitrarily assigned $A(x)$, $B(x)$, and $C(x)$. However, solving (10) where e^{v} is given by (14) and (15), seems to be an impossible task, but the special case of $B(x)/A(x) = \text{const}$ can be solved as follows

3. SOLUTION FOR A SPECIAL CASE

Take

$$B(x) = kA(x), \tag{16}$$

where k is a constant. Here, from (8) and (9),

$$\left.\frac{dw}{du}\right|_{t=\text{const}} + 1 + ke^{w} = 0,$$

which on integration gives

$$e^{-w} + k = e^{u}, \tag{17}$$

where the arbitrary function of t arising in the integration has been absorbed in e^{u}. This is possible due to Eq. (1). Equation (17) reduces Eq. (12) to

$$t = \int_{x=\text{const}} \frac{e^{w}\,dw}{(ke^{w} + 1)\{[A^2(x) - C^2(x)]e^{-4w} + D(x)e^{-3w} + B^2(x)e^{-2w}\}^{1/2}} + L(x)\cdots. \tag{18}$$

Using (16) one can integrate (15) to get

$$k^2[A^2(x) - C^2(x)] = D(x) - 2kA^2(x) + l, \tag{19}$$

where l is an arbitrary constant.

We shall now show that if w is defined by (18) and u by (17), $A(x)$, $C(x)$, and $D(x)$ are connected by (19), and v is defined through (10), then that provides the complete set of solutions for the special case $B(x)/A(x) = \text{const}$ which is under consideration here. That Eqs. (10), (17)–(19) are necessary for the present case has been shown already. That they are sufficient to provide a solution can be seen as follows.

Equations (17) and (18) together ensures (12). Equations (10), (12), and (19) together with the consistency condition $\partial w/\partial x = \partial w'/\partial t$ give (13). For the present case, (19) is equivalent to (14); (12) and (13) together ensure (15). Thus when (16) holds, Eqs. (10) and (17)–(19) together are sufficient to ensure (10), (12), (14), and (15), which in turn are sufficient to ensure that all the field equations are satisfied.

4. CONCLUSION

Therefore, in summary, the coupled Einstein–Maxwell equations for a nonstatic charge-carrying dust distribution in a comoving frame with the most general plane symmetric metric (1) reduces to Eq. (10), where u and v are given, respectively, by (12) and (15), and further $A(x)$, $B(x)$, $C(x)$, and $D(x)$

are connected by (14). The complete solution has been possible only when $B(x)/A(x) = \text{const}$, in which case the solution is given as follows, w is given by (18), u by (17), v by (10), and $A(x)$, $B(x)$, $C(x)$, and $D(x)$ satisfy (16) and (19). A special case of this solution was previously reported by one of the authors in the case $B(x) = 0$.

Although the equations have been completely integrated for $B/A = \text{const}$, the resulting equations still appear to be too involved for physical conclusions to be drawn. However, the fact that a class of exact solutions of Einstein's equations for the metric (1) has been obtained seems to be significant, because, although an extensive literature exists for the solutions for Einstein's equations for the plane symmetic metric, $ds^2 = e^{2u(x,t)}(dt^2 - dx^2) - e^{2w(x,t)}(dy^2 + dz^2)$, which is a special case of (1) for $u = v$, very few solutions are known for the metric (1). The present paper may open up some possibilities in that direction.

[1]U. K. De, J. Phys. A **6**, 478 (1973).
[2]U. K. De, Prog. Theor. Phys. **49**, 1546 (1973).
[3]U. K. De, Nuovo Cimento B **15**, 61 (1973).

Extended test particles in geometric fields

M. Toller
Department of Physics, University of Trento, I-38100 Trento, Italy

(Received 16 March 1982; accepted for publication 27 August 1982)

We treat the motion of an extended spinning test particle interacting with external fields belonging to a large class, which includes curvature, torsion, Maxwell, and Yang–Mills fields. We use a unified geometric description of all these fields based on a multidimensional space which generalizes the principal fiber bundle of the Lorentz frames. Guided by the work of Dixon, we give suitable definitions of the energy, the momentum, the angular momentum, the charge, and the isotopic spin of the extended particle, and we derive exact equations for the derivatives of these quantities. We discuss the pole–dipole approximation, and we compare our results with the special cases treated by other authors.

PACS numbers: 04.20.Me

I. INTRODUCTION

The motion of a spinning test particle in general relativity has been treated by several authors[1–3] and a very careful analysis of this problem has been done by Dixon,[4–7] who has given also a critical discussion of the preceding contributions. The effect of an electromagnetic field on a test particle with charge and electro-magnetic dipole moment has been examined by Dixon[4–7] and Souriau.[8] The case in which also torsion is present has been examined by Hehl[9] and by Trautman.[10] Particles provided with isotopic spin in a Yang–Mills field have been studied by Wong.[11] By isotopic spin we mean a set of charges corresponding to an arbitrary nonabelian internal gauge group.

The aim of the present paper is to give a unified treatment in which all the fields mentioned above are treated in a symmetric way. In fact, it is known that curvature, torsion, Maxwell and Yang–Mills fields, since they are gauge fields, can be described by means of a connection on a suitable principal fiber bundle.[12–15] Künzle[16] has used a principal fiber bundle to treat the motion of a spinning particle in gravitational and electromagnetic fields. Sternberg[17] has used fiber bundle techniques to treat the motion of a particle interacting with a Yang–Mills field. A further discussion of these problems can be found in Ref. 18.

Actually, one can see that the whole complicated mathematical structure of the principal fiber bundle is not essential for the treatment of the majority of our problems. Therefore, it is convenient to use a simpler and more general framework, based on an n-dimensional manifold $\mathscr{S}$, whose points are interpreted as orthonormal reference frames with a given choice of the internal gauge. The geometry of $\mathscr{S}$ is defined by n vector fields A_α, which represent infinitesimal parallel displacements, infinitesimal Lorentz transformations, and infinitesimal internal gauge transformations. All these fields can be treated in the same way and only at the end one has to remember their different physical meanings. A treatment of field theory based on these ideas has been given in Refs. 19–21 and the motion of test particles has been treated in Ref. 22 in the pole approximation.

The present treatment differs from the one of Ref. 22 because we use some of the ideas and the methods developed by Dixon,[4–7] which can be applied in a very natural way to the more general framework. In this way we extend a part of Dixon's results to the case in which torsion and Yang–Mills fields are present. Moreover, we clarify the essential geometric aspects of the problem and open the possibility of treating the case in which geometric fields of a more general (possibly nonlocal) nature are present.

In Sec. II we summarize the geometry of the space $\mathscr{S}$ and the balance equations of energy, momentum, relativistic angular momentum, electric charge and isotopic spin, which are the starting point of our approach. It is convenient to treat all these quantities in a symmetric way and to call them "the components of n-momentum." In Sec. III we treat the case in which all the geometric fields vanish, apart from a constant curvature, and the space $\mathscr{S}$ is isomorphic to a Lie group manifold. In Sec. IV we introduce some mathematical tools. In Sec. V we define the components of the n-momentum of an extended particle with respect to a given reference frame, and we find the exact equations which determine their derivatives in all the directions of the space $\mathscr{S}$.

In Sec. VI we consider a test particle, namely we disregard the geometric fields generated by the particle itself, and we write the equations derived in the preceding section in the pole–dipole approximation. Moreover, we restrict our attention to local theories. In Sec. VII we discuss the conditions which define the center of the particle and the approximations necessary in order to get a finite number of degrees of freedom. In Sec. VIII we rewrite the equations in the usual space–time formulation, and we compare them with the results of other authors.

II. THE BALANCE EQUATIONS

It has been shown[12–15] that the Einstein–Cartan theory of gravitation, Maxwell's theory of electromagnetism, and the nonabelian gauge field theories of the Yang–Mills type can be formulated in a very natural way on a principal fiber bundle $\mathscr{S}$.[23,24] The basis of $\mathscr{S}$ is the pseudo-Riemannian space–time $\mathscr{M}$ and its structural group is the product of the homogeneous Lorentz group and the internal gauge group. If we want to treat only the gravitational field, the space $\mathscr{S}$ is ten-dimensional and a point of $\mathscr{S}$ can be identified with an orthonormal tetrad in the space–time $\mathscr{M}$. If we consider also an internal gauge group of dimension k, the space $\mathscr{S}$ has

 0022-2488/83/030613-07$02.50

dimension $n = 10 + k$, and a point of $\mathscr{S}$ indicates a tetrad centered at a point of $\mathscr{M}$ and a choice of the internal gauge at this point.

Following Ref. 19, we introduce on the space $\mathscr{S}$ n vector fields A_α, which have a meaning even if $\mathscr{S}$ has not the whole structure of principal bundle. The infinitesimal parallel displacements of the tetrads and of the internal gauge along the tetrad vectors are generated by the vector fields $A_0,\ldots, A_3$. The infinitesimal rotations around the spacelike tetrad vectors are generated by the vector fields A_4, A_5, A_6 and the Lorentz boosts along the spacelike tetrad vectors are generated by A_7, A_8, A_9. The global internal gauge transformations are generated by the vector fields $A_{10},\ldots, A_{n-1}$. In the following, the indices $i,\ldots, z$ take the values $0,\ldots, 3$ and the indices $a,\ldots,h$ take the values $4,\ldots,(n-1)$. The Greek indices take all the values $0,\ldots,(n-1)$.

The Lie bracket of the vector fields A_α and A_β is given by (Refs. 19–22 use a different sign convention)

$$[A_\alpha, A_\beta] = L_\alpha A_\beta = -F^\gamma_{\alpha\beta} A_\gamma, \tag{2.1}$$

where $L_\alpha = L(A_\alpha)$ is the Lie derivative corresponding to the vector field A_α. The scalar fields $F^\gamma_{\alpha\beta}$ are called the structure coefficients of the space $\mathscr{S}$. They satisfy the generalized Jacobi identity

$$L_{[\alpha} F^\delta_{\beta\gamma]} = F^\eta_{[\alpha\beta} F^\delta_{\gamma]\eta}, \tag{2.2}$$

where $[\alpha\beta\gamma]$ indicates the antisymmetrization with respect to the indices α,β,γ.

We shall use also ten differential 1-forms ω^β defined by

$$i_\alpha \omega^\beta = \delta^\beta_\alpha, \tag{2.3}$$

where $i_\alpha = i(A_\alpha)$ is the inner product operator corresponding to the vector field A_α. They have the properties

$$L_\alpha \omega^\beta = F^\beta_{\alpha\gamma}\omega^\gamma, \tag{2.4}$$

$$d\omega^\gamma = \tfrac{1}{2} F^\gamma_{\alpha\beta}\omega^\alpha \wedge \omega^\beta. \tag{2.5}$$

When $\mathscr{S}$ is the principal bundle described above, the structure coefficients have simple geometric and physical meanings. If we indicate by $\hat{F}^\gamma_{\alpha\beta}$ the structure constants of the product of the Poincaré group and the internal gauge group, we have

$$F^\beta_{a\alpha} = -F^\beta_{\alpha a} = \hat{F}^\beta_{a\alpha}. \tag{2.6}$$

The quantities F^j_{ik} are the components of the torsion tensor; for $a = 4,\ldots,9$ the quantities F^a_{ik} represent the curvature and for $a \geqslant 10$, they are the field strengths of the internal gauge fields. The usual Riemann curvature tensor is given by

$$R^r{}_{sik} = -F^r_{as} F^a_{ik}. \tag{2.7}$$

From Eqs. (2.2) and (2.6) one gets the Bianchi identities, the homogeneous Maxwell and Yang–Mills equations and also the transformation properties of the quantities F^α_{ik} under Lorentz and internal gauge transformations. They are given by

$$L_a F^\alpha_{ik} = \hat{F}^\alpha_{a\beta} F^\beta_{ik} - \hat{F}^j_{ai} F^\alpha_{jk} - \hat{F}^j_{ak} F^\alpha_{ij}. \tag{2.8}$$

For a large part of our analysis it is not necessary and not even useful to assume that $\mathscr{S}$ has a structure of principal fiber bundle and that the structure coefficients have the property (2.6). We shall only assume that $\mathscr{S}$ is an infinitely differentiable n-dimensional manifold and that the n vector fields A_α are infinitely differentiable and linearly independent in every point of $\mathscr{S}$. The vector fields A_α define the geometry of the space $\mathscr{S}$.

We describe matter by means of a set of scalar fields f^A in the space $\mathscr{S}$. Note that the components of a tensor or spinor field in the space–time $\mathscr{M}$ can be considered as scalar fields in the space $\mathscr{S}$. The matter fields f^A interact with the geometric fields A_α, which we consider as known external fields.

Following Ref. 19, we derive the matter field equations from an action principle of the kind

$$\delta \int_{\mathscr{N}} \lambda^M = 0, \tag{2.9}$$

where $\mathscr{N}$ is an arbitrary four-dimensional oriented compact surface and the variations δf^A vanish on the boundary $\partial\mathscr{N}$ of $\mathscr{N}$. The Lagrangian form λ^M is a differential 4-form of the kind

$$\lambda^M = \tfrac{1}{24} \lambda^M_{\alpha\beta\gamma\delta}(f^A, L_\alpha f^A, F^\rho_{\mu\nu})\omega^\alpha \wedge \omega^\beta \wedge \omega^\gamma \wedge \omega^\delta. \tag{2.10}$$

If it does not depend on the structure coefficients $F^\rho_{\mu\nu}$, we say that the coupling between matter and geometry is minimal. The general form of the field equations and some specific examples have been considered in Refs. 19–21.

In all the interesting cases considered up to now, if the field equations are satisfied, we have

$$d\lambda^M = 0, \tag{2.11}$$

and the action integral does not change if the surface $\mathscr{N}$ is deformed keeping its boundary $\partial\mathscr{N}$ fixed. Then, if we consider a variation of the general kind which includes also a displacement of the surface $\mathscr{N}$ and of its boundary $\partial\mathscr{N}$ given by the infinitesimal vector field ϵB, we have instead of Eq. (2.9)

$$\delta \int_{\mathscr{N}} \lambda^M = \int_{\partial\mathscr{N}} [\delta f^A \rho_A - \epsilon \tau^M(B)], \tag{2.12}$$

where the canonical momenta ρ_A are differential 3-forms defined in Ref. 19 and the 3-form $\tau^M(B)$ is given by

$$\tau^M(B) = L(B) f^A \rho_A - i(B)\lambda^M. \tag{2.13}$$

We can always put

$$B = b^\alpha A_\alpha \tag{2.14}$$

and write

$$\tau^M(B) = b^\alpha \tau^M(A_\alpha) = b^\alpha \tau^M_\alpha. \tag{2.15}$$

The differential forms τ^M_α represent the density and the flow of energy, momentum, relativistic spin, electric charge, and isotopic spin, namely of the n-momentum of matter.

A symmetry transformation of the space $\mathscr{S}$ is a diffeomorphism which transforms the vector fields A_α into themselves. A vector field B generates an infinitesimal symmetry transformation of $\mathscr{S}$ if one of the following equivalent conditions is satisfied:

$$L(B)A_\alpha = [B, A_\alpha] = -L_\alpha B = 0, \tag{2.16}$$

$$L(B)\omega^\alpha = 0. \tag{2.17}$$

Then we have the conservation law (Noether theorem)

$$d\tau^M(B) = 0. \tag{2.18}$$

Otherwise, we have the balance equation[19,22]

$$d\tau^M(B) = L(B)\omega^\alpha \wedge \tau_\alpha^M - \frac{\partial \lambda^M}{\partial F^\rho_{\mu\nu}} L(B) F^\rho_{\mu\nu}. \tag{2.19}$$

The last term is absent if the coupling is minimal.

If the differential forms τ_α^M are given by

$$\tau_\alpha^M = T^i_{\ \alpha}\sigma_i, \tag{2.20}$$

where

$$\sigma_i = \tfrac{1}{6} e_{irst}\omega^r \wedge \omega^s \wedge \omega^t, \tag{2.21}$$

Eq. (2.19) gives the usual balance equations and also the transformation properties of the quantities $T^i_{\ \alpha}$ under the Lorentz and the internal gauge groups.[19]

III. TEST PARTICLE ON A GROUP MANIFOLD

We consider first the case in which the structure coefficients are constant, and therefore, as we see from Eq. (2.2), they are the structure constants of the Lie algebra of a group $\mathscr{P}$. This happens when all the geometric fields are absent or there is only a gravitational field with constant curvature. The group $\mathscr{P}$ is the product of the internal gauge group with the Poincaré or the de Sitter group. In this case the vector fields A_α generate a local action of the group $\mathscr{P}$ on the manifold $\mathscr{S}$. We assume that actually there is a global action of $\mathscr{P}$ on $\mathscr{S}$ and that, given two elements s and s' of $\mathscr{S}$, there is one and only one element g of $\mathscr{P}$ with the property

$$s = gs'. \tag{3.1}$$

We also write this relation in the form

$$g = s/s'. \tag{3.2}$$

If we fix the element s', these equations define an isomorphism between $\mathscr{S}$ and the manifold of the Lie group $\mathscr{P}$, which transforms the vector fields A_α into the generators of the infinitesimal left translations of the group $\mathscr{P}$.

In this case, there are n linearly independent vector fields which satisfy the conditions (2.16) or (2.17). They correspond to the generators of the infinitesimal right translations of the group $\mathscr{P}$. We indicate by B_α the vector field which satisfies Eq. (2.16) and coincides with A_α at the point s'. It can be written in the form

$$B_\alpha(s) = D^\beta_\alpha(s/s')A_\beta(s), \tag{3.3}$$

where

$$D^\beta_\alpha(e) = \delta^\beta_\alpha, \tag{3.4}$$

$$L_\theta D^\beta_\alpha(s/s') = F^\beta_{\theta\gamma}(s) D^\gamma_\alpha(s/s'). \tag{3.5}$$

These equations show that the matrices $D^\beta_\alpha(g)$ form the adjoint representation of the group $\mathscr{P}$. They have also the property

$$L'_\theta D^\beta_\alpha(s/s') = -D^\beta_\gamma(s/s')F^\gamma_{\theta\alpha}(s'), \tag{3.6}$$

where L'_θ indicates the derivative with respect to the argument s'.

The closed differential 3-forms

$$\tau^M(B_\alpha) = D^\beta_\alpha(s/s')\tau^M_\beta \tag{3.7}$$

represent the density and the flow of n-momentum defined in the fixed reference frame s', while the forms τ_α^M represent the same quantities defined with respect to the moving reference frame s.

In the space–time, the density of n-momentum of a small particle vanishes outside a thin neighborhood of a worldline. In the space $\mathscr{S}$ the differential forms τ_α^M vanish outside a region $\mathscr{U}$ which is a thin neighborhood of a $(n-3)$-dimensional submanifold of $\mathscr{S}$. A component of the total n-momentum of the particle measured with respect to the reference frame s' is given by the integral

$$p_\alpha(s') = \int_{\mathscr{R}} D^\beta_\alpha(s/s')\tau^M_\beta, \tag{3.8}$$

where $\mathscr{R}$ is a compact three-dimensional oriented submanifold of $\mathscr{S}$ whose boundary $\partial\mathscr{R}$ is contained in the complement $c\mathscr{U}$ of $\mathscr{U}$. We have to assume that $\mathscr{R}$ crosses the region $\mathscr{U}$ in a given way.

The choice of $\mathscr{R}$ can be discussed by means of the language of singular relative homology.[25] We indicate by $\mathscr{R}$ a singular relative cycle belonging to $Z_3(\mathscr{S}, c\mathscr{U})$. If we integrate a closed differential form vanishing on $c\mathscr{U}$, the integral depends only on the relative homology class of $\mathscr{R}$, and, in order to have a well-defined value of p_α, we have only to choose a suitable element of the relative homology group $H_3(\mathscr{S}, c\mathscr{U})$. We shall discuss this choice in Sec. V.

From Eq. (3.8) and the representation property of the matrices $D^\beta_\alpha(g)$ we obtain

$$p_\alpha(s) = D^\beta_\alpha(s'/s)p_\beta(s') \tag{3.9}$$

and from Eq. (3.6) we get the differential equation

$$L_\theta p_\alpha = -F^\beta_{\theta\alpha}p_\beta. \tag{3.10}$$

These equations show that when we change the reference frame, the n-momentum transforms according to the coadjoint representation of the group $\mathscr{P}$.

In order to describe the motion of the particle, we have to define its center. In some reference frames the center coincides with the origin when the time coordinate vanishes; as it has been discussed in Ref. 22, all these "central frames" form a $(n-3)$-dimensional submanifold $\mathscr{V}_c$ of $\mathscr{S}$. The simplest way to characterize them is to impose three conditions to the values of the quantities p_α.

$$\Gamma_i(p_\alpha) = 0, \quad i = 1,2,3. \tag{3.11}$$

If we substitute Eq. (3.9) into these conditions, we find three equations in the variable s, which define the manifold $\mathscr{V}_c$. Note that in the absence of external fields the motion of the center depends only on the initial n-momenta, namely it is not coupled with the internal degrees of freedom of the particle. This desirable feature would be lost if the conditions (3.11) contained other dynamical variables connected with the internal motions. The choice of the functions Γ_i will be discussed in Sec. VII.

IV. NORMAL COORDINATES

In order to define the n-momentum when the structure coefficients are not constant, we have to find a useful generalization of the matrices $D^\beta_\alpha(s/s')$ which appear in Eq. (3.8). We indicate by $\exp(\xi B)$ the one-parameter group of mappings of $\mathscr{S}$ onto itself generated (at least locally) by the vec-

tor field B. If the point s is given by

$$s = \exp(\xi^\alpha A_\alpha)s', \tag{4.1}$$

the parameters ξ^α are the "normal coordinates" of s with respect to the fixed point s'. They are defined univocally if s belongs to a sufficiently small neighborhood of s' and we indicate them by $\xi^\alpha(s,s')$.

One can easily show that

$$\xi^\alpha(s,s') = -\xi^\alpha(s',s), \tag{4.2}$$

$$\xi^\theta(s,s')L_\theta \xi^\alpha(s,s') = \xi^\alpha(s,s'), \tag{4.3}$$

$$\xi^\theta(s,s')L'_\theta \xi^\alpha(s,s') = -\xi^\alpha(s,s'), \tag{4.4}$$

where the derivatives L_θ and L'_θ operate, respectively, on the arguments s and s'. From Eqs. (4.3) and (4.4) we have

$$[L_\theta \xi^\alpha(s,s')]_{s=s'} = \delta^\alpha_\theta, \tag{4.5}$$

$$[L'_\theta \xi^\alpha(s,s')]_{s=s'} = -\delta^\alpha_\theta, \tag{4.6}$$

and we get also the useful formula

$$\omega^\alpha = d\xi^\alpha + \tfrac{1}{2}F^\alpha_{\beta\gamma}\xi^\beta d\xi^\gamma + \cdots. \tag{4.7}$$

From Eq. (4.5) we see that if s belongs to a sufficiently small neighborhood of s' the formula

$$L'_\beta \xi^\alpha(s,s') = -D^\gamma_\beta(s,s')L_\gamma \xi^\alpha(s,s') \tag{4.8}$$

defines the matrix $D^\gamma_\alpha(s,s')$ uniquely. From Eqs. (4.5) and (4.6) we get

$$D^\beta_\alpha(s,s) = \delta^\beta_\alpha. \tag{4.9}$$

If we act on both sides of Eq. (4.8) with the operator $\xi^\theta L_\theta$, after some calculations we get

$$\xi^\theta(s,s')[L_\theta D^\beta_\alpha(s,s') - F^\beta_{\theta\gamma}(s)D^\gamma_\alpha(s,s')] = 0. \tag{4.10}$$

A comparison of Eqs. (4.9) and (4.10) with Eqs. (3.4) and (3.5) shows that when the structure coefficients are constant the matrix $D^\beta_\alpha(s,s')$ coincides with the matrix $D^\beta_\alpha(s/s')$ defined in the preceding section.

In the general case the matrices $D^\beta_\alpha(s,s')$ do not have the representation property, but one can easily show that

$$D^\beta_\gamma(s,s')D^\gamma_\alpha(s',s) = \delta^\beta_\alpha. \tag{4.11}$$

From Eq. (4.10) we get

$$[L_\theta D^\beta_\alpha(s,s')]_{s=s'} = F^\beta_{\theta\alpha}. \tag{4.12}$$

If we put

$$L_\theta D^\beta_\alpha(s,s') = [C^\beta_{\theta\gamma}(s,s') + F^\beta_{\theta\gamma}(s)]D^\gamma_\alpha(s,s'), \tag{4.13}$$

from Eq. (4.12), we get

$$C^\beta_{\theta\gamma}(s,s) = 0 \tag{4.14}$$

and, using also Eqs. (2.1) and (2.2), after some calculations we obtain

$$[L_\rho C^\beta_{\theta\alpha}(s,s')]_{s=s'} = \tfrac{1}{2}L_\alpha F^\beta_{\rho\theta}. \tag{4.15}$$

From Eqs. (4.11) and (4.13), we have also

$$L'_\theta D^\beta_\alpha(s,s') = -D^\beta_\gamma(s,s')[C^\gamma_{\theta\alpha}(s',s) + F^\gamma_{\theta\alpha}(s')]. \tag{4.16}$$

Equations (4.9), (4.12), (4.14), and (4.15) give the first few terms of the power expansions

$$D^\beta_\alpha(s,s') = \delta^\beta_\alpha + F^\beta_{\rho\alpha}\xi^\rho(s,s') + \cdots, \tag{4.17}$$

$$C^\beta_{\theta\alpha}(s,s') = \tfrac{1}{2}L_\alpha F^\beta_{\rho\theta}\xi^\rho(s,s') + \cdots. \tag{4.18}$$

If the vector field

$$B = b^\alpha A_\alpha \tag{4.19}$$

generates an infinitesimal symmetry transformation of the space $\mathscr{S}$, namely, it satisfies the condition (2.16), we have

$$L_\theta b^\alpha = F^\alpha_{\theta\beta}b^\beta \tag{4.20}$$

and, therefore, from Eqs. (4.9) and (4.10)

$$b^\alpha(s) = D^\alpha_\beta(s,s')b^\beta(s'), \tag{4.21}$$

$$B(s) = b^\beta(s')D^\alpha_\beta(s,s')A_\alpha(s). \tag{4.22}$$

V. THE n-MOMENTUM OF A PARTICLE

In order to define the n-momentum of an extended particle in the presence of geometric fields, we have to generalize Eq. (3.8). The results of the preceding section suggest that we can put

$$p_\alpha(s') = \int_{\mathscr{R}(s')} D^\beta_\alpha(s,s')\tau^M_\beta. \tag{5.1}$$

In general the form under the integral is not closed and therefore this expression depends on the choice of the integration surface $\mathscr{R}(s')$ when it varies within the correct relative homology class.

However, if the vector field B given by Eq. (4.19) generates an infinitesimal symmetry transformation, the differential form

$$\tau^M(B) = b^\beta(s')D^\alpha_\beta(s,s')\tau^M_\alpha \tag{5.2}$$

is closed and the conserved quantity corresponding to the vector field B is given by

$$\int_{\mathscr{R}} \tau^M(B) = b^\beta(s')p_\beta(s'). \tag{5.3}$$

According the Dixon,[6,7] this is an important property which a correct definition of n-momentum must have.

In order to complete the definition (5.1), we have to specify the integration surface $\mathscr{R}(s')$. A natural choice is given by the equations

$$\begin{aligned} &\xi^\alpha(s,s') = 0, \qquad \alpha \neq 1,2,3, \\ &|\xi^\alpha(s,s')| \leqslant l, \qquad \alpha = 1,2,3. \end{aligned} \tag{5.4}$$

In other words, the points of $\mathscr{R}(s')$ are obtained by means of spatial parallel displacements of the point s'. Note that Eq. (5.4) defines a three-dimensional cube in the space of the coordinates ξ^α and that Eq. (4.1) gives a differentiable mapping of this cube onto $\mathscr{R}(s')$, namely we have actually defined a singular cubic symplex.[25]

We assume that there is an open set $\mathscr{U}'$, which contains $\mathscr{U}$ and has the following property: If $s' \in \mathscr{U}'$, one can choose the parameter l in such a way that the boundary of $\mathscr{R}(s')$ is contained in $c\mathscr{U}$, namely $\mathscr{R}(s')$ is a relative cycle. We assume also that all the relative cycles obtained in this way belong to the same relative homology class, which is by definition the "right" one. The quantities $p_\alpha(s')$ are defined only for $s' \in \mathscr{U}'$.

Now we have to generalize Eq. (3.10). The expression (5.1) depends on s' through the function under the integral and through the integration surface. If we describe the displacement of s' and of the generic point s of $\mathscr{R}(s')$ by means of the infinitesimal vector field

$$\eta^\theta A_\theta, \tag{5.5}$$

using Stoke's formula, we get

$$\delta p_\alpha = \eta^\theta(s')\int_{\mathscr{R}(s')} L'_\theta D^\beta_\alpha(s,s')\tau^M_\beta + \int_{\mathscr{R}(s')} \eta^\theta(s) i_\theta d\,[D^\beta_\alpha(s,s')\tau^M_\beta]. \tag{5.6}$$

From the definition of $\mathscr{R}(s')$ we have

$$\eta^\theta(s)L_\theta\xi^\alpha(s,s') + \eta^\theta(s')L'_\theta\xi^\alpha(s,s') = 0, \tag{5.7}$$

and Eq. (4.8) gives

$$\eta^\theta(s) = D^\theta_\rho(s,s')\eta^\rho(s'). \tag{5.8}$$

From Eqs. (2.19) and (4.13) we have

$$d\,[D^\beta_\alpha(s,s')\tau^M_\beta] = C^\beta_{\theta\gamma}(s,s')D^\gamma_\alpha(s,s')\omega^\theta\wedge\tau^M_\beta - D^\beta_\alpha(s,s')\frac{\partial\lambda^M}{\partial F^\rho_{\mu\nu}}L_\beta F^\rho_{\mu\nu}. \tag{5.9}$$

If we use Eqs. (4.16), (5.8), and (5.9), Eq. (5.6) becomes

$$\delta p_\alpha = \eta^\theta(-F^\beta_{\theta\alpha}p_\beta + f_{\theta\alpha}), \tag{5.10}$$

where

$$f_{\theta\alpha}(s') = \int_{\mathscr{R}(s')}[D^\sigma_\theta(s,s')C^\beta_{\sigma\gamma}(s,s')D^\gamma_\alpha(s,s') - D^\beta_\gamma(s,s')C^\gamma_{\theta\alpha}(s',s)]\tau^M_\beta - \int_{\mathscr{R}(s')} D^\sigma_\theta(s,s')C^\beta_{\rho\gamma}(s,s')D^\gamma_\alpha(s,s')\omega^\rho\wedge i_\sigma\tau^M_\beta - \int_{\mathscr{R}(s')} D^\sigma_\theta(s,s')D^\beta_\alpha(s,s')L_\beta F^\rho_{\mu\nu}i_\sigma\frac{\partial\lambda^M}{\partial F^\rho_{\mu\nu}}. \tag{5.11}$$

Equation (5.10) can also be written in the form

$$L_\theta p_\alpha = -F^\beta_{\theta\alpha}p_\beta + f_{\theta\alpha}, \tag{5.12}$$

which generalizes Eq. (3.10). The term $f_{\theta\alpha}$ is a correction which appears when the structure coefficients $F^\gamma_{\alpha\beta}$ are not constant.

The equations given in the present section are exact, and we have not assumed that $\mathscr{S}$ is a principal fiber bundle and that the structure coefficients have the property (2.6). The most delicate assumption concerns the properties of the intersection of the surfaces $\mathscr{R}(s')$ and the region $\mathscr{U}$. The definition of $\mathscr{R}(s')$ is the only point of our argument in which some of the vectors fields A_α play a special role.

VI. THE POLE–DIPOLE APPROXIMATION FOR A TEST PARTICLE

Now we disregard the geometric fields generated by the particle itself, and we assume that the external geometric fields vary smoothly in the region $\mathscr{U}$, in such a way that if s' is in $\mathscr{U}$ or at least near to it, we can replace all the geometric quantities in Eqs. (5.1) and (5.11) by a few terms of their power expansion in the normal coordinates $\xi^\alpha(s,s')$. In the pole–dipole approximation we disregard the squares and higher powers of the quantities ξ^α and in the integrals which describe the nonminimal couplings we disregard also the terms linear in ξ^α.

If we use Stokes' formula and Eq. (5.9), we can easily show that in the pole–dipole approximation the quantities p_α defined by Eq. (5.1) are not affected by a small deformation of the surface $\mathscr{R}(s')$ which leaves the point s' fixed. If we use Eq. (4.17), we get the following pole–dipole approximation of Eq. (5.1):

$$p_\alpha = \int_{\mathscr{R}(s')}\tau^M_\alpha + F^\beta_{\rho\alpha}\int_{\mathscr{R}(s')}\xi^\rho\tau^M_\beta. \tag{6.1}$$

The last term describes the orbital angular momentum, the contribution of the electric dipole moment to energy and momentum in the presence of an electromagnetic field and other less familiar contributions. We see that all these terms have a common origin, namely the noncommutativity of the transformations generated by the fields A_α.

With the same approximation, we can use Eqs. (4.7), (4.17), and (4.18) to write Eq. (5.11) in the form

$$f_{\theta\alpha} = \tfrac{1}{2}L_\alpha F^\beta_{\mu\nu}m^{\mu\nu}_{\beta\theta}, \tag{6.2}$$

where

$$m^{\mu\nu}_{\beta\theta} = \int_{\mathscr{R}(s')}(\xi^\mu\delta^\nu_\theta - \xi^\nu\delta^\mu_\theta)\tau^M_\beta - \tfrac{1}{2}\int_{\mathscr{R}(s')}(\xi^\mu d\xi^\nu - \xi^\nu d\xi^\mu)\wedge i_\theta\tau^M_\beta - 2\int_{\mathscr{R}(s')} i_\theta\frac{\partial\lambda^M}{\partial F^\beta_{\mu\nu}}. \tag{6.3}$$

The last equation defines the dipole moments in the pole–dipole approximation. If we consider higher multipoles, we have to use also a more accurate definition of the dipole moments. A consistent set of reduced multipole moments could be obtained by means of the methods developed by Dixon.[6]

The last integral in Eq. (6.3) is the contribution of the nonminimal couplings. If this term is absent and we use the local expression (2.20) for the differential forms τ^M_α, we obtain

$$m^{ik}_{\beta l} = \delta^0_l m^{ik}_\beta + \delta^k_l m^{i0}_\beta - \delta^i_l m^{k0}_\beta, \tag{6.4}$$

$$m^{ik}_{\beta a} = 0, \tag{6.5}$$

where

$$m^{ik}_\beta = \tfrac{1}{2}\int_{\mathscr{R}(s')}(\xi^i T^k_\beta - \xi^k T^i_\beta)\,d\xi^1 d\xi^2 d\xi^3. \tag{6.6}$$

Note that $\xi^0 = 0$ on the integration surface.

If $\mathscr{S}$ is a principal fiber bundle and we use Eqs. (2.6), (6.2), and (6.5), Eq. (5.12) gives

$$L_a p_\alpha = -\hat{F}^\beta_{a\alpha}p_\beta. \tag{6.7}$$

This equation means that the 4-momentum and the relativistic angular momentum are gauge-invariant Lorentz tensors and the charges are Lorentz scalars which transform according to the coadjoint representation of the internal gauge group. In a similar way, using also Eq. (2.8), we obtain

$$L_i p_k = F^\beta_{ki}p_\beta + \tfrac{1}{2}L_k F^\beta_{rs}m^{rs}_{\beta i}, \tag{6.8}$$

$$L_i p_a = \hat{F}^k_{ai}p_k + \tfrac{1}{2}(\hat{F}^\beta_{a\alpha}F^\alpha_{rs} - \hat{F}^j_{ar}F^\beta_{js} - \hat{F}^j_{as}F^\beta_{rj})m^{rs}_{\beta i}. \tag{6.9}$$

Note that the dipole moments defined by Eqs. (6.4) and (6.6) cannot transform as the components of a Lorentz tensor. In order to avoid a contradiction with Eqs. (6.7)–(6.9), we have to remember that these approximate equations are valid only near the central frames and one is not allowed to equate all the derivatives of their left- and right-hand sides.

VII. THE EQUATIONS OF MOTION

In order to discuss the motion of the particle, we have to consider with more detail the conditions (3.11) which define its center. A natural definition of the center-of-energy is

$$\int_{\mathscr{P}(s')} \xi^i \tau_0^M = 0, \quad i = 1,2,3. \tag{7.1}$$

If the energy density is positive, the center-of-energy lies inside the particle. If there is no spin density, from Eq. (6.1) we see that these conditions are equivalent to the conditions

$$p_7 = p_8 = p_9 = 0. \tag{7.2}$$

If we describe the angular momentum by means of the antisymmetric matrix p_{ik} defined by

$$p_a = \tfrac{1}{2}\hat{F}^i_{al} g^{lk} p_{ik}, \quad a = 4,...,9, \tag{7.3}$$

the conditions (7.2) can also be written in the form

$$p_{i0} = 0. \tag{7.4}$$

If there is a nonvanishing spin density, the conditions (7.1) are no longer equivalent to Eq. (7.2). Since the conditions (7.1) do not have the general form (3.11), which permits the decoupling of the motion of the center from the internal motions in the absence of external geometric fields, we have to adopt the conditions (7.2). In the absence of geometric fields, we can use Eq. (3.9) and see that, unless all the quantities p_i vanish, the conditions (7.2) define a $(n-3)$-dimensional manifold $\mathscr{V}_c$. This remains true as long as the geometric fields are sufficiently small, in a sense which should be stated more precisely. However, the center does not necessarily lie inside the particle.

Note that the conditions (7.4) are not Lorentz-invariant, and from Eq. (6.7) we see that if they are satisfied at a point s, in general they are not satisfied at the near points of the fiber which passes through s. In other words, the position of the center depends on the velocity of the observer, a fact which is well known in special relativity.[26] In order to obtain a well-defined trajectory in space–time, it is convenient to consider the zero-momentum central frames which are defined by the conditions

$$p_1 = p_2 = p_3 = p_7 = p_8 = p_9 = 0 \tag{7.5}$$

and form the $(n-6)$-dimensional manifold $\mathscr{V}_{c0}$. Equation (7.5) implies the Lorentz invariant conditions

$$p_{ik} p^k = 0, \tag{7.6}$$

which have been advocated by Tulczyjev[3] and by Dixon.[4–7] They can be used instead of Eq. (7.2) to give a different definition of the manifold $\mathscr{V}_c$.

If we are dealing with a spinning particle, we can choose a still smaller class of reference frames, in which the spin is directed along the z axis, namely we have, besides the conditions (7.5),

$$p_4 = p_5 = 0. \tag{7.7}$$

These "zero-momentum central directed frames" form an $(n-8)$-dimensional submanifold $\mathscr{V}_{c0d}$, which has been used by Künzle[16] to describe the motion of the particle.

In the pole approximation, Eq. (5.12) coincides with Eq. (3.10), which has been derived in the case of constant structure coefficients. In this case, Eq. (5.12), together with the conditions discussed above and with suitable initial conditions, determines the motion of the center of the particle and the n-momenta p_α. This is possible because these variables are not coupled with the internal degrees of freedom.

In the general case all the infinite degrees of freedom of the system are coupled and an exact determination of its motion is equivalent to the solution of the matter field equations. In order to get a useful particle theory, we have to introduce some approximation, in which the system can be described by means of a finite and small number of variables x_r. For instance, we can consider an adiabatic approximation, namely we can assume that all the other variables necessary for the complete description of the system vary very rapidly in such a way that all the interesting quantities can be replaced by their averages, which depend only on the variables x_r.

We do not try to formulate this assumption in a more rigorous way, since we are more interested in the corresponding quantum situation, in which most degrees of freedom are "frozen," namely they have a high frequency and cannot be excited with the available energy. We assume that the degrees of freedom x_r, which are not frozen, can be treated by means of classical mechanics. This is the approximation which gives a very good description of a monoatomic gas in normal conditions.

In the simplest case, the variables x_r are the coordinates of the center of the particle and the components p_α of n-momentum, which satisfy one of the conditions discussed above. Then the quantities $f_{\theta\alpha}$ or $m^{\mu\nu}_{\beta\theta}$ must be interpreted as average values which depend only on the quantities p_α and on the geometric fields near the center of the particle. For instance, the average magnetic dipole moment can be given by means of a constant gyromagnetic ratio. When these average values are known, the equations given in the preceding sections determine the average motion of the particle completely.

In this simple approximation, also the rotational degrees of freedom are averaged. If we want to treat them in detail as classical degrees of freedom, we have to use also other equations, which can be derived, for instance, from an assumption of "dynamical rigidity."[6]

We have seen that if the geometric fields are weak, the exact equation (5.12) together with the conditions (7.2) or (7.6) defines the manifold $\mathscr{V}_c$. The introduction of the multipole and of the adiabatic approximations may spoil the integrability conditions of these equations in such a way that the approximate system has no $(3-n)$-dimensional integral surfaces. In this case it is necessary to consider only the reference frames defined by the stronger conditions (7.5) or (7.7). Another advantage of this limitation is the simpler form of the phenomenological formulae which give the average multipole moments. A detailed discussion of these formulae will be given elsewhere.

VIII. CONCLUSION

Starting from the exact equations obtained in Sec. V, we have obtained a sufficiently simple approximate model for the motion of a test particle in external geometric fields. It is

based on Eqs. (6.8) and (6.9), on the conditions (7.5) which define the zero-momentum central frames and on the phenomenological equations which give the average values of the dipole moments as functions of the quantities p_α and of the structure coefficients F^α_{ik}. These equations determine the manifold $\mathscr{V}_{c0}$.

We see from Eqs. (6.7) and (7.5) that $\mathscr{V}_{c0}$ is invariant under rotations and transformations of the internal symmetry group. As a consequence, its projection on the space–time $\mathscr{M}$ is a line. We parametrize this line by means of the variable t, and we choose in a smooth way a zero-momentum central frame $s(t)$ at every point of the line. Then every function defined on $\mathscr{S}$ defines a function of t. We indicate by $v^i(t)$ the components of the velocity vector with respect to the frame $s(t)$ and by δ/dt the covariant derivative with respect to t. We remember[19] that the operators L_i applied to the components of a tensor field give the components of the covariant differential of the field. If the field transforms nontrivially under the internal gauge group, these covariant derivatives contain also the contribution of the internal gauge potentials.

Then from Eqs. (6.8) and (6.9) we obtain

$$\frac{\delta p_k}{dt} = F^\beta_{ki} v^i p_\beta + \tfrac{1}{2} D_k F^\beta_{rs} m^{rs}_{\beta i} v^i, \tag{8.1}$$

$$\frac{\delta p_a}{dt} = \hat{F}^k_{ai} p_k v^i + \tfrac{1}{2}(\hat{F}^\beta_{a\alpha} F^\alpha_{rs} - \hat{F}^j_{ar} F^\beta_{js} - \hat{F}^j_{as} F^\beta_{rj}) m^{rs}_{\beta i} v^i. \tag{8.2}$$

These equations are covariant with respect to a different choice of the frames $s(t)$ if the equations which define the dipole moments $m^{rs}_{\beta i}$ are covariant with respect to rotations and internal symmetry transformations. If these equations are Lorentz-covariant, Eqs. (8.1) and (8.2) can also be considered as Lorentz-covariant equations, but then the quantities $m^{rs}_{\beta i}$ are not the ones defined by Eqs. (6.4) and (6.6). In this case, the conditions (7.5) must be replaced by the Lorentz-invariant conditions (7.6).

Equations (8.1) and (8.2) can easily be compared with the results of other authors. If one considers only gravitational and electromagnetic fields, in the absence of torsion and of spin density, these formulae agree with the results of Dixon[6] and of Souriau.[8] The isospin pole term is the one described by Wong[11]; the isospin dipole terms have been discussed in Refs. 27 and 28. If we admit the existence of torsion and of spin density, there is one more pole term, which agrees with the one proposed in Refs. 9 and 10 and several "gravitational" dipole terms, which have been considered by Yasskin and Stoeger.[29] It has been shown[30,31] that a Dirac elementary particle has a gravitational dipole, similar to its magnetic dipole.

Yasskin and Stoeger[29] have discussed also the difficulties due to the ambiguities of the definition of the 4-momentum and of the center of a particle. We think that in the absence of torsion and of spin density the work of Dixon has indicated the definitions which are more convenient at least from the theoretical point of view. In the present paper we suggest that Dixon's arguments can be applied also in more general cases.

ACKNOWLEDGMENTS

It is a pleasure to thank G. Cognola, R. Soldati, L. Vanzo, and S. Zerbini for several useful suggestions and discussions.

[1] M. Mathisson, Acta Phys. Pol. **6**, 163 (1937).
[2] A. Papapetru, Proc. R. Soc. London A **209**, 248 (1951).
[3] W. Tulczyjev, Acta Phys. Pol. **18**, 393 (1959).
[4] W. G. Dixon, Nuovo Cimento **34**, 317 (1964).
[5] W. G. Dixon, J. Math. Phys. **8**, 1591 (1967).
[6] W. G. Dixon, Proc. R. Soc. London A **314**, 499 (1970); **319**, 509 (1970); Phil. Trans. R. Soc. London Ser. A **277**, 59 (1974).
[7] W. G. Dixon, in *Proceedings of the International School of Physics "Enrico Fermi"*, Course 67, edited by J. Ehlers (North–Holland, Amsterdam, 1979).
[8] J. M. Souriau, Ann. Inst. Henri Poincaré A **20**, 315 (1974).
[9] F. W. Hehl, Phys. Lett. A **36**, 225 (1971).
[10] A. Trautman, Bull. Acad. Pol. Sci. **20**, 895 (1972).
[11] S. K. Wong, Nuovo Cimento A **65**, 689 (1970).
[12] A. Trautman, Rep. Math. Phys. **1**, 29 (1970).
[13] Y. M. Cho, J. Math. Phys. **16**, 2029 (1975); Phys. Rev. D **14**, 2521, 3335, 3341 (1976).
[14] L. N. Chang, K. Macrae, and F. Mansouri, Phys. Lett. B **57**, 59 (1975); Phys. Rev. D **13**, 235 (1976).
[15] F. Mansouri and L. N. Chang, Phys. Rev. D **13**, 3129 (1976).
[16] H. P. Künzle, J. Math. Phys. **13**, 739 (1972).
[17] S. Sternberg, Proc. Natl. Acad. Sci. USA **74**, 5253 (1977).
[18] R. Giachetti, R. Ricci, and E. Sorace, J. Math. Phys. **22**, 1703 (1981).
[19] M. Toller, Nuovo Cimento B **44**, 67 (1978).
[20] G. Cognola, R. Soldati, L. Vanzo, and S. Zerbini, J. Math. Phys. **20**, 2613 (1979).
[21] G. Cognola, R. Soldati, M. Toller, L. Vanzo, and S. Zerbini, Nuovo Cimento B **54**, 325 (1979).
[22] R. Soldati and S. Zerbini, Lett. Nuovo Cimento **27**, 575 (1980).
[23] Y. Choquet-Bruhat, *Géométrie différentielle et systèmes extérieurs* (Dunod, Paris, 1968).
[24] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1969).
[25] W. S. Massey, *Singular Homology Theory* (Springer, Berlin, 1980).
[26] C. Møller, *The Theory of Relativity* (Oxford U.P., Oxford, 1972).
[27] A. P. Balachandran, P. Salmonson, B. S. Skagerstam, and J. O. Winnberg, Phys. Rev. D **15**, 2308 (1977).
[28] Ch. Duval, preprint, centre de Physique Théorique CNRS, Marseille, 1978.
[29] P. B. Yasskin and W. R. Stoeger, Phys. Rev. D **21**, 2081 (1980).
[30] H. Rumpf, in *Cosmology and Gravitation*, edited by P. G. Bergmann and V. De Sabbata (Plenum, New York, 1980).
[31] J. Audretsch, Phys. Rev. D **24**, 1470 (1981).

Boundary effects in the dimer problem on a non-Bravais lattice

D. Grensing and G. Grensing
Universität Kiel, Fachbereich Physik, Kiel, West Germany

(Received 23 July 1981; accepted for publication 27 August 1982)

The dimer problem on the hexagonal lattice is solved for various boundary shapes. The number of arrangements of dimers in the asymptotic limit of large lattices proves to be very sensitive to the precise form of the lattice boundary. If N is the number of lattice points and (for large N) $W^{N/2}$ the number of dimer configurations, then the molecular freedom W reduces to 1 in the case of standard boundaries (parallelogram, rectangle), and there is, in the thermodynamic limit, no freedom in placing a dimer at all. But in the case of a hexagonal lattice which has been given the shape of a macroscopic honeycomb and which thus exhibits the symmetry properties of the unit cell, the molecular freedom is found to tend to $W = \frac{3}{4}\sqrt{3} = 1.299$ for $N \to \infty$. Investigating various other boundary forms, the regular hexagon proves to be a convenient surface shape with a maximum molecular freedom which differs markedly from the known value $W = 1.381$ for the hexagonal lattice wrapped on a torus.

PACS numbers: 05.50. + q

I. INTRODUCTION

Since the discovery of the Pfaffian method as an elegant tool for solving the dimer problem on the square lattice,[1-4] the number of arrangements of dimers can be evaluated in closed form for all two-dimensional lattices and even some exotic lattices, e.g., the Kagomé lattice and the Fisher lattice.[5] A common feature of all known solutions on planar lattices is that the number of dimer configurations Z_N on a lattice with N vertices is of the order e^N, and hence it is convenient to define the number of configurations per dimer

$$W = \lim_{N\to\infty} Z_N^{2/N}, \tag{1}$$

which is called the molecular freedom.[6] As a representative example for the known 2D solutions we only quote the result for the honeycomb lattice[7-9]

$$W = \exp\left\{(2\pi)^{-2} \int_0^{2\pi}\int_0^{2\pi} \ln[\phi(e^{i\theta_1}, e^{i\theta_2})]\, d\theta_1\, d\theta_2\right\} = 1.3814, \tag{2}$$

where $\phi(x, y)$ denotes the generating function for random walks on the hexagonal lattice.[10] This relationship indicates that the dimer problem is connected intimately with the random walk problem, but this is not the point. What is of interest is the fact that the derivation of (2) requires the assumption of periodic boundary conditions. Further, the known analytic methods available for the solution of the dimer problem[1,3,11,12] prove ineffective, if not inapplicable, in the case of non-Bravais lattices with free boundary conditions.

At first sight one is tempted to believe that the exact nature of boundary conditions has little influence on the solution of the dimer problem in the thermodynamic limit; hence it might be thought that (2) also represents a solution of the hexagonal lattice with free edges. For comparison, in the case of the square lattice it has been shown explicitly that the molecular freedom is insensitive to the precise form of the boundary conditions in the limit of a large lattice.[1] But on the hexagonal lattice we are faced with the phenomenon that boundary effects cannot be neglected, and the application of cyclic boundary conditions is no longer justified.

To demonstrate the restrictive influence of the boundary of the non-Bravais lattice at hand, we study the special case of a honeycomb lattice which has been given the shape of a parallelogram (see Fig. 1) and which anticipates characteristic features of the lattice forms which will be dealt with in the following sections. The two opposed sited vertices with coordination number 1 (marked as circles in Fig. 1) are by no means superfluous lattice sites, but have been introduced for two reasons. First, both vertices are necessary if one requires that the lattice is composed of unit cells. Secondly, they ensure the possibility of a periodic repetition of the lattice and thus are necessary if one wants to employ periodic boundary conditions.

Assuming the lattice to be wound on a torus, we find that Eq. (2) is true, and in the case of free boundary conditions we find $Z_N = 1$, independent of the lattice size. The latter result is obvious because a dimer, placed on the lattice such that it covers one of the two exceptional points with coordination number 1, generates another lattice site with one exit only, and so forth. Thus, if one starts to place dimers on vertices with a minimal coordination number, one enforces the occupation of the whole vertical boundary line,

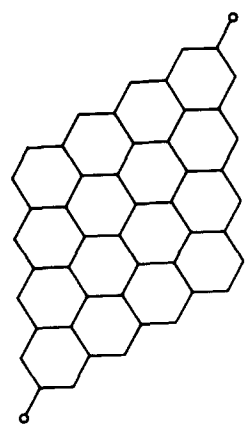


FIG. 1. Hexagonal lattice with two additional lattice sites necessary for the device of wrapping the lattice on a torus.

 0022-2488/83/030620-11$02.50

and the resulting lattice with unoccupied vertices exhibits the same structure as the lattice we started with. Repeating this procedure, one finds that the lattice in Fig. 1 permits only a single dimer configuration.

In view of this drastic reduction of the molecular freedom, one is compelled to provide the hexagonal lattice with a less restrictive boundary in order to obtain a nontrivial dimer problem. But the evaluation of the generating function for dimer configurations in a closed form, using for example the Pfaffian method on those lattices with free edges, is in general a difficult problem. For certain domains, being regular in the sense of van Hove,[13,14] however, it is possible to carry through a rigorous analysis. The problems we pose are the following:

1. Does any regular domain exist which permits a nontrivial molecular freedom on the hexagonal lattice ($W > 1$)?

2. Is W the same for free and periodic boundary conditions, provided one finds such a domain?

Our interest in these questions is not entirely academic and arises from the natural discomfort when using periodic boundary conditions. Actually, one must admit that cyclic boundary conditions cannot be achieved physically. In two dimensions a network of honeycomb cells drawn on the surface of a torus is still imaginable, but for a three-dimensional lattice the imagination breaks down and, indeed, the cyclic boundary conditions cannot be satisfied by any topological contortions.

In view of the fact that the natural boundary condition to be imposed physically is the free one, it seemed worthwhile to try to find the molecular freedom on the hexagonal lattice avoiding the trick of winding the lattice on a torus. Investigating a variety of possible boundary shapes of the hexagonal lattice, we discover that the molecular freedom given by Eq. (2) is unattainable in a lattice with free edges. Our result is that the greatest molecular freedom is to be found on a lattice shaped as a regular hexagon, the numerical value being

$$W = \tfrac{3}{4}\sqrt{3} = 1.29904 \tag{3}$$

in the limit of a large lattice. This obvious reduction of the molecular freedom is due to the rough lattice boundary, the restrictive influence of which survives when performing the limiting process $N \to \infty$ and would have been suppressed by using cyclic boundary conditions in order to get rid of the difficulties brought about by a cumbersome surface.

II. RECTANGULAR DOMAINS

The preliminary remarks in Sec. I indicate that the two vertices with the minimal coordination number 1 on the hexagonal lattice shown in Fig. 1 enforce a single dimer configuration. Detaching these two vertices responsible for this drastic restriction of possible dimer coverings leads to another trivial dimer problem on the hexagonal lattice, with the result

$$Z_N = \binom{2L}{L}, \tag{4}$$

where N denotes the number of lattice sites of a lattice composed of $L \times L$ honeycombs. Thus N depends on L via

$$N = 2(L+1)^2 - 2. \tag{5}$$

The number of dimer configurations given by Eq. (4) proves to be a special case of a general solution of the dimer problem on domains which will be studied in detail in the next section. What is important for the moment is the fact that the binomial representation of Z_N in Eq. (4) entails $W = 1$ in the limit $N \to \infty$. Thus the hexagonal lattice with the shape of a parallelogram must be excluded as a candidate for a lattice with a nonvanishing molecular freedom. This should not cause any surprise in view of the fact that even on a square lattice with a similar boundary shape there is no freedom in placing a dimer when $N \to \infty$.[15] This property and the known solutions on the $n \times m$ square lattice suggest that the rectangular domain permits much larger numbers of dimer coverings even in the case of a hexagonal structure.

The lattice shape we intend to study in this section is shown in Fig. 2. If this lattice consists of $L_1 \times L_2$ honeycombs (L_1 horizontally and L_2 vertically arranged cells), the number of vertices is

$$N = 2(L_1 L_2 + L_1 + L_2). \tag{6}$$

Let $Z(L_1, L_2)$ denote the number of ways of covering the lattice with $N/2$ dimers. This quantity can be evaluated quite simply if one exploits the fact that horizontal dimers placed at the boundary induce a mechanism observed already on the lattice discussed in Sec. I (cf. Fig. 1). To illustrate this mechanism, we begin to place a first dimer and choose the vertex marked as a circle in Fig. 2 as a starting point. The coordination number of this boundary site is 2; thus it has two next-neighbor bonds which will be called exits. Being interested in the total number of possible dimer configurations, we must study the subsequent covering history for both cases, one with a dimer placed horizontally [cf. Fig. 2(a)] and the other with a dimer occupying the edge representing the second exit [Fig. 2(b)].

The horizontal dimer generates a vertex with one exit only. In order to avoid the subsequent covering of the lattice constituting isolated lattice sites, which are usually called monomers, it appears convenient to place dimers immediately wherever such vertices with one exit appear. For these dimers placed under constraint we introduce the notation c-dimers. Setting the first c-dimer there emerges another ver-

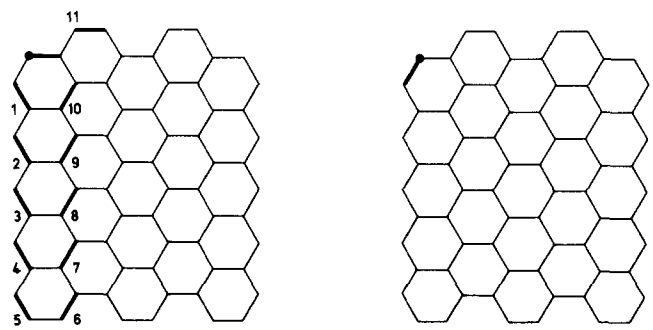


FIG. 2. Hexagonal lattice of rectangular shape where the occupation of a boundary site (marked by a circle) with a first dimer is studied. The bonds marked as heavy lines in Fig. 2(a) represent c-dimers and are numbered consecutively.

tex which necessitates placing a second c-dimer and so on. In Fig. 2(a) these c-dimers are numbered consecutively from 1 to $2L_2 + 1$ in order to illustrate the sequence of events. The resulting lattice is seen to be a $(L_1 - 2)\times L_2$ lattice, which is again of rectangular shape and permits $Z(L_1 - 2,L_2)$ dimer coverings.

The second possibility in starting with a first dimer [see Fig. 2(b)] has no such spectacular consequences and requires a repetition of the procedure described above. To this end, we choose a new starting point for a second dimer. This vertex being a boundary site with the low coordination number 2, there are again two possibilities: (a) a horizontal dimer which actuates an avalanche of c-dimers [see Fig. 3(a)]; (b) a dimer which covers a boundary edge [see Fig. 3(b)] and gives rise to the choice of a new starting point for a third dimer in order to bring this procedure to an end.

Possibility (a) results in a lattice which can be thought of as being composed of $(L_1 - 2)\times L_2$ honeycombs and one additional cell, which spoils the rectangular shape. This is a special case of the lattice shown in Fig. 4. Let $Z(k,L_1 - 1,L_2)$ denote the number of ways of covering this imperfect lattice, where k indicates the existence of only k cells in the first column of honeycombs ($k < L_2$).

The continuation of the procedure indicated in Figs. 2 and 3 is now apparent. At every stage of this procedure there are two cases to be considered, one of which yields a contribution $Z(k,L_1 - 1,L_2)$ ($k = 1,2,...,L_2 - 1$) and the second of which requires a continuation of the procedure provided $k < L_2$.

After L_2 steps of this kind, the procedure terminates in a natural way because of an enforced horizontal dimer. Placing this c-dimer, one is faced with a lattice the left vertical boundary of which is completely filled with dimers. Thus the total number of dimer configurations on the lattice we started with is given by the sum

$$Z(L_1,L_2) = Z(L_1 - 2,L_2) + \sum_{k=1}^{L_2 - 1} Z(k,L_1 - 1,L_2) + Z(L_1 - 1,L_2), \tag{7}$$

and we are left with the more general problem of finding the number of configurations on the imperfect lattice in Fig. 4. Obviously the first and third term on the right side of Eq. (7) are special cases of the general combinatorial quantity

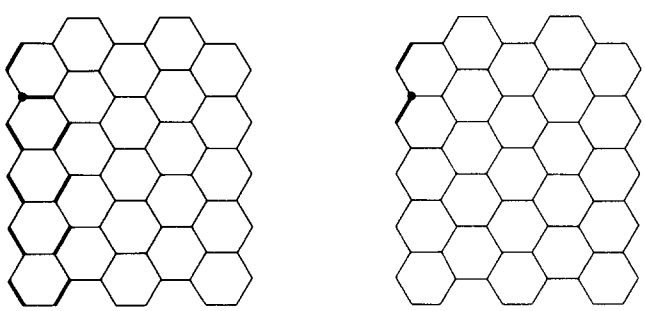


FIG. 3. The possible cases of placing a second dimer on the lattice shown in Fig. 2(b).

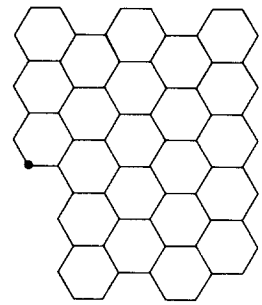


FIG. 4. "Incomplete" hexagonal lattice as a result of a partial covering of the hexagonal lattice, the covering history of which is shown in Figs. 2 and 3.

$Z(k,L_1 - 1,L_2)$ with $k = 0$ and $k = L_2$, respectively. Adopting the useful convention

$$Z(k,L_1 - 1,L_2) = \begin{cases} Z(L_1 - 2,L_2) & \text{for } k = 0, \\ Z(L_1 - 1,L_2) & \text{for } k = L_2, \end{cases} \tag{8}$$

both terms can be incorporated into the sum in Eq. (7) which now runs from 0 to L_2.

As matters stand, it will be necessary to solve the general problem of calculating $Z(k,L_1,L_2)$ rather than the original dimer problem itself. Fortunately, this quantity satisfies a simple recurrence relation, namely

$$Z(k,L_1,L_2) = Z(k - 1,L_1,L_2) + Z(L_2 - k,L_1 - 1,L_2), \tag{9}$$

which is easily verified if one places alternatively a first dimer on the imperfect lattice such that it covers the vertex marked as a circle in Fig. 4.

The obvious advantage of Eq. (9) is that by iteration it leads to

$$Z(k,L_1,L_2) = \sum_{l = L_2 - k}^{L_2} Z(l,L_1 - 1,L_2), \tag{10}$$

where use has been made of the definition $Z(0,L_1,L_2) = Z(L_2,L_1 - 1,L_2)$ [cf. Eq. (8)]. For $k = L_2$ this basic recurrence relation reproduces Eq. (7) which serves to determine the required number $Z(L_1,L_2)$.

To get the feel of things, we start with the case $L_1 = 1$ and obtain the trivial result

$$Z(k,1,L_2) = Z(1,k) = k + 1, \tag{11}$$

which represents the number of dimer configurations on a linear chain of k honeycombs. Substituting this result into Eq. (10), where L_1 is set equal to 2, yields for $Z(k,2,L_2)$ a sum of binomial coefficients

$$Z(k,2,L) = \binom{L+2}{2} - \binom{L+1-k}{2}. \tag{12}$$

In deriving (12) use is made of the identity

$$\sum_{l=0}^{L} \binom{n+l}{n} = \binom{n+L+1}{n+1}, \tag{13}$$

which ensures that in the following procedure of mathematical induction the quantity $Z(k,L_1,L_2)$ for arbitrary L_1 will be expressible in terms of binomial coefficients. We omit the details here and give only the result for perfect lattices ($k = L_1$) with even L_1 ($= 2n$) and with arbitrary L_2 ($= L$):

$$Z(2n,L) = \sum_{k=0}^{n-1} (-1)^k (n-k)! \times \sum_{\pi(n)} \prod_{i=1}^{\nu} (r_i!)^{-1} \binom{L+n_i+1}{2n_i}, \quad (14)$$

where $\pi(n)$ stands for a partition of n, usually denoted by $n_1^{r_1} n_2^{r_2} \cdots n_\nu^{r_\nu}$, with the restriction $\Sigma_1^\nu r_i = n - k$.

Using mathematical induction, it may be shown that (14) is the only solution of (10) which satisfies the boundary condition $Z(k,0,L) = 1$ and the conditions summarized in Eq. (8). Analogously one finds a similar expression for $Z(2n+1,L)$. But it must be admitted that the representation of the solution of the dimer problem given by (14) proves to be inefficient if one wants to present numerical values for finite but large lattices.

A more convenient method for evaluating $Z(L,L)$ for large L is based on the recurrence relation (10) itself. Starting with the formula for a linear chain of honeycombs [cf. Eq. (11)], one finds by a simple procedure of adding, according to (10), the dimer coverings of imperfect lattices which consist of two columns of honeycombs. By repeated application of (10) one works through the lattice column by column from right to left and finally finds, at the bottom of the L_2th column, the quantity $Z(L_1,L_2)$, if L_1 is held fixed during the whole procedure. This concept provides, in conjunction with the definition in the second of Eqs. (8), an easy way to compute the total number of dimer coverings on a rectangular $L_1 \times L_2$ hexagonal lattice. For numerical concreteness, a continuable array of numbers $Z(k,n,6)$ $(k<n)$, a modified version of the Pascal scheme, is given in Table I. Note how easily one column may be filled in from the next on the right using relation (10). The number which is underscored in Table I represents $Z(6,6) = 15\,106$. Analogously one finds that for square lattices with $L = 1,2,\ldots,5$ the number of dimer coverings is equal to 2, 6, 30, 190, 1547.

The quantity which is then of primary interest is the molecular freedom defined by

$$W_N = Z_N^{2/N}, \quad (15)$$

where $Z_N \equiv Z(L_1,L_2)$ and N denotes the number of lattice sites of an $L_1 \times L_2$ lattice given by Eq. (6).

In Fig. 5 we have summarized the numerical values of W_N on square lattices. The steep descent of the curve, which connects the discrete values of W_N for $N \to \infty$, indicates that the molecular freedom reduces to 1 in the thermodynamic limit.

It is again the recurrence formula (10) which enables us to give a rigorous proof. Setting $k = L_2$ and replacing each term on the right of (10) by $Z(L_2,L_1-1,L_2) \equiv Z(L_1-1,L_2)$, one finds the inequality

$$Z(L_2,L_1,L_2) \equiv Z(L_1,L_2) < (L_2+1) Z(L_1-1,L_2), \quad (16)$$

and, if this is repeated until the appearance of $Z(0,L_2)$, which is unity, one obtains the following upper bound for the number of dimer coverings of a lattice composed of $L_1 \times L_2$ honeycombs:

$$Z(L_1,L_2) < (L_2+1)^{L_1}. \quad (17)$$

Hence, in the limit of a large lattice, the upper bound behaves as $\exp(L_1 \ln L_2)$. Combining this with the asymptotic version of Eq. (6) leads to the inquality for the quantity required:

$$W_N < \exp(L_2^{-1} \ln L_2), \quad N \sim 2L_1L_2 \to \infty, \quad (18)$$

which predicts that the molecular freedom is bound to approach the limiting value $\lim_{N\to\infty} W_N = 1$ no matter how the edge length L_1 goes to infinity.

To sum up, we succeeded in solving the dimer problem on a hexagonal lattice with rectangular shape. With this choice of a free boundary the number of dimer coverings is found to be augmented when compared with that on an equal-sized lattice which is formed like a parallelogram [cf. Eq. (4)]. However, in the thermodynamic limit, the rough surface leads in both cases to the same drastic reduction of the molecular freedom. The meaning of $W = 1$ is hard to grasp, the only interpretation being that one finds exactly one possibility to place a dimer, wherever on the infinite lattice, whenever in the course of the history of covering. This striking feature of the hexagonal lattice seems to render the dimer problem on this non-Bravais lattice an unphysical model. But one ray of hope exists which becomes obvious by comparison of the profiles of both lattice forms, where the

TABLE I. Numbers of dimer coverings, $Z(k,n,6)$, of imperfect hexagonal lattices (cf. Fig. 4), the correspondence being indicated by a dashed line.

15106		658		28		1	
	3164		140		7		
29554		1288		55		2	
	6188		273		13		
42714		1863		80		3	
	8939		393		18		
54011		2358		102		4	
	11297		495		22		
		2751		120		5	
	13160		575		25		
		3024		133		6	
	14448		630		27		
		3164		140		7	
	15106		658		28		

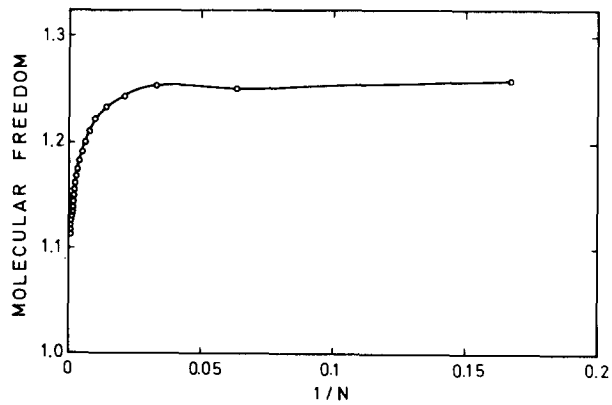


FIG. 5. Molecular freedom of the hexagonal lattice with rectangular shape in case of free boundary conditions as a function of lattice size (N = total number of lattice sites).

phenomenon of a molecular confinement has been observed [cf. Figs. 6(a) and 6(b)]. We found that it is (a) a rough surface of the hexagonal structure and (b) the parallel run of the horizontal boundary lines which restricts and rules the number of dimer coverings. The first point is a peculiarity inherent to the lattice type, whereas the second is a feature we can dispose of. This leads us to investigate in the next section a lattice with a trapezoidal profile [cf. Fig. 6(c)].

III. HEXAGONAL DOMAINS

The purpose of the present section is to verify that providing the hexagonal lattice with a trapezoidal profile prevents the molecular freedom from running into the trivial value $W = 1$. To be able to compare numbers of dimer coverings for both free and cyclic boundary conditions, we combine two equal-sized trapezoids to a hexagon, which still involves the possibility of a nontrivial behavior of the molecular freedom in the limit $N \to \infty$. As a first step we solve the dimer problem on a regular hexagon and, later on, discuss the modifications necessary for more general hexagonal domains.

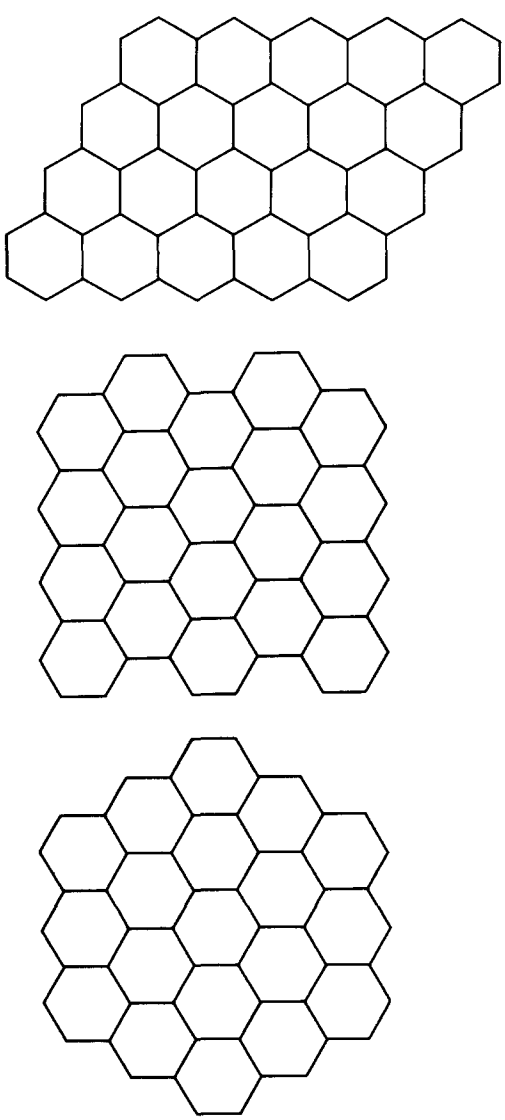


FIG. 6. Domains of the hexagonal lattice with the shape of a parallelogram (a), rectangle (b), and honeycomb (c). The last boundary form leads to a nonvanishing asymptotic molecular freedom ($W = 3\sqrt{3}/4$) whereas the lattices in Figs. 6(a) and 6(b) exhibit the phenomenon of molecular confinement ($W = 1$).

The proposed lattice represents a macroscopic reproduction of the honeycomb cell. It can hence be thought of as grown out of the central cell by attaching one-dimensional layers of honeycombs at the circumference. Let the macroscopic hexagon consist of L layers. Then the total number of lattice sites, N, is related to L by

$$N = 6(L + 1)^2. \tag{19}$$

To simplify the subsequent discussion, we recall that the hexagonal lattice represents an AB lattice, where all next neighbors of an A site belongs to the B sublattice and vice versa. This fact and the symmetry properties of the lattice obviously entail that one-third of the dimers of a complete covering will occupy one type of edge along one of the three principal axes. Hence a determination of the generating function advantageous in the case of the square lattice[1-4] proves to be superfluous for the lattice at hand, and it is advisable again to calculate the total number of coverings directly. This can be done by using the same technique developed in the last section.

We start by placing a first dimer in the vertex marked as a circle in Fig. 7. Occupying as a first possibility two horizontally adjacent vertices, we encounter a standard situation discussed above. That means, we release an avalanche of c-dimers as indicated in Fig. 7. The result is a lattice the left vertical boundary of which is completely occupied by dimers, or what is equivalent, a lattice reduced by one column of honeycombs and with a single monomer sitting on a boundary site of type A.

The other possibility of starting with a first dimer is represented by the second term in Fig. 7. Placing this dimer is equivalent to dismounting a complete honeycomb. This enables us to choose the lattice site in the adjacent honeycomb which is again marked by a circle as a new starting point. Proceeding in the same manner as described above, we produce again a lattice which is reduced by one column of cells and which accommodates a single monomer at a vertex neighboring, in the sublattice of A sites, on the first monomer position. Moreover, we obtain a lattice with two dismounted cells which requires a repeated application of the same pro-

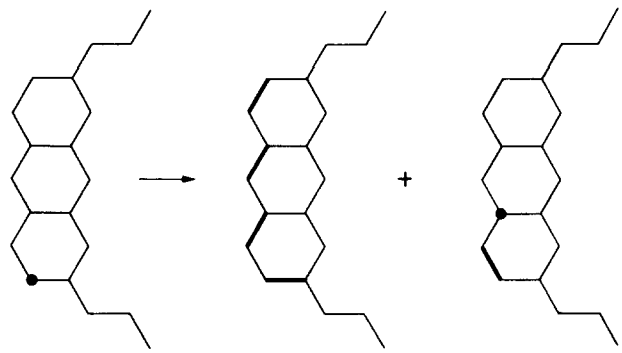
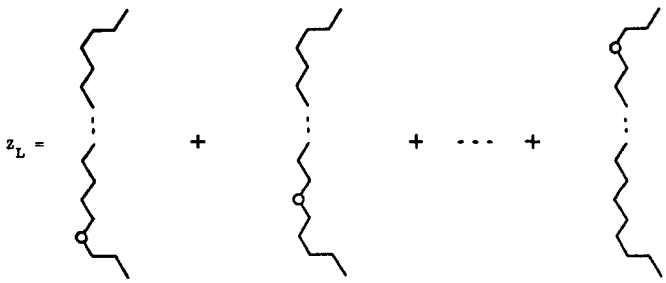


FIG. 7. The two possibilities of placing a first dimer on a hexagonal lattice with the shape of a macroscopic honeycomb ($L = 2$).

cedure. L steps of this kind finally yield the following sum of configurations:



(20)

where we have indicated each configuration by the left vertical boundary line and the position of the emerging monomer which occupies one of the $(L+1)$ possible A sites.

Thus we are faced with a monomer–dimer (MD) problem where the number of monomers after this first step is equal to 1. A next step produces lattices reduced by another column of honeycombs and decorated by two monomers. Generally, we find after n steps all possible configurations of n monomers distributed over $L+n$ boundary sites of type A.

For $n=1$ we found that each configuration appears with the weight 1. But in the case of $n=2$, an arbitrary two-monomer configuration results from various different one-monomer configurations, which will be called in the following ancestor configurations. Denoting the positions of the two monomers by x_i $(i=1,2)$, where $x_1<x_2$, it is easily found that a two-monomer configuration appears with the weight $W(x_1,x_2)=x_2-x_1$.

For general n, the weight of an n-monomer configuration proves to be a function of the relative distances x_i-x_k $(i,k=1,...,n)$, where x_k represents the position of the k th monomer and the monomers are numbered consecutively from 1 to n. The functional dependence of W on $x_1,...,x_n$ is given by

$$W(x_1,...,x_n)=\prod_{l>m}(x_l-x_m)\prod_{k=1}^{n-1}\frac{1}{k!}. \tag{21}$$

The proof of (21) is simply done by the use of mathematical induction and is given in Appendix A.

The procedure of reducing the honeycomb-shaped lattice of "size" L ceases after L steps and yields a set of L-monomer configurations, where the L monomers are distributed over $2L+1$ boundary sites in $\binom{2L+1}{L}$ different ways. A typical configuration with $L=2$ is shown in Fig. 8. This special configuration appears with the weight $W=1$. The rest lattice is of a taperwise shape and gives rise to c-dimers as indicated in Fig. 8. An actual placing of these dimers, however, proves to be disadvantageous and would complicate the state of affairs, as will become obvious in the following discussion.

Now we are left with an MD problem which can be solved on this special lattice owing to the restrictive shape of the boundary and the fact that the monomers are all placed on lattice sites of the same parity. To this end, we perform a last step of reduction in order to reduce the lattice by an

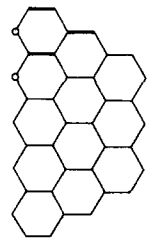


FIG. 8. Allowed two-monomer configurations as a result of partial covering of the hexagonal lattice shaped as a macroscopic honeycomb ($L=2$). The heavy lines indicate c-dimers.

additional column of cells which provides $(L+1)$-monomer configurations on an "imperfect" lattice. A representative lattice with this new feature is shown in Fig. 9, where for reasons of illustration we have again assumed $L=2$ and chosen one of four possible three-monomer configurations which are related to the ancestor configuration of Fig. 8 by the usual act of generation. The weight of this three-monomer configuration is $W(2,5,6)=6$.

The lattice we have now to deal with is characterized by two vertices of coordination number 1, which facilitates the solution of the MD problem on this lattice. Once this solution is found, we are able to calculate the combinatorial quantity Z_L by summing over all allowed $(L+1)$-monomer configurations.

The determination of possible dimer configurations on a lattice of the form shown in Fig. 9 is easily accomplished by employing again the procedure of generation of monomer configurations which will be started now at the right bordering column of L cells ($L=2$ in the special case of Fig. 9). The progenitive act is of the same mechanism as described above and finally yields a zigzag chain of $4L+3$ lattice sites with $L+1$ monomers occupying A sites and L monomers on B sites of this linear chain.

Clearly, the possible B-monomer configurations which permit a complete covering with $L+1$ dimers are exactly those which are able to generate the $(L+1)$-monomer configuration on the A sites, that is, a possible B-monomer configuration must represent an ancestor configuration of the A monomers we generated starting at the left-hand side of the lattice. To get the feel of things, let us work through all the allowed two-monomer configurations compatible with the positions of the three monomers in Fig. 9. Obviously one monomer is bound to occupy the isolated B site surrounded by the two neighboring A monomers. Positioning the second

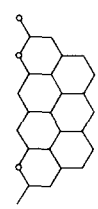


FIG. 9. Possible three-monomer configurations the ancestor configuration of which is shown in Fig. 8.

B monomer, one is faced with the same confinement. In order to ensure a complete covering of dimers, one is forced to reduce the uneven number of unoccupied sites in the interval defined by the second pair of A monomers. Thus we find three possible conjucated B-monomer configurations, as is shown in Fig. 10. The relative distance of these three pairs of B monomers (marked as full circles) being 1, 2, 3, respectively, the sum of weights of these configurations gives rise to a factor $W = 1 + 2 + 3 = 6$, which thus proves equal to the weight $W = 6$ of the A monomer configuration (indicated by open circles in Fig. 10).

Generally, we find that the number of dimer coverings of a honeycomb lattice of extent L is given by the following sum of squared weights:

$$Z_L = \sum_{\substack{\text{all configurations } X \\ \text{of } L+1 \text{ monomers}}} [W(x_1,x_2,\ldots,x_{L+1})]^2, \tag{22}$$

where x_i denotes the position of the ith monomer, it being assumed that the monomers are numbered consecutively from 1 to $(L + 1)$. The summation in (22) runs over all configurations of $(L + 1)$ monomers on A sites of a linear lattice, where the total number of A sites is

$$M = 2(L + 1). \tag{23}$$

Thus we succeeded in translating the dimer problem on a 2D lattice into a combinatorial problem which is now one-dimensional in character.

The closed expression in Eq. (22) provides an easy way to compute the combinatorial quantities Z_L with pen and paper. By way of example, the number of ways of covering a hexagon with $N = 150$ lattice sites ($L = 4$) has been found to be

$$Z_4 = 267\,227\,532.$$

We can obtain considerable insight into the form of the solution in the asymptotic limit of a large lattice by resorting to a related physical model of interacting electrons the partition function of which can be identified with (22).

Expressing the weight W in (22) in terms of Vandermonde's determinant [cf. Eq. (21)] gives the above representation of Z_L the form

$$Z_L = \prod_{k=1}^{L} (k!)^{-2} \sum_X \prod_{l>m} (x_l - x_m)^2, \tag{24}$$

where Σ_X means a sum over $\binom{M}{L+1}$ monomer configurations $X = (x_1,x_2,\ldots,x_{L+1})$. Now (24) can be written as

$$Z_L = \prod_{k=1}^{L} (k!)^{-2} \sum_X \exp\phi \tag{25}$$

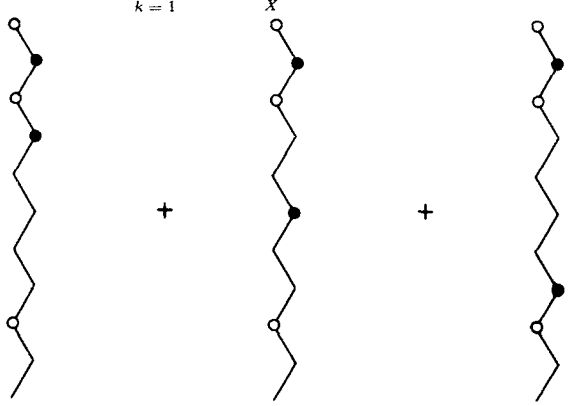


FIG. 10. Possible two-monomer configurations (full circles) compatible with the three-monomer configuration on the A sublattice shown in Fig. 9.

with

$$\phi = 2 \sum_{l>m} \ln(x_l - x_m) \tag{26}$$

and the combinatorial quantity Z_L is found to represent the partition function of a system of $(L + 1)$ particles freely movable on a linear lattice with M sites, if one ignores for the moment the unusual factor $\Pi_1^L (k!)^{-2}$. The interaction potential, which is a two-particle one, is given by $V(x_l - x_m) \propto \ln|x_l - x_m|$.

It is worth pointing out that this problem is related to various models of one-dimensional systems with Coulomb forces, e.g., Dawson's model where the potential energy is quadratic in displacements[16] and a model simultaneously solved by Prager[17] and Lenard[18] which represents a system of charged plane sheets with motion restricted to one dimension. The difference in the analytic form of the potential (being logarithmic in displacements in the model at hand) is not essential. What is important from the physical point of view is the fact that Dawson considers particles of one sign of charge embedded in an oppositely charged, uniform background, and Lenard and Prager deal with particles of both signs of charge, whereas we are faced with a system of equally charged particles without a neutralizing background. While for algebraic purposes this difference is relatively insignificant, it is quite crucial for the existence of a partition function in the thermodynamic limit.

As a matter of fact, the quantity Z_L in Eq. (25) associated to a system of $(L + 1)$ particles with Coulomb forces represents a partition function which is found to diverge in the bulk limit. Hence we conclude that there is, from the physical point of view, no virtue in dealing with this system. But, recalling that we started with a combinatorial problem on a lattice with $N \sim 6L^2$ lattice sites, a partition function of the order of $\exp(L^2)$ is just what is needed to assure a real molecular freedom.

In the following a rough sketch of the calculation of the partition function will be given and the existence of the thermodynamic limit $-L^{-2} \lim_{L\to\infty} \ln Z_L$ will only be indicated.

If we introduce an averaging notation

$$\langle\cdots\rangle = \prod_1^L \left(\frac{1}{k!}\right)^2 \sum_J, \tag{27}$$

the quantity Z_L may be expressed by $Z_L = \langle e^{\phi} \rangle$. The right-hand side can be computed by expanding $\exp\phi$ and operating with $\langle\cdots\rangle$ on each term. Then using the device of a cumulant expansion, taking the limit of large L, and omitting terms like $L \ln L$, etc., we arrive at

$$\log Z_L = \langle\phi\rangle + \tfrac{1}{2}(\langle\phi^2\rangle - \langle\phi\rangle^2) + \cdots$$
$$\sim 3L^2(\tfrac{3}{2}\log 3 - \log 4) + o(L^2), \quad L\to\infty. \tag{28}$$

To interpret the above thermodynamic limit in particle language, the associated resulting free energy per particle is proportional to L and the system of electrons with Coulomb forces is found to be unstable.

From the combinatorial point of view, this is a desired result, for it leads to a molecular freedom

$$W = \tfrac{3}{4}\sqrt{3} \tag{29}$$

on an infinite hexagonal domain.

IV. NONREGULAR HEXAGONAL DOMAINS

The last result in Eq. (29) indicates that, even on domains which deviate slightly from the regular hexagon discussed above, we will encounter a molecular freedom $W > 1$. In this section we wish to relax the restriction on regular hexagonal domains where the "lengths" L and M were connected by $M = 2(L+1)$. In the following, a hexagonal lattice of arbitrary shape is characterized by two independent lengths L and M. Then the total number of lattice sites is

$$N = 2(L+1)(2M - L - 1). \tag{30}$$

Although the choice of the "lengths" L and M is somewhat clumsy, its meaning is clear. We assume a nonregular hexagon to be composed of two trapezoids with $b = M - 1$ and $h = L$, where b and h denote the numbers of honeycombs constituting base and height, respectively.

The obvious advantage of the above definition of "lengths" L and M is that it leaves the representation of the combinatorial quantity Z [cf. Eq. (24)] unaltered, the only necessary modification being that the symbol $Z(L)$ has to be replaced by $Z(L,M)$ in order to specify the geometry of the underlying lattice. Thus, generally,

$$Z(L,M) = \prod_{k=1}^{L} (k!)^{-2} \sum_{X} \prod_{l>m} (i_l - i_m)^2. \tag{31}$$

The sum in (31) has to be taken over all $(L+1)$-monomer configurations on a linear lattice with M lattice sites, where M and L are now to be considered as independent lengths.

A study of more general domains is of no use, because attaching a column of honeycombs to an $L \times M$ lattice amounts to adding an uneven number of lattice points, which prohibits a complete covering of the resulting lattice. Moreover, a periodical repeating of those lattices and thus a comparison with solutions of toroidal lattices would be no longer possible.

Equation (31) represents a combinatorial problem which at first sight seems to be quite intricate. It turns out, however, that the quantity $Z(L,M)$ is simply expressible as a product of binomial coefficients. As an example, for $L = 1,2,3$, $M =$ fixed integer, one readily finds that

$$Z(1,M) = \frac{\binom{M}{1}\binom{M+1}{3}}{\binom{2}{1}\binom{3}{3}},$$

$$Z(2,M) = \frac{\binom{M}{1}\binom{M+1}{3}\binom{M+2}{5}}{\binom{3}{1}\binom{4}{3}\binom{5}{5}},$$

$$Z(3,M) = \frac{\binom{M}{1}\binom{M+1}{3}\binom{M+2}{5}\binom{M+3}{7}}{\binom{4}{1}\binom{5}{3}\binom{6}{5}\binom{7}{7}}, \tag{32}$$

and one is tempted to conclude that, generally,

$$Z(L,M) = \prod_{k=0}^{L} \frac{\binom{M+k}{2k+1}}{\binom{L+1+k}{2k+1}}. \tag{33}$$

The result in (33) is so simple as to invite proofs of equal simplicity. One way to verify (33) is to use mathematical induction.

We decompose the binomial coefficients in (33) into factorials and rearrange them in the following way:

$$Z(L,M) = \prod_{k=0}^{L} \frac{\binom{M+k}{M-L-1}}{\binom{M-1-k}{M-L-1}}. \tag{34}$$

Taking the logarithm on both sides, replacing summation by integration and using Stirling's formula, we get, after some tedious calculations, the result

$$\lim_{L\to\infty} \frac{1}{L^2} \ln[Z(L,M)]$$
$$= \tfrac{1}{2}(\xi+1)^2 \ln(\xi+1) + \tfrac{1}{2}(\xi-1)^2 \ln(\xi-1) - \xi^2 \ln\xi - \ln 4, \tag{35}$$

where we have introduced the parameter

$$\xi = M/L. \tag{36}$$

As a next step we express N, the number of lattice sites [cf. Eq. (30)] in terms of L and ξ and find that N behaves roughly as

$$N \sim 2L^2(2\xi - 1) \tag{37}$$

in the limit of an infinite lattice [$L\to\infty$, $M\to\infty$, $M/L = \xi$ (finite)]. Then the combinatorial quantity Z proves to be of the order of e^N assuring a nontrivial molecular freedom. Combination of (35) and (37) finally yields the molecular freedom as a function of ξ:

$$W(\xi) = [(\xi+1)^{(\xi+1)^2/2}(\xi-1)^{(\xi-1)^2/2}/4\xi^{\xi^2}]^{1/(2\xi-1)}. \tag{38}$$

The function $W(\xi)$ is represented in Fig. 11. As follows from (38), the molecular freedom has a maximum as a function of the ratio $\xi = M/L$ if $\xi = 2$. Inserting this value into Eq. (38), we finally obtain

$$W_{\max} = W(2) = \tfrac{3}{4}\sqrt{3}, \tag{39}$$

which reveals the regular hexagon ($M = 2L$) to be the domain with greatest possible molecular freedom.

Now let us turn to the two special cases $\xi\to\infty$ and $\xi\to 1$. The ratio $\xi = M/L = \infty$ is realized on a single column of honeycombs ($M\to\infty$, $L\to 0$). The number of dimer coverings of this strip is M and thus depends linearly on the total number of lattice sites ($N \sim 4M$), which entails $W = 1$ as $N\to\infty$.

The second limiting case $\xi\to 1$ corresponds to a hexagonal lattice with the shape of a parallelogram (lattice depicted in Fig. 1, where the two vertices with coordination number $q = 1$ must the two vertices with coordination number $q = 1$ must be omitted). Thus the rigorous results in Eqs. (34)–(38) established for hexagonal domains include the case of a lattice which has been mentioned in Sec. II as being representative for a lattice with the trivial molecular free-

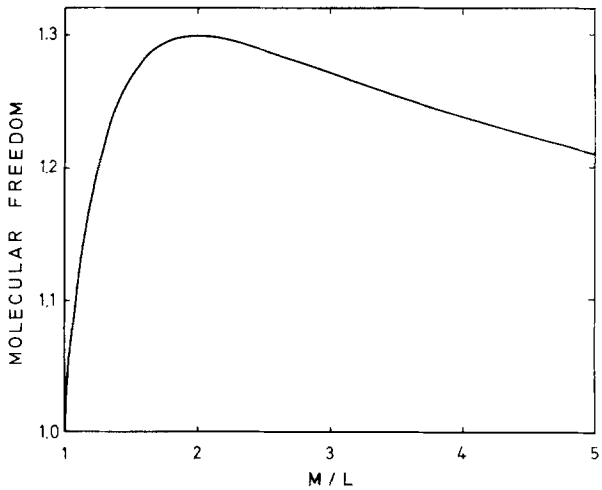


FIG. 11. Molecular freedom of a hexagonal lattice with honeycomb shape in case of free boundary conditions. M and L denote the characteristic lengths of the honeycomb so that $M = 2L$ corresponds to the regular hexagon.

dom $W = 1$ in the thermodynamic limit. Indeed, replacing M by $L + 2$ in Eq. (34) reproduces the result cited in Eq. (4). Furthermore, Eq. (38) yields the predicted molecular freedom $W(1) = 1$, as can be seen in Fig. 11.

V. CONCLUDING REMARKS

In this paper the dimer problem has been solved exactly for hexagonal lattices equipped with various boundary shapes. The important result is that the greatest molecular freedom attainable on a lattice with edges is $\frac{3}{4}\sqrt{3}$. This maximal value is found to be realized on an infinite domain shaped as a regular hexagon.

As soon as one employs periodic boundary conditions, the molecular freedom on a regular hexagon jumps from $W_{\max} = 1.299$ to the value $W_{\text{torus}} = 1.381$, which follows already from geometric considerations: In Fig. 12 we have indicated how to truncate the original domain and rearrange the constituent parts to a new domain which obviously exhibits the shape of a parallelogram. [This resulting domain consists of $L \times (3L + 2)$ honeycombs, where L denotes as usual the characteristic length of a regular hexagon with $N = 6(L + 1)^2$ lattice sites.] Thus, under the protection of periodic boundary conditions, these two lattice forms are shown to be indistinguishable and hence yield the same molecular freedom in the thermodynamic limit, namely that given by Eq. (2).

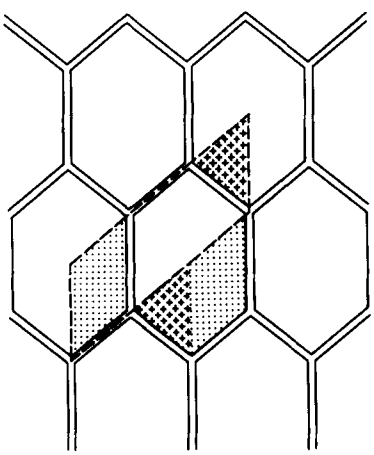


FIG. 12. Decomposition of a periodically repeated hexagonal domain and subsequent rearrangement. The resulting (repeated) lattice exhibits the shape of a parallelogram.

To sum up, we are faced with drastic boundary effects in several respects. The solution of the dimer problem on the hexagonal lattice depends decisively on the actual shape of the boundary. Moreover, we are faced with the phenomenon that, in contrast to the situation in Bravais lattices, free-end boundary conditions and periodic boundary conditions lead to different results in the bulk limit. The method of proof used for the free-end case indicates that this discrepancy arises from the existence of a rough surface appearing in a natural way on non-Bravais lattices. The striking result of this paper is that the restrictive influence of this rough surface survives even in the limit of a lattice with infinite extent and cannot be reduced completely by any contortion of the boundary shape.

Now the question arises of which boundary condition to use. Admittedly, as long as we do not intend to discuss boundary effects themselves, cyclic conditions are by far the most convenient. But on the lattice at hand we encounter the extreme situation that the application of periodic boundary conditions produces artificially a molecular freedom $W > 1$, where actually there is no freedom at all (e.g., rectangle and parallelogram). In view of this discrepancy, we recall that the natural boundary condition dictated by physical grounds is the free one, whereas the trick of periodic repetition of the lattice must be regarded as a pure mathematical device to get rid of complications brought about by a troublesome lattice surface. This leads us to conclude that at least on the hexagonal lattice the application of periodic boundary conditions is no longer justified.

Then we are still left with the unphysical sensitivity to the precise form of boundary. There are reasons to believe that this phenomenon results from the restriction to the only solvable case where the dimers completely fill the lattice and is thus a specific feature of the close-packed limit and might disappear as soon as monomers are present. Indeed, the results given in the last section indicate that already two monomers, placed on the two opposed sited corners of a finite parallelogram, augment the molecular freedom on this lattice, which predicts that taking account of all possible monomer positions and passing to the bulk limit, a finite ratio of monomers to lattice sites will lead away from the molecular freedom $W = 1$.

But this more general monomer–dimer problem is far from being solvable on two-dimensional lattices, and we must leave the question unanswered whether the presence of monomers will lead to a uniform molecular freedom or even permit the application of periodic boundary conditions.

APPENDIX: DEPENDENCE OF THE WEIGHT $W(x_1,\ldots,x_n)$ ON THE MONOMER POSITIONS x_i

The procedure which leads to an n-monomer configuration in the nth column of the hexagonal lattice can be ex-

pressed in mathematical terms by the following recursion relation:

$$W(x_1,\ldots,x_n) = \sum_{y_1=x_1}^{x_2-1}\sum_{y_2=x_2}^{x_3-1}\cdots\sum_{y_{n-1}=x_{n-1}}^{x_n-1} W(y_1,\ldots,y_{n-1}), \tag{A1}$$

where $W(y_1,\ldots,y_{n-1})$ represents the weight of a $(n-1)$-monomer configuration in the $(n-1)$th column of the lattice with positions $y_1<y_2<\cdots<y_{n-1}$. The restrictions when summing over y_i result from the condition that the configuration $X=(x_1,\ldots,x_n)$ can be generated by monomer configurations $Y=(y_1,\ldots,y_{n-1})$ only if

$$x_1\leqslant y_1<x_2,\quad x_2\leqslant y_2<x_3,\quad \ldots,\quad x_{n-1}\leqslant y_{n-1}<x_n. \tag{A2}$$

As a first step, we study the case $n=2$. The allowed positions y_1 of the one monomer generating the configuration $X=(x_1,x_2)$ in the next column of the lattice are $x_1\leqslant y_1<x_2$. Then, performing the sum in (A1), we find

$$W(x_1,x_2)=x_2-x_1, \tag{A3}$$

where $W(y_1)=1$ has been used.

The common feature of both results for $n=1,2$ is that W can be written as a determinant:

$$W(x_1,\ldots,x_n)=\begin{vmatrix} 1\binom{x_1}{1}\binom{x_1}{2}\cdots\binom{x_1}{n-1} \\ 1\binom{x_2}{1}\binom{x_2}{2}\cdots\binom{x_2}{n-1} \\ \cdots \\ 1\binom{x_n}{1}\binom{x_n}{2}\cdots\binom{x_n}{n-1} \end{vmatrix}. \tag{A4}$$

We now give an inductive proof that (A4) holds for general n. Our inductive premise for the weight $W(y_1,\ldots,y_{n-1})$ of an $(n-1)$-monomer configuration is

$$W(y_1,\ldots,y_{n-1}) = \det\binom{y_i}{k-1} = \sum_P \delta_P \binom{y_1}{p_1-1}\binom{y_2}{p_2-1}\cdots\binom{y_{n-1}}{p_{n-1}-1}, \tag{A5}$$

where the sum is on $(n-1)!$ permutations $P=(p_1,p_2,\ldots,p_{n-1})$. Now inserting (A5) into (A1) and performing the indicated sums on $y_1,\ldots,y_{n-1}$, one finds

$$W(x_1,\ldots,x_n)=\sum_P \delta_P \prod_{r=1}^{n-1}\left[\binom{x_{r+1}}{p_r}-\binom{x_r}{p_r}\right], \tag{A6}$$

where use has been made of the summation rule

$$\sum_{y=x_1}^{x_2-1}\binom{y}{p-1}=\binom{x_2}{p}-\binom{x_1}{p}. \tag{A7}$$

It is important to recall that the sum in (A6) is still on $(n-1)!$ permutations $P=(p_1,\ldots,p_{n-1})$, which at first sight makes it impossible to rewrite (A6) as an $n\times n$ determinant.

When we multiply out the factors in (A6) we observe that there are some terms in which, for example, x_2 occurs twice, namely in,

$$-\sum_P \sigma_P\binom{x_2}{p_1}\binom{x_2}{p_2}\prod_{r=3}^{n-1}\left[\binom{x_{r+1}}{p_r}-\binom{x_r}{p_r}\right], \tag{A8}$$

which will vanish when performing the sum over all permutations. The terms surviving the summation Σ_P in (A6) are those where x_i occurs only once $(i=1,\ldots,n)$. Thus (A6) can be reduced to

$$W(x_1,\ldots,x_n)=\sum_{k=1}^{n}(-1)^{k-1}\times\sum_P \delta_P\binom{x_1}{p_1}\cdots\binom{x_{k-1}}{p_{k-1}}\binom{x_{k+1}}{p_k}\cdots\binom{x_n}{p_{n-1}}. \tag{A9}$$

As a representative of all terms, we study

$$(-1)^{k-1}\sum_P \delta_P\binom{x_1}{p_1}\cdots\binom{x_{k-1}}{p_{k-1}}\binom{x_k}{0}\binom{x_{k+1}}{p_k}\cdots\binom{x_n}{p_{n-1}}, \tag{A10}$$

where we have inserted superfluously as k th factor $\binom{x_k}{0}$. As a next step we define a special element of the permutation group S_n, namely the permutation $P'=(p_1',\ldots,p_n')$, which is related to P by

$$p_r'=p_r+1,\quad r\neq k,$$
$$p_k'=1, \tag{A11}$$

and (A10) can be rewritten as

$$+\sum_{P'}\delta_{P'}\binom{x_1}{p_1'-1}\binom{x_2}{p_2'-1}\cdots\binom{x_n}{p_n'-1}, \tag{A12}$$

where $P'=(p_1',p_2',\ldots,p_n')$ are permutations of the numbers $(1,2,\ldots,n)$ with the restriction $p_k'=1$. Since the sum in (A9) is over all k's, $W(x_1,\ldots,x_n)$ can be expressed as a single sum over all permutations p' without any restriction so that

$$W(x_1,\ldots,x_n)=\sum_{P'}\delta_{P'}\prod_{r=1}^{n}\binom{x_n}{p_r'-1}, \tag{A13}$$

and we have found exactly what we want, namely the weight W of an n-monomer configuration in the determinantal form (A4) for general n.

As a last step towards a proof of Eq. (21) in Sec. III we recall the known identity[19]

$$\begin{vmatrix} 1\binom{x_1}{1}\binom{x_1}{2}\cdots\binom{x_1}{n-1} \\ 1\binom{x_2}{1}\binom{x_2}{2}\cdots\binom{x_2}{n-1} \\ \cdots \\ 1\binom{x_n}{1}\binom{x_n}{2}\cdots\binom{x_n}{n-1} \end{vmatrix} = \frac{\Pi_{i>k}(x_i-x_k)}{1!2!\cdots(n-1)!}. \tag{A14}$$

[1] P. W. Kasteleyn, Physica **27**, 1209 (1961).
[2] H. N. V. Temperley and M. E. Fisher, Phil. Mag. **6**, 1061 (1961).
[3] M. E. Fisher, Phys. Rev. **124**, 1664 (1961).
[4] E. W. Montroll, "Lattice Statistics," in *Applied Combinatorial Mathematics*, edited by E. F. Beckenbach (Wiley, New York, 1964), p. 105.
[5] M. E. Fisher, J. Math. Phys. **7**, 1776 (1966).
[6] R. H. Fowler and G. S. Rushbrooke, Trans. Faraday Soc. **33**, 1272 (1937).

[7]P. W. Kasteleyn, J. Math. Phys. **4**, 287 (1963).
[8]E. H. Lieb and F. Y. Wu, "Two-Dimensional Ferroelectric Models," in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, London, 1971), Vol. I, p. 440.
[9]J. F. Nagle, Phys. Rev. **152**, 190 (1966).
[10]C. Domb, Adv. Phys. **9**, 245 (1960).
[11]E. H. Lieb, J. Math. Phys. **8**, 2339 (1967).
[12]E. H. Lieb, "Model in Statistical Mechanics," in *Statistical Mechanics and Quantum Field Theory, Summer School of Theoretical Physics, Les Houches 1970*, edited by C. DeWitt and R. Stora (Gordon and Breach, New York, 1971), p. 283.
[13]L. van Hove, Physica **15**, 951 (1949).
[14]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969), p. 13.
[15]D. Grensing, J. Carlsen, and H. C. Zapp, Phil. Mag. **41**, 777 (1980).
[16]J. M. Dawson, Phys. Rev. **113**, 383 (1959).
[17]S. Prager, *Advances in Chemical Physics* (Wiley, New York, 1962), Vol. IV, p. 201.
[18]A. Lenard, J. Math. Phys. **2**, 682 (1961).
[19]W. Gröbner, *Matrizenrechnung* (Bibliographisches Institut, Mannheim, 1966), p. 66.

Factorized S matrices and generalized Baxter models

Anton Bovier[a)]
Physikalisches Institut, Universität Bonn, D-5300 Bonn 1, West Germany

(Received 29 December 1981; accepted for publication 2 September 1982)

We prove that solutions of the factorization equations for S matrices with Z_n symmetry can be given in terms of theta functions. We present a class of vertex models based on A-symmetric S matrices (A being an abelian group) as natural generalizations of the Baxter model. The mapping of these models into spin systems is described.

PACS numbers: 05.50. + q, 11.20. − e, 02.30. + g

I. INTRODUCTION

The existence of nontrivial interacting systems which are completely integrable, i.e., exactly solvable, is one of the most interesting problems in theoretical physics. For one thing they present excellent tools for studying various physical phenomena, as, for instance, phase transitions, or for testing approximate techniques; for the other, it is interesting in itself to understand the reasons for complete integrability.

Integrable systems are found in various parts of physics:

(1) A number of two-dimensional lattice models in statistical mechanics has been solved exactly. The most prominent ones are the Ising model[1] and the Baxter model.[2] In a sufficiently general formulation the latter contains almost all known solvable statistical models as special cases.[3]

(2) Some types of nonlinear partial differential equations are solvable by means of the inverse scattering transform method.[4,5] They appear in physics, for instance, in hydrodynamics (Korteweg–de Vries equation[6]) and as equations of motion of certain 1 + 1-dimensional field theories.[7]

In most cases, the quantitized versions of these field theories have also been shown to be completely integrable using the quantum method of the inverse scattering transform.[7–9] Among the best known of such models are the nonlinear Schrödinger model, the sine–Gordon model, the massive Thirring model, the Gross–Neveu model, and the Heisenberg spin chain.[10]

The understanding of these integrable systems has been largely improved during the past years by the realization that their existence is deeply related to the existence of factorized scattering theories in 1 + 1 dimensions.[11–14] A scattering theory is called factorized if its n-particle S matrices factorize into products of two-particle ones. The existence of such theories arises from the existence of solutions of certain *cubic functional equations*, the so-called factorization equations.

Factorized S matrices reappear as matrices of vertex weights in integrable statistical models and as the S matrices in the direct problem of the inverse scattering transform. In both cases, the factorization equations are crucial for the complete integrability.

In particular, from solutions of the factorization equations it is possible to *construct* integrable systems. Thus finding new solutions becomes an extremely interesting problem.

Due to the complexity of the factorization equations only relatively few solutions are rigorously known until now. The most important ones are the S matrix found by Baxter[2] in the course of the solution of the famous eight-vertex model and the O(n) symmetric factorized S matrices found by Zamolodchikov.[12] In addition, some rather special solutions are known.[15,16] In a recent paper, however, Beljavin[17] proposed a way which might allow for finding in a systematic way large classes of factorized S matrices. His principal idea was the conjecture that the factorization equation represents nothing but an addition formula for certain quasiperiodic (matrix) functions. It would then suffice to solve the quasiperiodicity equations, which is a far simpler problem. Beljavin was thus able to conjecture a class of new solutions, including the Baxter solution as the simplest special case. However, Beljavin could give no rigorous justification for his ansatz, and could not verify that the S matrices he found were indeed factorized ones.

In this paper we take up Beljavin's idea and give *rigorous* arguments for its validity. We give the corresponding S matrices explicitly in terms of theta functions and *prove* that they are solutions of the factorization equations.

As a second point we show that these S matrices are related to certain statistical models which are natural generalizations of the Baxter model. As in that case, they possess a representation both as a vertex model and as a spin system. In that part of the paper we keep our analysis more general considering models based on an arbitrary abelian group A whereas we give solutions of the factorization equations only in the case where A is a cyclic group, $A = Z_n$.

Solutions for the general case, which are just the naive generalizations of our solutions in the Z_n case, have been conjectured recently by Chudnowsky and Chudnowsky.[18] However, our proof cannot be generalized to that situation, and numerical tests, which have been done by Öttinger and Honerkamp[19] indicate that the conjecture is indeed false.

We have endeavored to keep this paper as readable as possible. To this end, we have included some introductory chapters which should provide the necessary background.

The paper is organized as follows: In Sec. II we derive, following Zamolodchikov,[12] the factorization equations. In Sec. III we explain the connection between factorized scattering theories and statistical vertex models. We give a sim-

a) Supported by Studienstiftung des deutschen Volkes. Present address: Theoretische Physik, ETH-Hönggerberg, CH-8093 Zürich, Switzerland.

 0022-2488/83/030631-11$02.50

ple method for the calculation of the free energy of such vertex models, which has been derived by Stroganov.[20] In Sec. IV we review some aspects of finite abelian groups and introduce the groups $\mathscr{D}(A)$, which we will need in what follows. In Sec. V we define a large class of vertex models as generalizations of the Baxter model. We also give their spin-system representation. In Sec. VI we find solutions of the factorization equations and present our proof.

II. FACTORIZED S MATRICES

In this chapter we briefly review the basic features of factorized scattering in $1+1$ dimensions and derive the factorization equations. For a more detailed review we refer the reader to the paper[12] by Zamolodchikov and Zamolodchikov.

The basic properties to be satisfied by any factorized scattering theory are[12,21,23]

(1) conservation of particle number,
(2) conservation of the total set of rapidities present in the initial state.

These properties reflect the existence of an infinite number of conservation laws in field theories possessing a factorized S matrix.

As has been shown by Zamolodchikov such scattering theories are most conveniently described in the following way. We introduce an algebra of states generated by one-particle states $A_k(u)$. Here k is an additive quantum number, distinguishing between the various types of particles in the theory, and u is the rapidity of the particle.

An n-particle state is then denoted by

$$A_{k_1}(u_1)A_{k_1}(u_2)\cdots A_{k_n}(u_n). \tag{II.1}$$

The order of the $A_k(u)$ in this expression reflects the spatial ordering of the particles. Of course, (II.1) is uniquely defined only if the algebra of $A_k(u)$'s is associative. We will now show that from the associativity of this algebra it follows that the S matrix is a factorized one and that the requirement of associativity leads to the factorization equations for the two-particle S matrix.

In terms of the $A_k(u)$'s, the two-particle S matrix is defined through

$$A_k(u_1)A_l(u_2) = S^{ij}_{kl}(u_1-u_2)A_j(u_2)A_i(u_1) \tag{II.2}$$

and the n-particle S matrix through

$$A_{j_1}(u_1)\cdots A_{j_n}(u_n) = S^{i_n\cdots i_1}_{j_1\cdots j_n}(u_1,\ldots,u_n)A_{i_n}(u_n)\cdots A_{i_1}(u_n), \tag{II.3}$$

where $u_1 \geqslant u_2 \geqslant \cdots \geqslant u_n$, that is, the left-hand side represents an in state and the right-hand one an out state.

If the algebra of A's is associative, in and out states are uniquely defined by the above expressions and the out state can be obtained from the in state by subsequent transmutations of neighboring rapidities, i.e., two-particle scatterings. The n-particle S matrix is thus a product of two-particle ones:

$$S^{i_1\cdots i_n}_{j_1\cdots j_n}(u_1,\ldots,u_n) = S^{k_1k_2}_{j_1j_2}(u_1-u_2)S^{k_3k_4}_{k_1j_3}(u_1-u_3)\cdots S^{i_{n-1}i_n}_{k_{n-1}k_n}(u_{n-1}-u_n). \tag{II.4}$$

Now, the two-particle S matrix determines the "commutator" (the only part of the product rule relevant for our purposes) of our algebra. Associativity of the algebra thus imposes certain conditions on the two-particle S matrix. To obtain these, it suffices to consider a three-particle in state and to calculate the corresponding out state in the two different possible ways. Equating the results yields the following cubic functional equations which are known as the Baxter–Yang–Zamolodchikov factorization equations:

$$S^{k_1k_2}_{i_1i_2}(u_1-u_2)S^{j_1k_3}_{k_1i_3}(u_1-u_3)S^{j_2j_3}_{k_2k_3}(u_2-u_3) = S^{k_2k_3}_{i_2i_3}(u_2-u_3)S^{k_1j_3}_{i_1k_3}(u_1-u_3)S^{j_1j_2}_{k_1k_2}(u_1-u_2). \tag{II.5}$$

It is useful to visualize the two processes considered in obtaining (II.5) by a space–time diagram (see Fig. 1). One sees from this picture that the factorization equations express the independence of the scattering process of the absolute distances of the particles involved. Only the relative order of the particles is relevant. From the definition (II.2) of the two-particle S matrix, it is clear that the natural "initial conditions" are

$$S^{ij}_{kl}(0) = \delta_{i,l}\delta_{j,k}. \tag{II.6}$$

It is important to notice that any solution to (II.5) satisfying (II.6) is "almost" unitary, i.e., it holds

$$S^{k_1k_2}_{ij}(u)S^{lm}_{k_2k_1}(-u) = \delta_{i,m}\delta_{j,l}f(u) \tag{II.7}$$

with some arbitrary function $f(u)$.

Thus any solution of (II.5) can be normalized to obtain a unitary S matrix. The necessary normalizing factor is determined by solving the functional equation:

$$n(u)n(-u) = f(u). \tag{II.8}$$

Apparently, then, $\tilde{S}(u) \equiv S(u)/n(u)$ is unitary.

If an additional symmetry, usually crossing symmetry, is imposed on the S matrix, the solution to (II.8) satisfying this additional requirement is unique (up to CDD ambiguities[23]).

In the next section we will show an interesting connection between this normalizing factor and the partition function of statistical models constructed using factorizing S matrices.[20]

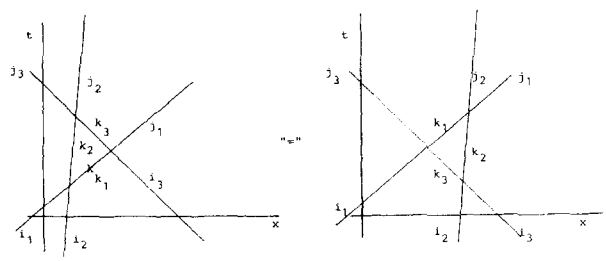


FIG. 1.

III. FACTORIZED S MATRICES AND STATISTICAL MODELS

In this section we will explain how the S matrices from Sec. II can be used to construct a statistical vertex model. In particular, we indicate why the models obtained are solvable if the original S matrix satisfies the factorization equations.

Consider a space–time diagram of the world lines of an n-particle scattering process from the preceding section[3,13] (see Fig. 2). Performing a Weyl rotation into Euclidean two-dimensional space, the world lines constitute an irregular two-dimensional lattice. The rapidities of the particles are now the angles of the lines forming the lattice. On this lattice we now introduce link variables which take the same values as the quantum numbers of the particles in the former scattering theory. To each configuration of link variables around one vertex, we now assign a distinct statistical weight (vertex weight) $w({}^{j}_{k}\times^{i}_{l})$, which may also depend on the angle u between the two crossing lines at this vertex. We take this vertex weight equal to the corresponding S matrix element:

$$w\left({}^{j}_{k}\times^{i}_{l}\right) = S^{ij}_{kl}(u). \tag{III.1}$$

We will use the notation $S^{ij}_{kl}(u)$ for vertex weights in the future rather than $w({}^{j}_{k}\times^{i}_{l})$ and call $S(u)$ the matrix of vertex weights or simply the weight matrix. The partition function of the vertex model defined above is now given by

$$Z = \sum_{\{i,j,k,l\}} \prod_{\text{vertices}} (S^{ij}_{kl}(u))_{\text{vertex}}. \tag{III.2}$$

Here the sum is to be taken over all the configurations of the variables on the links. We will now show why this vertex model is solvable, if the weight matrix $S(u)$ satisfies the factorization equations.

It is irrelevant for what follows that u represents the angle between lines of the lattice. We will thus consider a regular square $M\times N$ lattice in what follows, and let u be an arbitrary parameter. For convenience, we take the same u at any vertex. The usual way to solve such a model is to introduce transfer operators between rows of vertical links by[24]

$$\hat{T}^{k}_{l}{}^{\{i\}}_{\{j\}}(u) = S^{ki_1}_{li_1}(u)S^{ki_2}_{li_2}(u)\cdots S^{ki_N}_{li_N}(u). \tag{III.3}$$

Here $\{i\}$ represents the N-tuple $(i_1,i_2,\ldots,i_N)$. The trace of this operator with respect to the indices k and l is the usual transfer matrix:

$$T^{\{i\}}_{\{j\}}(u) = \sum_{k}\hat{T}^{k}_{k}{}^{\{i\}}_{\{j\}}(u), \tag{III.4}$$

which is connected to the partition function through the well-known relation

$$Z(u) = \operatorname{tr}(T^{M}(u)). \tag{III.5}$$

In the thermodynamic limit $N\to\infty$, $M\to\infty$, this simplifies to

$$Z(u) = t^{M}_{\max}(u), \tag{III.6}$$

where $t_{\max}(u)$ is the maximum eigenvalue of $T(u)$. Our problem is thus the determination of the eigenvalues of the transfer matrix (III.4). The most powerful technique known to do this is the Bethe ansatz[25] and its generalizations.[2,7,8,24,26] In its most algebraic version, discovered in the development of the quantum inverse scattering transform method, one makes the following ansatz for the eigenvectors of the transfer matrix:

$$\hat{T}^{k_1}_{l_1}(u_1)\hat{T}^{k_2}_{l_2}(u_2)\cdots\hat{T}^{k_M}_{l_M}(u_M)|0\rangle \tag{III.7}$$

with some known, simple eigenvector $|0\rangle$ of $T(u)$, whose eigenvalue is known. (In general, it can be pretty difficult to find such a state; for an example, see the calculations necessary in the case of the Baxter model.[2,24])

If simple commutation relations between the elements $\hat{T}^{k}_{l}(u)$ of the transfer operator are known, the action of $T(u)$ on the ansatz vectors (III.1) can be calculated by commuting $T(u)$ through all the $\hat{T}^{k_i}_{l_i}(u_i)$. One obtains a system of equations which determines the u_i and the eigenvalues. Now, the necessary commutation relations in our models are a straightforward consequence of the factorization equations (II.5). Namely from them and the definition (III.3) follows:

$$S^{k_1k_2}_{n_1n_2}(u-u')\hat{T}^{m_1}_{k_1}{}^{\{i'\}}_{\{i\}}(u)\hat{T}^{m_2}_{k_2}{}^{\{j\}}_{\{i'\}}(u') = \hat{T}^{k_2}_{n_2}{}^{\{i'\}}_{\{i\}}(u')\hat{T}^{k_1}_{n_1}{}^{\{j\}}_{\{i'\}}(u)S^{m_1m_2}_{k_1k_2}(u-u'). \tag{III.8}$$

In all known integrable systems the transfer-operator elements satisfy commutation relations of this form. (In the classical inverse scattering transform method, these commutation relations are present as Poisson-bracket relations between the scattering data.) The investigation of the infinite-dimensional algebras defined through relations of the form (III.8) appears to be a very interesting problem in itself; it lies, however, beyond the scope of the present paper. From Eq. (III.8) it follows easily that the transfer matrices for different values of the parameter u commute, i.e.,

$$[T(u),T(u')] = 0. \tag{III.9}$$

This equation was the starting point in Baxter's original work. He had observed this property of the transfer-matrix in the previously solved six-vertex model[27] and conjectured that it was the clue for the working of the Bethe ansatz. From (III.9) he then derived (III.8) and finally (II.5) as sufficient conditions.

The commuting of the transfer matrices is, however, also a physically very suggestive feature. Namely, it reflects directly the existence of infinitely many conserved quantities in Hamiltonian systems associated with their transfer matrices via

$$H_n = \frac{d^n}{du^n}\ln(T(u))\Big|_{u=0}. \tag{III.10}$$

For example, from the transfer matrix of the Baxter model one obtains the Hamiltonian of the Heisenberg XYZ model

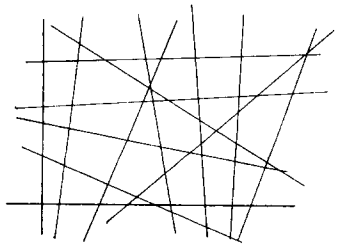


FIG. 2.

as H_1.[28]

The actual solution of such a model by Bethe ansatz is in general a very difficult problem, and we refer the reader to the excellent reviews by Takthajan and Faddeev[24] or Thacker[7] for details and examples. For the calculation of the free energy only one can, however, apply a far simpler method which has been proposed recently by Stroganov.[20] In this, one observes that from (II.7) it follows that

$$T(u)T(-u) = 1 f^N(u). \tag{III.11}$$

On the other hand, by virtue of (III.9), $T(u)$ and $T(-u)$ commute, and thus both matrices can be diagonalized simultaneously. Hence the eigenvalues $t(u)$ of $T(u)$ satisfy

$$t(u)t(-u) = f^N(u). \tag{III.12}$$

In particular, this equation holds for the maximum eigenvalue of $T(u)$. Since that is the M th root of the partition function, we have for the latter

$$Z^{1/MN}(u)Z^{1/MN}(-u) = f(u). \tag{III.13}$$

With $F(u) = -(kT/MN)\ln Z(u)$ it follows for the free energy per vertex $F(u)$:

$$F(u) + F(-u) = -kT \ln f(u). \tag{III.14}$$

(III.14), of course, only determines the even part of $F(u)$, so we need an additional relation to determine $F(u)$ uniquely. Such a relation is obtained from the invariance of the partition function under rotations and reflections of the lattice.

IV. THE GROUPS $\mathscr{D}(A)$

This section is intended to provide some concepts of abelian groups and of the nonabelian groups $\mathscr{D}(A)$ which are needed in what follows. Readers familiar with the theory of finite groups may skip most of this section and content themselves with some glimpses on definitions (IV.3), (IV.5), and (IV.7).

The building blocks in the following discussions are finite abelian groups.[29] The most simple types of them are cyclic groups, that is, groups, in which any element can be obtained as a power of one certain group element, called the generator. Cyclic groups are fully characterized by their order, that is, the number of their elements. We will denote the cyclic group of order n by $\mathbf{Z}_n$. The nth power of each element is the unit element, and, on the other hand, there is at least one element that doesn't give the unit element if it is raised to any power smaller than n.

The group $\mathbf{Z}_n$ is realized through the residue classes of integers modulo n with addition mod n as a group operation. Observe that multiplication mod n defines a second operation on this set and turns $\mathbf{Z}_n$ into a ring. We will identify $\mathbf{Z}_n$ with this ring if suitable. An arbitrary abelian group A is always obtained as a direct product of cyclic groups, that is,

$$A = \mathbf{Z}_{n_1} \times \cdots \times \mathbf{Z}_{n_k}. \tag{IV.1}$$

This group is realized by k tuples $\alpha = (\alpha_1,\ldots,\alpha_k)$ with α_i an element of $\mathbf{Z}_{n_i}$. The group operation is then given by componentwise addition mod n_i. It is important to keep in mind that the identity $\mathbf{Z}_n \times \mathbf{Z}_l = \mathbf{Z}_{kl}$ holds if and only if k and l are relatively prime, i.e., their largest common divisor is 1. In particular, it always holds $\mathbf{Z}_n \times \mathbf{Z}_n \neq \mathbf{Z}_{n^2}$. On the other hand, we have

$$\mathbf{Z}_n = \mathbf{Z}_{p_1^{l_1}} \times \cdots \times \mathbf{Z}_{p_s^{l_s}} \tag{IV.2}$$

if $n = p_1^{l_1} \cdots p_s^{l_s}$, where the p_i are different prime numbers. The most general abelian group thus has the decomposition

$$A = \mathbf{Z}_{p_1^{l_1}} \times \cdots \times \mathbf{Z}_{p_r^{l_r}}$$

with arbitrary (not necessarily different) primes p_i. Based on an abelian group A we now define nonabelian groups $\mathscr{D}(A)$. These are central extensions of the group $A \times A$ with a center $\mathbf{Z}_m$, m being the least common multiple (lcm) of all the n_j if A is given by (IV.1). To do this, we define a scalar product from $A \times A$ into $\mathbf{Z}_m$ by

$$(\alpha,\beta) = \sum_{i=1}^{m} \frac{\alpha_i \beta_i}{n_i} m. \tag{IV.3}$$

An element of $\mathscr{D}(A)$ is now denoted by

$$g^z_{\alpha,\beta}, \quad z \in \mathbf{Z}_m, \quad \alpha,\beta \in A, \tag{IV.4}$$

and the multiplication rule is defined through

$$g^z_{\alpha,\beta}\, g^{z'}_{\alpha',\beta'} = g^{z+z'+(\alpha,\beta')}_{\alpha+\alpha',\beta+\beta'}. \tag{IV.5}$$

Apparently, $g^0_{0,0}$ is the unit element of $\mathscr{D}(A)$ and the m elements $g^z_{0,0}$ form the center of the group. (The center of a group is the subgroup all of whose elements commute with all the elements of the whole group.)

As an example for definitions (IV.4) and (IV.5), consider the case $A = \mathbf{Z}_2$. The group $\mathscr{D}(\mathbf{Z}_2)$ has eight elements, and is generated by the products of the two generators $g^0_{1,0}$ and $g^0_{0,1}$. These generators satisfy

$$\begin{aligned} &(g^0_{1,0})^2 = (g^0_{0,1})^2 = g^0_{0,0} = id, \\ &(g^0_{1,0}\, g^0_{0,1})^4 = (g^1_{1,1})^4 = id. \end{aligned} \tag{IV.6}$$

From these generator relations[29] we see that $\mathscr{D}(\mathbf{Z}_2)$ is isomorphic to the dihedral group D_4.[29,30] In the next sections we will need representations of $\mathscr{D}(A)$, to be precise, one faithful irreducible representation.

The matrix elements of one such representation can be given explicitly. To do this, we label rows and columns by elements a, b from A. The matrix elements of $d = n_1 n_2 \cdots n_k$-dimensional representations are then given by

$$[T(g^z_{\alpha,\beta})]_{a,b} = \epsilon^{z+(\alpha,b)} \delta^A_{(a,b+\beta)}, \quad \epsilon \equiv e^{\pi i/m}. \tag{IV.7}$$

Here $\delta^A_{(a,b)}$ is one if, for all i, $a_i = b_i$, mod n_i, and zero otherwise. It is easy to check that (IV.7) gives a unitary irreducible representation of $\mathscr{D}(A)$. Besides the representation given by (IV.7) the groups $\mathscr{D}(A)$ possess still d^2 one-dimensional representations, and, except in the case $A = \mathbf{Z}_2 \times \mathbf{Z}_2 \times \cdots \times \mathbf{Z}_2$, further higher-dimensional representations. For the case $A = \mathbf{Z}_n$ a detailed discussion can be found in Ref. 30, and the results obtained in this case can easily be generalized. For the rest of this paper, however, the results given here will suffice.

V. GENERALIZED BAXTER MODELS

In this section we look for generalizations of the Baxter model and thus for candidates for new integrable systems. This section will be concerned with the symmetry properties of the corresponding weight matrices only, while the investi-

gation of the factorization equations will be the subject of Sec. VI.

Our considerations start from the observation of Wu and of Kadanoff and Wegner[31] that the Baxter model can be represented in the form of a spin system as well as of a vertex model. We will briefly repeat their construction (see Fig. 3). Consider two interlacing square lattices. Let the lines forming the lattice be numbered such that lines with even numbers belong to one lattice, those with odd numbers to the other. On every lattice site of the two lattices we place $\mathbf{Z}_2$ spin variables σ and define an interaction by the following Lagrangian:

$$L = \sum_{\substack{x,y \\ x\neq y \bmod 2}} \mathscr{L}_{xy}(\sigma_{x-1,y},\sigma_{x+1,y},\sigma_{x,y-1},\sigma_{x,y+1}), \quad \text{(V.1)}$$

where

$$\begin{aligned}\mathscr{L}_{x,y}&(\sigma_{x-1,y},\sigma_{x+1,y},\sigma_{x,y-1},\sigma_{x,y+1}) \\ &= g_1\delta(\sigma_{x-1,y}+\sigma_{x+1,y}) + g_2\delta(\sigma_{x,y-1}+\sigma_{x,y+1}) \\ &\quad + g_3\delta(\sigma_{x-1,y}+\sigma_{x,y+1}+\sigma_{x+1,y}+\sigma_{x,y-1}). \quad \text{(V.2)}\end{aligned}$$

We will call $\mathscr{L}_{x,y}$ the local Lagrangian of the vertex xy (mind that $x\neq y$, mod 2).

This is the spin system representation of the Baxter model. We will now transform this into the vertex-model representations. To do this, we introduce an intermediate lattice (dotted lattice in Fig. 3) with vertices xy. We will consider one such vertex in what follows and denote the surrounding spins by σ^1, σ^2, σ^3, and σ^4 for convenience. We place $\mathbf{Z}_2$ variables on the *links* of the new lattice (denoted by i,j,k,l). Link and site variables are connected by

$$i=\sigma^3+\sigma^2, \quad j=\sigma^2+\sigma^1, \quad k=\sigma^1+\sigma^4, \quad l=\sigma^4+\sigma^3. \quad \text{(V.3)}$$

To each configuration of the link variables at a vertex, we now assign a vertex weight according to the prescription

$$S^{ij}_{kl} = \begin{cases} e^{-(1/kT)\mathscr{L}_{xy}(\sigma^1,\sigma^2,\sigma^3,\sigma^4)} & \text{if } i=\sigma^3+\sigma^2, \\ & j=\sigma^2+\sigma^1, \text{ etc.,} \\ 0 \quad \text{otherwise, i.e., if } i+k\neq j+l. \end{cases} \quad \text{(V.4)}$$

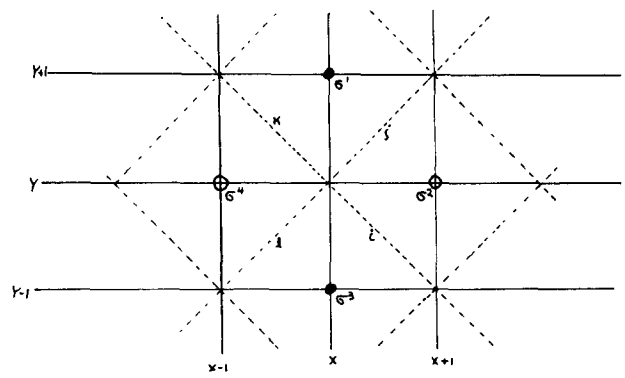


FIG. 3.

It is easy to see that the partition functions of the two models are equal:

$$\begin{aligned}Z_{\text{vertex}} &= \sum_{\{i,j,k,l\}} \prod_{\substack{x,y \\ x\neq y,\ \text{mod } 2}} (S^{ij}_{kl})_{xy} = \sum_{\{\sigma\}} \prod_{\substack{x,y \\ x\neq y,\ \text{mod } 2}} e^{-(1/kT)\mathscr{L}_{xy}(\sigma)} \\ &= \sum_{\{\sigma\}} e^{-(1/kT)L(\sigma)} = Z_{\text{spin}}. \end{aligned} \quad \text{(V.5)}$$

The vertex model we have obtained is just the Baxter model. The local Lagrangian (V.2) is not the most general function of the variables $\sigma^1,\sigma^2,\sigma^3,\sigma^4$, but is a function of the variables $\sigma^1+\sigma^3$ and $\sigma^2+\sigma^4$ only. This fact is reflected in the vertex representation by the fact that S^{ij}_{kl} depends on $i+k$ and $k+l$ only, if $i+k=j+l$ while otherwise $S^{ij}_{kl}=0$.

The spin-system version can be viewed as two identical anisotropic Ising models which are coupled by a four-spin interaction which involves the variables of the decoupled model only. Besides integrability, one of the most remarkable properties of the Baxter model is the fact that the critical indices vary continuously with the coupling constant of the four-spin interaction. As a first step in generalizing this model, let us replace all the $\mathbf{Z}_2$ variables by variables taking values in $\mathbf{Z}_2^k=\mathbf{Z}_2\times\cdots\times\mathbf{Z}_2$. Thus we denote by σ the k tupel $\sigma=(\sigma_1,\ldots,\sigma_k)$ with components σ_i in $\mathbf{Z}_2$, and correspondingly $i=(i_1,\ldots,i_k)$, etc.

We can keep up the relations (V.3) between link and site variables and take as the local Lagrangian again the most general function of $\sigma^1+\sigma^3$ and $\sigma^2+\sigma^4$, i.e.,

$$\begin{aligned}\mathscr{L}_{xy}(\sigma^1+\sigma^3,\sigma^2+\sigma^4) &= g_1\delta(\sigma_1^1+\sigma_1^3)+g_2\delta(\sigma_2^1+\sigma_2^3)+\cdots \\ &\quad + g_{2k}\delta(\sigma_k^2+\sigma_k^4)+g_{2k+1}\delta(\sigma_1^1+\sigma_1^3+\sigma_2^1+\sigma_2^3)+\cdots \\ &\quad + g\delta(\sigma_k^1+\sigma_k^3+\sigma_k^2+\sigma_k^4)+\cdots \\ &\quad + g_{2^{2k}-1}\delta(\sigma_1^1+\sigma_2^1+\cdots+\sigma_k^1+\sigma_1^2+\cdots+\sigma_k^2+\sigma_1^3+\cdots+\sigma_k^3+\sigma_1^4+\cdots+\sigma_k^4). \end{aligned} \quad \text{(V.6)}$$

For illustration we write down the complete Lagrangian in the case $k=2$. It describes two Ashkin–Teller models[32] which are coupled by four-, six-, and eight-spin couplings:

$$\begin{aligned}\mathscr{L}_{xy} &= g_1\delta(\sigma_1^1+\sigma_1^3)+g_2\delta(\sigma_2^1+\sigma_2^3)+g_3\delta(\sigma_1^1+\sigma_1^3+\sigma_2^1+\sigma_2^3) \\ &\quad + g_4\delta(\sigma_1^2+\sigma_1^4)+g_5\delta(\sigma_2^2+\sigma_2^4)+g_6\delta(\sigma_1^2+\sigma_1^4+\sigma_2^2+\sigma_2^4) \\ &\quad + g_7\delta(\sigma_1^1+\sigma_1^2+\sigma_1^3+\sigma_1^4)+g_8\delta(\sigma_2^1+\sigma_2^2+\sigma_2^3+\sigma_2^4) \\ &\quad + g_9\delta(\sigma_1^1+\sigma_2^2+\sigma_1^3+\sigma_2^4)+g_{10}\delta(\sigma_2^1+\sigma_1^2+\sigma_2^3+\sigma_1^4) \\ &\quad + g_{11}\delta(\sigma_1^1+\sigma_1^2+\sigma_1^3+\sigma_1^4+\sigma_2^1+\sigma_2^3)+g_{12}\delta(\sigma_2^1+\sigma_2^2+\sigma_2^3+\sigma_2^4+\sigma_1^1+\sigma_1^3) \\ &\quad + g_{13}\delta(\sigma_1^1+\sigma_1^2+\sigma_1^3+\sigma_1^4+\sigma_2^2+\sigma_2^4)+g_{14}\delta(\sigma_2^1+\sigma_2^2+\sigma_2^3+\sigma_2^4+\sigma_1^2+\sigma_1^4) \\ &\quad + g_{15}\delta(\sigma_1^1+\sigma_1^2+\sigma_1^3+\sigma_1^4+\sigma_2^1+\sigma_2^2+\sigma_2^3+\sigma_2^4). \end{aligned} \quad \text{(V.7)}$$

The weights of the corresponding vertex models are again introduced through (V.4) only with $\mathscr{L}_{xy}$ now given by (V.6). We obtain thus a 2^{3k}-vertex model with but 2^{2k} different vertex weights. Again the weight S^{ij}_{kl} is zero if $i+k \neq j+l$ and depends on $i+k$ and $k+l$ only, analogous to the situation of the Baxter model.

Before we proceed to introduce variables from a general abelian group A rather than $\mathbb{Z}_2^k$ only, we will find a more algebraic expression of the peculiar symmetries of the weight matrices resulting from the special choice of interaction Lagrangian.

Baxter in his original paper already mentioned that his weight matrix could be expanded in terms of tensor "squares" $\sigma_i \otimes \sigma_i$ of Pauli matrices. We will show that the weight matrices of our models can be expanded in a similar way using the representations of the groups $\mathscr{D}(\mathbb{Z}_2^k)$, introduced in the preceding section. Namely, we can write

$$S = \sum_{\gamma,\delta \in \mathbb{Z}_2^k} w_{\gamma,\delta} T(g^0_{\gamma,\delta}) \otimes T(g^0_{\gamma,\delta}). \tag{V.8}$$

Inserting the expression (IV.7) for the representation matrices, we arrive at

$$S^{ij}_{kl} = \sum_{\gamma,\delta \in \mathbb{Z}_2^k} w_{\gamma,\delta} \epsilon^{(\gamma,k+l)} \delta^{\mathbb{Z}_2^k}_{(i,k+\beta)} \delta^{\mathbb{Z}_2^k}_{(j,l+\beta)}. \tag{V.9}$$

We see that indeed $S^{ij}_{kl} = 0$ if $i+k \neq j+l$ is satisfied automatically by (V.8) and S^{ij}_{kl} depends only on $i+k$ and $k+l$. Thus the matrices (V.9) for arbitrary $w_{\gamma,\delta}$ represent just the possible weight matrices of our models. We will call matrices of the form (V.9) $\mathbb{Z}_2^k$-symmetric.

As a simple special case, using (V.8), we obtain for $k=1$ Baxter's result[2]:

$$S_{\text{Baxter}} = w_{00} \mathbb{1} \otimes \mathbb{1} - w_{01} \sigma_1 \otimes \sigma_1 + w_{10} \sigma_3 \otimes \sigma_3 - w_{11} \sigma_2 \otimes \sigma_2. \tag{V.10}$$

We will regard the symmetry expressed through (V.8) as the crucial property of our models. Hence it is clear how to further generalize them. Namely, taking variables now from an arbitrary finite abelian group A, we will require that the weight matrices of the resulting vertex models can be expanded using the groups $\mathscr{D}(A)$ in the form

$$S = \sum_{\gamma,\delta \in A} w_{\gamma,\delta} T(g^0_{\gamma,\delta}) \otimes T(g^0_{-\gamma,-\delta}) \tag{V.11}$$

with $g^0_{\gamma,\delta} \in \mathscr{D}(A)$. [The minus signs in the second term have been introduced to include the "initial conditions" (II.6) in our class of matrices. They result upon taking $w_{\gamma,\delta} = \epsilon^{(\gamma,\delta)}$.]

To these weight matrices again there corresponds a spin system. However, this time we have to be careful about signs, so that link and site variables are now related by

$$i = \sigma^3 - \sigma^2, \quad j = \sigma^2 + \sigma^1, \quad k = \sigma^1 - \sigma^4, \quad l = \sigma^4 + \sigma^3. \tag{V.12}$$

The possible form of the Lagrangian $\mathscr{L}_{xy}$ is now dictated by the requirement that S^{ij}_{kl} depends on $i-k$ and $k-l$ only. Consequently, we find that $\mathscr{L}_{xy}$ may depend on the variables $\sigma^2 + \sigma^4$ and $\sigma^4 + \sigma^3 - \sigma^2 - \sigma^1$ alone! Thus except in the case $A = \mathbb{Z}_2^k$, where $2\sigma^4 = 0$, no pure two-spin interactions in both directions on the lattice are allowed. In particular, ordinary Potts models will not be included as special cases of our models.

In the next section we will search for solutions of the factorization equations among A-symmetric S matrices. The A symmetry turns out to be very helpful in finding solutions.

VI. Z_n-SYMMETRIC SOLUTIONS OF THE FACTORIZATION EQUATIONS

This section is concerned with finding A-symmetric solutions of the factorization equations where $A = \mathbb{Z}_n$. In the first part we follow Beljavin[17] and derive an explicit expression of $\mathbb{Z}_n$-symmetric S matrices in terms of theta functions. In the second part we prove that these S matrices indeed satisfy the factorization equations. Beljavin's idea can easily be applied to a more general situation, where $\mathbb{Z}_n$ is replaced by an arbitrary abelian group $A = \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$ and the complex variable n by k complex variables $u_1,\ldots,u_k$. The naive generalization of the procedure leads to Chudnowsky and Chudnowsky's conjecture.[18]

However, a generalization of our proof does not seem possible. To show why this is the case, we give all the necessary auxiliary theorems on theta functions which we will use in their general form (i.e., as applying to theta functions of several variables) and indicate at what point of the argumentation the restriction to one complex variable is crucial.

It is remarkable that due to numerical calculations done by Öttinger and Honerkamp[19] not only does our proof not apply in this situation but also the conjecture seems to be false in the case of several variables.

To be able to write the factorization equations in a more transparent way, we introduce for any two-particle S matrix $S(u)$ the (formal) three-particle S matrix $S^{ij}(u)$, which operates nontrivial on the subspaces indicated by i and j only, for example,[17]

$$S^{12}(u) = \sum_{\gamma,\delta=0}^{n-1} w_{\gamma,\delta}(u) T(g^0_{\gamma,\delta}) \otimes T(g^0_{-\gamma,-\delta}) \otimes \mathbb{1}. \tag{VI.1}$$

Using this notation, the factorization equations take the form[17]

$$S^{12}(u) S^{13}(u+u') S^{23}(u') = S^{23}(u') S^{13}(u+u') S^{12}(u). \tag{VI.2}$$

In the following calculations we will need two properties of the groups $\mathscr{D}(A)$ and of A-symmetric S matrices. The first is the fact that the representations of $\mathscr{D}(A)$ are projective representations[29] of $A \times A$, i.e.,

$$T(g^0_{\alpha,\beta}) T(g^0_{\alpha',\beta'}) = \epsilon^{(\alpha,\beta')} T(g^0_{\alpha+\alpha',\beta+\beta'}). \tag{VI.3}$$

The second follows from this and from the A symmetry of $S(u)$:

$$T^1(g^0_{\alpha,\beta}) S^{12}(u) T^1(g^0_{\alpha,\beta})^{-1} = T^2(g^0_{\alpha,\beta})^{-1} S^{12}(u) T^2(g^0_{\alpha,\beta}). \tag{VI.4}$$

Here again the upper index to T indicates the subspace where T acts.

We will use these equations to construct first a solution on a lattice $\tilde{\Omega}$ in $\mathbb{C}$. The most general lattice has the form

$$\tilde{\Omega} = \{u \in \mathbb{C} : u = \omega\alpha + \omega'\beta; \alpha, \beta \in \mathbb{Z}\} \tag{VI.5}$$

with arbitrary complex numbers ω and ω'. On $\tilde{\Omega}$ we can define an S matrix by

$$S^{12}(\omega\alpha+\omega'\beta;\omega,\omega',f)=T^1(g^0_{\hat{\alpha},\hat{\beta}})S^{12}(0)T^1(g^0_{\hat{\alpha},\hat{\beta}})^{-1}\times f(\alpha,\beta;\omega,\omega') \tag{VI.6}$$

with

$$S^{12}(0)=\sum_{\gamma,\delta}^{n-1}\epsilon^{(\gamma,\delta)}T(g^0_{\gamma,\delta})\otimes T(g^0_{\gamma,\delta})\otimes \mathbf{1},$$

$$\text{i.e., } S^{ij}_{kl}=\delta_{i,l}\delta_{k,j}. \tag{VI.7}$$

Here $\hat{\alpha}=\alpha$, mod n, etc. In what follows we will drop the caret and write the same symbols for elements of $\mathbf{Z}$ and elements of $\mathbf{Z}_n$, understanding that they are related in the above manner.

$f(\alpha,\beta,\omega,\omega')$ is an arbitrary function. To avoid misinterpretations, we have written S explicitly as a function of ω,ω', and f besides u; that means we have defined a family of S matrices, distinguished by the choice of the lattice and the function f. In what follows we will drop these parameters, but it is important to keep in mind their implicit presence.

Inserting (VI.6) into (VI.2) and using Eqs. (VI.3) and (VI.4), it is easy to see that the S matrices defined by (VI.6) are indeed a solution on the corresponding lattice $\tilde{\Omega}$.

Now, the question is how we can use these solutions on a lattice to obtain solutions which are defined on all $\mathbb{C}$. As the most natural possibility one would rewrite Eq. (VI.6) as a quasiperiodicity condition for an S matrix defined on all $\mathbb{C}$:

$$S^{12}(u+\omega\alpha+\omega'\beta)=T^1(g^0_{\alpha,\beta})S^{12}(u)T^1(g^0_{\alpha,\beta})^{-1}\times f(\alpha,\beta;\omega,\omega';u). \tag{VI.8}$$

If again initial conditions (VI.7) are required, any solution to (VI.8) coincides on the lattice $\tilde{\Omega}$ with one of the S matrices from (VI.6), that means, satisfies (VI.2) on $\tilde{\Omega}$.

The next question to ask is when do solutions to (VI.8) exist and what choices of $\tilde{\Omega}$ and f yield independent solutions. These questions have been studied in the mathematical literature on quasiperiodic functions and allow for a definite answer.[33,34]

For the period lattice we can always take the choice

$$\Omega=\{u\in\mathbb{C}:u=\alpha+\tau\beta;\alpha,\beta\in\mathbf{Z}\}. \tag{VI.9}$$

Here τ is a complex number with positive imaginary part. The function f can be chosen in the form

$$f(\alpha,\beta,\tau,u)=e^{-i\pi\beta\tau\beta-2\pi i\beta(u+\eta)} \tag{VI.10}$$

with some arbitrary $\eta\in\mathbb{C}$.

Now (VI.8) takes the form

$$S^{12}(u+\alpha+\tau\beta)=T^1(g^0_{\alpha,\beta})S^{12}(u)T^1(g^0_{\alpha,\beta})^{-1}\times e^{-i\pi\beta\tau\beta-2\pi i\beta(u+\eta)}. \tag{VI.11}$$

This equation implies on the coefficient functions $w_{\gamma,\delta}(u)$:

$$w_{\gamma,\delta}(u+\alpha+\tau\beta)=\epsilon^{\alpha\delta-\beta\gamma}e^{-i\pi\beta\tau\beta-2\pi i\beta(u+\eta)}w_{\gamma,\delta}(u),\quad \epsilon=e^{2\pi i/n}. \tag{VI.12}$$

The solutions of (VI.13) can easily be found in terms of theta functions. Taking into account the initial conditions

$$w_{\gamma,\delta}(0)=\epsilon^{\gamma\cdot\delta}, \tag{VI.13}$$

one finds

$$w_{\gamma,\delta}(u)=\frac{\Theta_\tau[\bar{\delta}\,\bar{\gamma}](u+\eta)}{\Theta_\tau[\bar{\delta}\,\bar{\gamma}](\eta)}\epsilon^{\gamma\cdot\delta}. \tag{VI.14}$$

Here we have put

$$\bar{\delta}=\delta/n,\ \bar{\gamma}=\gamma/n. \tag{VI.15}$$

$\Theta_\tau[\bar{\delta}\,\bar{\gamma}](u)$ are the well-known theta functions with period matrix τ and characteristic $[\bar{\delta}\,\bar{\gamma}]$. They can be given as a theta series:

$$\Theta_\tau[\bar{\delta}\,\bar{\gamma}](u)=\sum_{k\in\mathbf{Z}}e^{i\pi(t+\bar{\delta})\tau(t+\bar{\delta})+2\pi i(t+\bar{\delta})(u+\bar{\gamma})}. \tag{VI.16}$$

Apparently, the series converges only if the imaginary part of τ is positive, hence the restriction on the period lattices. In this case, the theta function is an entire function in $\mathbb{C}$.

What has been achieved so far? We have obtained the explicit expression for an S matrix which is defined for all $u\in\mathbb{C}$ and which coincides with a solution of the factorization equations on the lattice Ω, that is, Eq. (VI.2) is satisfied, provided elements of Ω only are inserted for u and u'. However, at this stage it is absolutely not clear that these S matrices are solutions for all u and u'!

In the remainder of this section we will show that we have indeed found solutions of the factorization equations everywhere. To do this, we will consider the difference between the left- and right-hand side of (VI.2) as a function of u and u'.

We will use the $\mathbf{Z}_n$ symmetry of $S(u)$ to show that this function is zero for $u=0$ or $u'=0$. The quasiperiodicity of $S(u)$ implies certain quasiperiodicity relations for this difference function. This suffices to prove the identical vanishing of the difference. Let us now formulate our statement:

Theorem: A $\mathbf{Z}_n$-symmetric S matrix $S(u)$, which satisfies the initial conditions

$$S^{ij}_{kl}(0)=\delta(i,l)\delta(j,k)$$

and the quasiperiodicity relations

$$S^{12}(u+\alpha+\tau\beta)=T^1(g^0_{\alpha,\beta})S^{12}(u)T^1(g^0_{\alpha,\beta})^{-1}\times e^{-i\pi\beta\tau\beta-2\pi i\beta(u+\eta)}$$

is the solution of the factorization equations.

Proof: We first define a (matrix-valued) function $H(u,u')$ by

$$H(u,u')=S^{12}(u)S^{13}(u+u')S^{23}(u')-S^{23}(u')S^{13}(u+u')S^{12}(u). \tag{VI.17}$$

Some inspection shows that this function can be expanded in the form

$$H(u,u')=\sum_{\gamma\in\mathbf{Z}_n^4}h^\gamma(u,u')T(g^0_{\gamma,\delta})\otimes T(g^0_{-\gamma-\gamma',-\delta-\delta'})\otimes T(g^0_{\gamma,\delta'}). \tag{VI.18}$$

We have put $\gamma=(\gamma,\delta,\gamma',\delta')$ for convenience. To prove the theorem, we will have to show that all the $h^\gamma(u,u')$ vanish identically if $S(u)$ satisfies the given assumptions. We show now a first lemma, which uses the $\mathbf{Z}_n$ symmetry of the S matrix:

Lemma 1: For any A-symmetric S matrix we have, with definition (VI.17), $H(u,0) = H(0,u') = 0$.

Proof:

$$S^{12}(u)S^{13}(u)S^{23}(0)$$

$$= \sum_{\substack{\gamma_1,\gamma_2,\gamma_3\\ \delta_1,\delta_2,\delta_3}} w_{\gamma_1,\delta_1}(u)w_{\gamma_2,\delta_2}(u)\epsilon^{(\gamma_3,\delta_3)}T(g^{(\gamma_2,\delta_2)}_{\gamma_1+\gamma_2,\delta_1+\delta_2}) \otimes T(g^{-(\gamma_1,\delta_3)}_{\gamma_3-\gamma_1,\delta_3-\delta_1}) \otimes T(g^{(\gamma_2,\delta_3)}_{-\gamma_2-\gamma_3,\delta_2-\delta_3})$$

and

$$S^{23}(0)S^{13}(u)S^{12}(u)$$

$$= \sum_{\substack{\gamma_1,\gamma_2,\gamma_3\\ \delta_1,\delta_2,\delta_3}} w_{\gamma_1,\delta_1}(u)w_{\gamma_2,\delta_2}(u)\epsilon^{(\gamma_3,\delta_3)}T(g^{(\gamma_2,\delta_1)}_{\gamma_2+\gamma_1,\delta_2+\delta_1}) \otimes T(g^{-(\gamma_3,\delta_1)}_{\gamma_3-\gamma_1,\delta_3-\delta_1}) \otimes T(g^{(\gamma_3,\delta_2)}_{-\gamma_3-\gamma_2,-\delta_3-\delta_2}).$$

Performing the following transformation of summation indices,

$$\gamma_1\to\gamma_2,\quad \delta_3\to\delta_2,\quad \gamma\to\gamma_1,\quad \delta_2\to\delta_1,\quad \gamma_3\to\gamma_3+\gamma_2-\gamma_1,$$
$$\delta_3\to\delta_3+\delta_2-\delta_1,$$

and using that $T(g^z_{\gamma,\delta}) = \epsilon^z T(g^0_{\gamma,\delta})$, we see that both expressions coincide and thus $H(u,0) = 0$. $H(0,u') = 0$ follows correspondingly.

We will use the following definitions and properties of theta-functions in the next steps of the proof:

(1) *Definition*[33]: An analytic solution of the quasiperiodicity equation

$$j(u+\alpha+\tau\beta) = j(u) - \epsilon^{n(\alpha,\delta)-n(\beta,\gamma)}e^{-n(i\pi\beta\tau\beta+2\pi i\beta(u)}$$

is called a theta function of order n with period matrix τ and characteristic $\left[\begin{smallmatrix}\bar\delta\\ \bar\gamma\end{smallmatrix}\right]$.

(2) *Proposition*[33]: A theta function of order n with period matrix τ and characteristic $\left[\begin{smallmatrix}\bar\delta\\ \bar\gamma\end{smallmatrix}\right]$ can be expressed through theta functions of order 1 with period matrix $n\tau$:

$$\theta^n_\tau\left[\begin{smallmatrix}\bar\delta\\ \bar\gamma\end{smallmatrix}\right](u) = \sum_{\lambda\in\mathbb{Z}^k_n} c_\lambda \Theta_{n\tau}\begin{bmatrix}\bar\delta+\lambda\,/n\\ n\bar\gamma\end{bmatrix}(nu) \tag{VI.19}$$

with some set of constants C_λ. The n^k functions $\Theta_{n\tau}\left[\begin{smallmatrix}\bar\delta+\lambda\,/n\\ n\bar\gamma\end{smallmatrix}\right]$ are linear independent for different λ.

(3) *Transformation rule*[33,35]: Theta functions of order n with different characteristic are related by the equation

$$\theta^n_\tau\begin{bmatrix}\bar\delta+\bar l_1\\ \bar\gamma+\bar l_2\end{bmatrix}(u) = e^{i\pi n\bar l_1\tau\bar l_1+2\pi i n l_1(u+\bar\gamma+\bar l_2)}\theta^n_\tau\begin{bmatrix}\bar\delta\\ \bar\gamma\end{bmatrix}(u+\bar l_2+\tau\bar l_1). \tag{VI.20}$$

In particular,

$$\Theta_\tau\begin{bmatrix}\bar\delta+\bar l_1\\ \bar\gamma+\bar l_2\end{bmatrix}(u) = e^{i\pi\bar l_1\tau\bar l_1+2\pi i\bar l_1(u+\bar\gamma+\bar l_2)}\Theta_\tau\begin{bmatrix}\bar\delta\\ \bar\gamma\end{bmatrix}(u+\bar l_2+\tau\bar l_1). \tag{VI.21}$$

One easily verifies (VI.21) from the series expansion (VI.16) and (VI.20) follows from (VI.21) using (VI.19).

We now proceed to the proof of the theorem. In what follows it is always assumed that $S(u)$ is quasiperiodic, i.e., satisfies (VI.12).

Lemma 2: The functions $h^\gamma(u,u')$ are theta functions of order 2 with period τ in the variable u as well as in u'. As functions of u they have characteristic $\left[\begin{smallmatrix}\bar\delta/2\\ \bar\gamma/2\end{smallmatrix}\right]$ and as functions of u' characteristic $\left[\begin{smallmatrix}-\bar\delta'/2\\ -\bar\gamma'/2\end{smallmatrix}\right]$.

Proof of Lemma 2: The lemma results from the periodicity relations satisfied by $S(u)$. We have

$$H(u,u'+\alpha+\tau\beta) = T^3(g^0_{\alpha,\beta})^{-1}H(u,u')T^3(g^0_{\alpha,\beta}) \times e^{-2\pi i\beta\tau\beta-2\pi i\beta(u+2u'+2\eta)}$$

and

$$H(u+\alpha+\tau\beta,u') = T^1(g^0_{\alpha,\beta})H(u,u')T^1(g^0_{\alpha,\beta})^{-1} \times e^{-2\pi i\beta\tau\beta-2\pi i\beta(2u+u'+2\eta)}.$$

Hence it follows that

$$h^\gamma(u,u'+\alpha+\tau\beta) = \epsilon^{-\alpha\delta'+\beta\gamma}e^{-2\pi i\beta\tau\beta-2\pi i\beta(u+2u'+2\eta)} \times h^\gamma(u,u'), \tag{VI.22}$$

$$h^\gamma(u+\alpha+\tau\beta,u') = \epsilon^{\alpha\delta-\beta\gamma}e^{-2\pi i\beta\tau\beta-2\pi i\beta(2u+u'+2\eta)} \times h^\gamma(u,u'). \tag{VI.23}$$

According to the definition above, these equations define theta functions of order 2 as proposed.

A direct consequence of Lemmas 1 and 2 is that $h^\gamma(u,u')$ vanishes not only for $u=0$ or $u'=0$, but even for all $u\in\Omega$ or $u'\in\Omega$.

Together with the proposition quoted above, Lemma 2 allows us to expand $h^\gamma(u,u')$ in the two following forms:

$$h^\gamma(u,u') = \sum_{\lambda=0}^{1} C^\gamma_\lambda(u)\Theta_{2\tau}\begin{bmatrix}(-\bar\delta'+\lambda)/2\\ -\bar\gamma'\end{bmatrix}(2u'+u+2\eta), \tag{VI.24}$$

$$h^\gamma(u,u') = \sum_{\lambda=0}^{1} \tilde C^\gamma_\lambda(u')\Theta_{2\tau}\begin{bmatrix}(\bar\delta+\lambda)/2\\ \bar\gamma\end{bmatrix}(2u+u'+2\eta). \tag{VI.25}$$

Here $C^\gamma_\lambda(u)$ and $\tilde C^\gamma_\lambda(u')$ are analytic functions of u and u', respectively. (Of course, they depend on η and τ as well; however, we do not write this dependence explicitly as we are principally interested in the properties of the C^γ_λ as functions of u.)

We will use the two possible ways to represent $h^\gamma(u,u')$ crucially. First we derive from the knowledge of zeros of the functions $h^\gamma(u,u')$ a lemma about zeros of C^γ_λ and $\tilde C^\gamma_\lambda$.

Lemma 3: The functions C^γ_λ and $\tilde C^\gamma_\lambda$ have zeros for all $u\in\Omega$.

This lemma follows immediately from the vanishing of $h^\gamma(u,u')$ if one of the entries is a lattice point and from the linear independence of the theta functions $\Theta_{2\tau}\left[\begin{smallmatrix}(\bar\delta+\lambda)/2\\ \bar\gamma\end{smallmatrix}\right]$ for different λ.

A second important consequence of Lemma 1 is obtained upon inserting $u'=0$ in (VI.24) and $u=0$ in (VI.25). Namely, we get

$$0 = \sum_{\lambda=0}^{1} C^\gamma_\lambda(u)\Theta_{2\tau}\begin{bmatrix}-(\bar\delta'+\lambda)/2\\ \bar\gamma'\end{bmatrix}(u+2\eta), \tag{VI.26}$$

$$0 = \sum_{\lambda=0}^{1} \tilde C^\gamma_\lambda(u')\Theta_{2\tau}\begin{bmatrix}(\bar\delta+\lambda)/2\\ \bar\gamma\end{bmatrix}(u'+2\eta). \tag{VI.27}$$

$$C_0^\gamma(u) = -\left\{1 \quad \Theta_{2\tau}\begin{bmatrix} -\bar\delta'/2 \\ -\bar\gamma' \end{bmatrix}(u+2\eta)\right\} C_1^\gamma(u)$$
$$\times \Theta_{2\tau}\begin{bmatrix} (-\bar\delta'+1)/2 \\ -\bar\gamma' \end{bmatrix}(u+2\eta). \qquad \text{(VI.28)}$$

Introducing for $\lambda \neq 0$ the analytic functions[36]

$$g^\gamma(u) = C_1^\gamma(u) \quad \Theta_{2\tau}\begin{bmatrix} -\bar\delta'/2 \\ -\bar\gamma' \end{bmatrix}(u+2\eta), \qquad \text{(VI.29)}$$

we may expand $h^\gamma(u,u')$ in the form

$$h^\gamma(u,u') = g^\gamma(u)\left\{\Theta_{2\tau}\begin{bmatrix} -\bar\delta'/2 \\ -\bar\gamma' \end{bmatrix}(u+2\eta)\right.$$
$$\times \Theta_{2\tau}\begin{bmatrix} (-\bar\delta'+1)/2 \\ -\bar\gamma' \end{bmatrix}(2u'+u+2\eta)$$
$$- \Theta_{2\tau}\begin{bmatrix} (-\bar\delta'+1)/2 \\ -\bar\gamma' \end{bmatrix}(u+2\eta)$$
$$\left.\times \Theta_{2\tau}\begin{bmatrix} -\bar\delta'/2 \\ -\bar\gamma' \end{bmatrix}(2u'+u+2\eta)\right\}. \qquad \text{(VI.30)}$$

Using this expression together with (VI.25), we obtain periodicity relations for $g_\lambda^\gamma(u)$:

$$h^\gamma(u+\alpha+2\tau\beta,u') = g^\gamma(u+\alpha+2\tau\beta)\epsilon^{2\beta\gamma'-\alpha\delta'}(-1)^\alpha$$
$$\times e^{-2\pi i\beta 2\tau\beta - 2\pi i\beta(2u+2u'+4\eta)}$$
$$\times\left\{\Theta_{2\tau}\begin{bmatrix} -\bar\delta'/2 \\ -\bar\gamma' \end{bmatrix}(u+2\eta)\Theta_{2\tau}\begin{bmatrix} (-\bar\delta'+1)/2 \\ -\bar\gamma' \end{bmatrix}\right.$$
$$\times(2u'+u+2\eta) - \Theta_{2\tau}\begin{bmatrix} (-\bar\delta'+1)/2 \\ -\bar\gamma' \end{bmatrix}(u+2\eta)$$
$$\left.\times\Theta_{2\tau}\begin{bmatrix} -\bar\delta'/2 \\ -\bar\gamma' \end{bmatrix}(2u'+u+2\eta)\right\}$$
$$= \epsilon^{-2\beta\gamma+\alpha\delta} e^{-i\pi 4\beta 2\tau\beta - 2\pi i 2\beta(u'+2u+2\eta)} h^\gamma(u,u').$$

Hence

$$g^\gamma(u+\alpha+2\tau\beta) = \epsilon^{-2\beta(\gamma+\gamma')+\alpha(\delta+\delta')}(-1)^\alpha$$
$$\times e^{-i\pi 2\beta 2\tau\beta - 2\pi i 2\beta u} g^\gamma(u). \qquad \text{(VI.31)}$$

Thus we have shown:

Lemma 4: The functions $g^\gamma(u)$ are theta functions of order 2 with period matrix 2τ and characteristic $\begin{bmatrix} (\bar\delta+\bar\delta'+\frac12)/2 \\ \bar\gamma+\bar\gamma' \end{bmatrix}$. In addition, from Lemma 3 and definition (VI.29) it follows that $g^\gamma(u)$ has zeros for $u = \alpha + \tau\beta$. Otherwise, $\Theta_{2\tau}\begin{bmatrix} -\bar\delta'/2 \\ -\bar\gamma' \end{bmatrix}(2\eta)$ had to be zero, which is wrong for almost all η. But this means that we know two zeros of $g^\gamma(u)$ in each cell of the period lattice. According to Lemma 4, we can write $g^\gamma(u)$ in the form:

$$g_\lambda^\gamma(u) = \sum_{\mu=0}^{1} C_\mu^\gamma \Theta_{4\tau}\begin{bmatrix} (\bar\delta+\bar\delta'+1/2+\mu)/2 \\ 2(\bar\gamma+\bar\gamma') \end{bmatrix}(2u). \qquad \text{(VI.32)}$$

The two zeros of $g^\gamma(u)$ we know allow us to impose the following two equations on the two constants C_μ^γ:

$$\sum_{\mu=0}^{1} C_\mu^\gamma \Theta_{4\tau}\begin{bmatrix} (\bar\delta+\bar\delta'+1/2)/2+\mu/2 \\ 2(\bar\gamma+\bar\gamma') \end{bmatrix}(2\nu\tau) = 0, \quad \nu\in\mathbb{Z}_2. \qquad \text{(VI.33)}$$

This system of equations has nonzero solutions C_μ^γ only if the determinant of the matrix $(M_{\mu,\nu})$ with elements

$$M_{\mu,\nu} \equiv \Theta_{4\tau}\begin{bmatrix} (\bar\delta+\bar\delta'+1/2)/2+\mu/2 \\ 2(\bar\gamma+\bar\gamma') \end{bmatrix}(2\nu\tau) \qquad \text{(VI.34)}$$

is zero.

Thus we have to investigate that determinant. Using Eq. (VI.21), we get

$$M_{\mu,\nu} = \Theta_{4\tau}\begin{bmatrix} (\bar\delta+\bar\delta'+1/2)/2+\mu/2 \\ 2(\bar\gamma+\bar\gamma') \end{bmatrix}(2\nu\tau)$$
$$= \Theta_{4\tau}\begin{bmatrix} (\bar\delta+\bar\delta'+1/2)/2+\mu/2+\nu/2 \\ 2(\bar\gamma+\bar\gamma') \end{bmatrix}(0) e^{-i\pi\nu\tau\nu - 2\pi i\nu(\bar\gamma+\bar\gamma')}. \qquad \text{(VI.35)}$$

Hence the determinant of $M_{\mu,\nu}$ vanishes if and only if the determinant of

$$\left(\Theta_{4\tau}\begin{bmatrix} (\bar\delta+\bar\delta'+1/2)/2+\mu/2+\nu/2 \\ 2(\bar\gamma+\bar\gamma') \end{bmatrix}(0)\right)_{\mu,\nu}$$

is zero. The latter matrices have been investigated in mathematical literature. Therefore, we can refer to a paper by Koizumi,[35] in which the following equation is proven:

$$\det\left(\Theta_\tau\begin{bmatrix} k_1+p_1/n-p_2/n \\ k_2 \end{bmatrix}(u)\right)_{p_1,p_2\in\mathbb{Z}_n^k}$$
$$= \prod_{q\in\mathbb{Z}_n^k} e^{-2\pi i k_1 q}\Theta_{2/n^2}\begin{bmatrix} nk_1 \\ k_2/n+q/n \end{bmatrix}(u/n). \qquad \text{(VI.36)}$$

For our problem this gives

$$\det\left(\Theta_{4\tau}\begin{bmatrix} (\bar\delta+\bar\delta'+1/2)/2+\mu/2+\nu/2 \\ 2(\bar\gamma+\bar\gamma') \end{bmatrix}(0)\right)_{\mu,\nu\in\mathbb{Z}_2}$$
$$= \prod_{q=0}^{1} e^{-i\pi q(\bar\delta+\bar\delta'+1/2)}\Theta_\tau\begin{bmatrix} \bar\delta+\bar\delta'+1/2 \\ \bar\gamma+\bar\gamma'+q/2 \end{bmatrix}(0). \qquad \text{(VI.37)}$$

By this equation we have reduced our problem to the question for which $\bar\delta+\bar\delta'+1/2$ and $\bar\gamma+\bar\gamma'$ one of the factors on the right-hand side of (VI.37) vanishes. A well-known theorem, which may be found, for instance, in the textbook by Igusa,[34] contains the answer to this question: It says that $\Theta_\tau\begin{bmatrix} \bar\delta \\ \bar\gamma \end{bmatrix}(0) \equiv 0$ if and only if $2\bar\delta = 0 \bmod 1$ and $2\bar\gamma = 0 \bmod 1$ and $\frac12\bar\gamma\bar\delta = \frac12 \bmod 1$. This gives us the following lemma:

Lemma 5: If $2(\bar\gamma+\bar\gamma') \neq 0 \bmod 1$ or $\bar\delta+\bar\delta' = \neq 0 \bmod 1$, then $g^{\gamma,\delta,\gamma',\delta'}(u) \equiv 0$.

By now the proof is almost complete and we have only to consider the two remaining cases $\bar\delta+\bar\delta' = 0, \bar\gamma+\bar\gamma' = 0$, and $\bar\delta+\bar\delta' = 0, \bar\gamma+\bar\gamma' = \frac12$.

Before we do this, we give a well-known proposition on the zeros of nth order theta functions of *one* complex variable. We mention that this proposition can be used as an alternative to the investigation of the determinant $M_{\mu\nu}$ in order to derive Lemma 5.

Proposition[34]: A theta function of order n and characteristic $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ in one variable has n zeros in each period cell and

the sum of these zeros in the elementary cell (spanned by 1 and τ) is equal to $\frac{1}{2}n(1+\tau)$. We now come to the proof that even the "exceptional" $g^{\gamma}(u)$ must vanish. To do this, we introduce, in analogy to $g^{\gamma}(u)$, the functions $\tilde{g}^{\gamma}(u')$. One easily verifies that they are theta functions of the same type as $g^{\gamma}(u)$ but have characteristic $\begin{bmatrix} (-\bar{\delta}-\bar{\delta}'+\lambda\ /2)/2 \\ -\bar{\gamma}-\bar{\gamma}' \end{bmatrix}$.

Thus, in the exceptional cases (only) g^{γ} and $\tilde{g}^{\gamma}$ have *same* characteristic! We have the general expression for these functions:

$$g^{\gamma}(u) = C_0^{\gamma}\Theta_{4\tau}\begin{bmatrix}1/4\\0\end{bmatrix}(2u) + C_1^{\gamma}\Theta_{4\tau}\begin{bmatrix}3/4\\0\end{bmatrix}(2u),$$
$$\tilde{g}^{\gamma}(u') = C_0^{\gamma}\Theta_{4\tau}\begin{bmatrix}1/4\\0\end{bmatrix}(2u') + \tilde{C}_1^{\gamma}\Theta_{4\tau}\begin{bmatrix}3/4\\0\end{bmatrix}(2u'). \tag{VI.38}$$

Taking into account that both functions vanish at $u=0$, resp. $u'=0$, and observing that, e.g.,

$$\Theta_{4\tau}\begin{bmatrix}1/4\\0\end{bmatrix}(0) = \Theta_{4\tau}\begin{bmatrix}3/4\\0\end{bmatrix}(0)$$

we have even:

$$g^{\gamma}(u) = C^{\gamma}\left\{\Theta_{4\tau}\begin{bmatrix}1/4\\0\end{bmatrix}(2u) - \Theta_{4\tau}\begin{bmatrix}3/4\\0\end{bmatrix}(2u)\right\},$$
$$\tilde{g}^{\gamma}(u') = \tilde{C}^{\gamma}\left\{\Theta_{4\tau}\begin{bmatrix}1/4\\0\end{bmatrix}(2u') - \Theta_{4\tau}\begin{bmatrix}3/4\\0\end{bmatrix}(2u)\right\}. \tag{VI.39}$$

In particular from this it follows that $g^{\gamma}(u)$ and $\tilde{g}^{\gamma}(u)$ are proportional! One may now write $h^{\gamma}(u,u')$ once in the form (VI.22) and once in form (VI.23) and equate the results. This gives

$$\begin{aligned} h^{\gamma}(u,u') &= g^{\gamma}(u)\left\{\Theta_{2\tau}\begin{bmatrix}\bar{\delta}/2\\\bar{\gamma}\end{bmatrix}(u+2\eta+\rho/2)\right. \\ &\quad\times\Theta_{2\tau}\begin{bmatrix}(\bar{\delta}+1)/2\\\bar{\gamma}\end{bmatrix}(2u'+u+2\eta+\rho/2) \\ &\quad-\Theta_{2\tau}\begin{bmatrix}(\bar{\delta}+1)/2\\\bar{\gamma}\end{bmatrix}(u+2\eta+\rho/2) \\ &\quad\left.\times\Theta_{2\tau}\begin{bmatrix}\bar{\delta}/2\\\bar{\gamma}\end{bmatrix}(2u'+u+2\eta+\rho/2)\right\} \\ &= k^{\gamma}g^{\gamma}(u')\left\{\Theta_{2\tau}\begin{bmatrix}\bar{\delta}/2\\\bar{\gamma}\end{bmatrix}(u'+2\eta)\right. \\ &\quad\times\Theta_{2\tau}\begin{bmatrix}(\bar{\delta}+1)/2\\\bar{\gamma}\end{bmatrix}(2u+u'+2\eta) \\ &\quad-\Theta_{2\tau}\begin{bmatrix}(\bar{\delta}+1)/2\\\bar{\gamma}\end{bmatrix}(u'+2\eta) \\ &\quad\left.\times\Theta_{2\tau}\begin{bmatrix}\bar{\delta}/2\\\bar{\gamma}\end{bmatrix}(u'+2u+2\eta)\right\}. \end{aligned} \tag{VI.40}$$

Here we have put $\rho/2 = \bar{\gamma}+\bar{\gamma}'$, $\rho\in\mathbb{Z}_2$ and $\tilde{g}^{\gamma}(u) = k^{\gamma}g^{\gamma}(u)$. We have to distinguish two cases, namely $\rho=0$ and $\rho=1$.

Case 1: $\rho=1$: Put $u=\frac{1}{2}$ in (VI.40). We arrive at

$$\begin{aligned} &g^{\gamma}(1/2)\left\{\Theta_{2\tau}\begin{bmatrix}\bar{\delta}/2\\\bar{\gamma}\end{bmatrix}(2\eta)\Theta_{2\tau}\begin{bmatrix}(\bar{\delta}+1)/2\\\bar{\gamma}\end{bmatrix}(2u'+2\eta)\right. \\ &\left.-\Theta_{2\tau}\begin{bmatrix}(\bar{\delta}+1)/2\\\bar{\gamma}\end{bmatrix}(2\eta)\Theta_{2\tau}\begin{bmatrix}\bar{\delta}/2\\\bar{\gamma}\end{bmatrix}(2u'+2\eta)\right\}e^{2\pi i(\bar{\delta}+1/2)} \\ &= k^{\gamma}g^{\gamma}(u')\left\{-\Theta_{2\tau}\begin{bmatrix}\bar{\delta}/2\\\bar{\gamma}\end{bmatrix}(u'+2\eta)\Theta_{2\tau}\begin{bmatrix}(\bar{\delta}+1)/2\\\bar{\gamma}\end{bmatrix}(u'+2\eta)\right. \\ &\left.-\Theta_{2\tau}\begin{bmatrix}(\bar{\delta}+1)/2\\\bar{\gamma}\end{bmatrix}(u'+2\eta)\Theta_{2\tau}\begin{bmatrix}\bar{\delta}/2\\\bar{\gamma}\end{bmatrix}(u'+2\eta)\right\}e^{i\pi\bar{\delta}}. \end{aligned} \tag{VI.41}$$

Using the proposition above and the transformation rule for theta functions (VI.20), we find that the right-hand side of (VI.41) vanishes with $\Theta_{2\tau}\left[\begin{smallmatrix}\bar{\delta}/2\\\bar{\gamma}\end{smallmatrix}\right](u'+2\eta)$ at

$$u_0' = \tfrac{1}{2}+\tau-2\eta-\tau\bar{\delta}-\bar{\gamma}.$$

The left-hand side, however, is a theta function of order 2, one zero of which we see immediately, namely $u'=0$. The second zero can now be calculated using again our proposition yielding

$$u_2' = 1+\tau-2\eta-\tau\bar{\delta}-\bar{\gamma}.$$

Since $u_2' = u_0'$, we must have $g^{\gamma}(\frac{1}{2})=0$. But this again implies $g^{\gamma}(u)\equiv 0$.

Case 2: Here a procedure similar to that in Case 1 would not give any result. But, fortunately, $h^{\gamma,\delta,-\gamma,-\delta}(u,u')$, which corresponds to this case, can be shown to be zero as a direct consequence of the factorization equations, regardless even what S matrices are inserted. This holds not only for $\mathbb{Z}_n$ but for arbitrary A.

To show this, we write down the factorization equations (VI.1) with (VI.2) inserted:

$$\begin{aligned} 0 = \sum_{\substack{\gamma_1,\gamma_2,\gamma_3\\ \delta_1,\delta_2,\delta_3}} & w_{\gamma_1,\delta_1}(u)w_{\gamma_2,\delta_2}(u+u')w_{\gamma_3,\delta_3}(u')\{T(g^{(\gamma_1,\delta_2)}_{\gamma_1+\gamma_2,\delta_1+\delta_2}) \\ &\otimes T(g^{-(\gamma_1,\delta_3)}_{\gamma_3-\gamma_1,\delta_3-\delta_1}) \otimes T(g^{(\gamma_2,\delta_3)}_{-\gamma_2-\gamma_3,-\delta_2-\delta_3}) \\ &- T(g^{(\gamma_2,\delta_1)}_{\gamma_2+\gamma_1,\delta_3+\delta_1}) \\ &\otimes T(g^{-(\gamma_3,\delta_1)}_{\gamma_3-\gamma_1,\delta_3-\delta_1}) \otimes T(g^{(\gamma_3,\delta_2)}_{-\gamma_3-\gamma_2,-\delta_3-\delta_2})\}. \end{aligned} \tag{VI.42}$$

Introducing new summation indices

$$\gamma=\gamma_1+\gamma_2,\quad \gamma''=\gamma_1,\quad \gamma'=-\gamma_2-\gamma_3,$$
$$\delta=\delta_1+\delta_2,\quad \delta''=\delta_1,\quad \delta'=-\delta_2-\delta_3,$$

we obtain

$$\begin{aligned} 0 = \sum_{\gamma'',\delta''\in A} & w_{\gamma'',\delta''}(u)w_{\gamma-\gamma'',\delta-\delta''}(u+u') \\ &\times w_{\gamma''-\gamma-\gamma',\delta''-\delta-\delta'}(u') \\ &\times\epsilon^{-3(\gamma'',\delta'')-(\gamma,\delta)}\{\epsilon^{3(\gamma'',\delta)+2(\gamma'',\delta')+(\gamma,\delta'')-(\gamma,\delta')} \\ &-\epsilon^{3(\gamma,\delta'')+2(\gamma',\delta'')+(\gamma'',\delta)-(\gamma',\delta)}\}. \end{aligned} \tag{VI.43}$$

If $\gamma=-\gamma'$ and $\delta=-\delta'$, this equation is trivially satisfied and thus $h^{\gamma,\delta,-\gamma,-\delta}(u,u')\equiv 0$ automatically. This proves the theorem.

ACKNOWLEDGMENTS

I want to thank Professor V. Rittenberg for stimulating this work and for helpful discussions during its completion. I also thank U. Abresch for his assistance in some mathemat-

ical questions. I am indebted to Professor J. Honerkamp and Dr. H.-C. Öttinger for communicating the results of their calculations to me prior to publication.

[1]L. Onsager, Phys. Rev. **65**, 117 (1944); T. D. Schultz, D. C. Mattis, and E. H. Lieb, Rev. Mod. Phys. **36**, 856 (1964).
[2]R. J. Baxter, Ann. Phys. (N.Y.) **70**, 193 (1972); **76**, 1, 25, 48 (1973).
[3]R. J. Baxter, Phil. Trans. Roy. Soc. **289**, 315 (1978).
[4]V. E. Zakharov, *The Method of the Inverse Scattering Problem*, Lecture Notes in Mathematics (Springer-Verlag, New York, 1978).
[5]V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].
[6]L. S. Gardener, J. M. Greene, M. D. Kruskal, and R. M. Muira, Phys. Rev. Lett. **19**, 1095 (1967).
[7]H. B. Thacker, Rev. Mod. Phys. **53**, 253 (1981).
[8]L. D. Faddeev, Leningrad Preprint LOMI-P-2, 1979.
[9]F. K. Sklyanin, L. A. Takhtajan, and L. D. Faddeev, Theor. Math. Phys. **40**, 688 (1980).
[10]P. P. Kulish and E. K. Sklyanin, Phys. Lett. A **70**, 461 (1979).
[11]A. B. Zamolodchikov and Al. B. Zamolodchikov, Nucl. Phys. B **133**, 525 (1978).
[12]A. B. Zamolodchikov and Al. B. Zamolodchikov, Ann. Phys. (N.Y.) **120**, 253 (1979).
[13]A. B. Zamolodchikov, Comm. Math. Phys. **69**, 165 (1979).
[14]A. B. Zamolodchikov, Zh. Eksp. Teor. Fiz. **79**, 641 (1980) [Sov. Phys. JETP **52**, 325 (1980)].
[15]V. V. Bazhanov and Yu. G. Stroganov, Preprint N.O.B.80-14, 1980.
[16]V. A. Fateev, Preprint CERN TH-2963, 1980.
[17]A. A. Belavin, Pis'ma Zh. Eksp. Teor. Fiz. **32**, 182 (1980) [Sov. Phys. JETP Lett. **32**, 169 (1980)].
[18]D. V. Chudnowsky and G. V. Chudnowsky, Preprint Saclay DPh-T/80/131, 1980.
[19]H. C. Öttinger and J. Honerkamp, Phys. Lett. A **88**, 339 (1982). See also R. Shrock, Stony Brook Preprint ITP-SB-82-17, 1982.
[20]Yu. G. Stroganov, Phys. Lett. A **74**, 116 (1979).
[21]C. N. Yang, Phys. Rev. A **16**, 1920 (1968).
[22]D. Iagollnitzer, *Lecture Notes in Physics* (Springer-Verlag, New York, 1980), Vol. 126.
[23]L. Castillejo, R. H. Dalitz, and F. J. Dyson, Phys. Rev. **101**, 453 (1956).
[24]L. A. Takhtajan and L. D. Faddeev, Usp. Mat. Nauk **34**, 13 (1979).
[25]H. A. Bethe, Z. Phys. **71**, 205 (1931).
[26]M. Gaudin, Saclay Report CFA-N-155 CEA-N-1559 (1), (2), (3), 1973.
[27]E. H. Lieb, Phys. Rev. Lett. **18**, 692, 1046 (1967); **19**, 108 (1967).
[28]R. J. Baxter, Ann. Phys. (N.Y.) **70**, 229 (1972).
[29]H. S. M. Coxeter and W. O. S. Moser, *Generators and Relations for Discrete Groups* (Springer-Verlag, Berlin, 1972).
[30]U. Abresch, A. Bovier, O. Lechtenfeld, M. Luling, V. Rittenberg, and G. Weymans, J. Math. Phys. **21**, 2481 (1980).
[31]F. Y. Wu, Phys. Rev. B **4**, 3212 (1971);L. P. Kadanoff and F. J. Wegner, Phys. Rev. B **4**, 3898 (1971).
[32]J. Ashkin and E. Teller, Phys. Rev. **64**, 178 (1943).
[33]A. Krazer, *Lehrbuch der Thetafunktionen* (Teubner, Leipzig, 1903).
[34]J. Igusa, *Theta Functions* (Springer-Verlag, New York, 1972).
[35]S. Koizumi, J. Fac. Sci. Tokyo Sec. 1 **24**, 115 (1977).
[36]The analyticity of $g^{\gamma}(u)$ is a consequence of (VI.28) together with the fact that the functions $C^{\gamma}_{\lambda}(u)$ are analytic. This conclusion can be drawn in the case of one complex variable only. If we were dealing with functions of several variables, the analog of (VI.28) would read

$$C^{\gamma}_{0}(u) = -\frac{1}{\Theta_{2\tau}\begin{bmatrix} -\bar{\delta}'/2 \\ -\delta' \end{bmatrix}(u+2\eta)} \sum_{\lambda \in Z^{L}_{1}\setminus\{0\}} \times C^{\gamma}_{\lambda}(u)\Theta_{2\tau}\begin{bmatrix} -(\bar{\delta}'+\lambda)/2 \\ -\bar{\gamma}' \end{bmatrix}(u+2\eta)$$

From this it cannot be concluded that all the

$$g^{\gamma}_{\lambda}(u) \equiv C^{\gamma}_{\lambda}(u) \quad \Theta_{2\tau}\begin{bmatrix} -\bar{\delta}'/2 \\ -\bar{\gamma}' \end{bmatrix}(u+2\eta)$$

are analytic. Since the analyticity of $g^{\gamma}_{\lambda}(u)$ is crucial for the rest of the proof, the generalization is not possible.

Symmetry and fermion degeneracy on a lattice

H. Raszillier
Physikalisches Institut der Universität Bonn, Bonn, Federal Republic of Germany

(Received 3 May 1982; accepted for publication 27 August 1982)

We consider the general form of the finite difference approximation to the Dirac (Weyl) Hamiltonian on a lattice and investigate systematically the dependence on symmetry of the number of particles described by it. To a lattice with given symmetry, expressed by its crystallographic space group, there corresponds a minimal number of particles, which are associated with prescribed points of momentum space (the unit cell of the reciprocal lattice). We show in tables, using the existing detailed descriptions of the space groups, our results for all the relevant (symmorphic) symmetry groups. Only for lattice Hamiltonians with a momentum dependent mass term can this degeneracy be reduced or eliminated without reducing the symmetry.

PACS numbers: 05.50. + q, 11.30. − j, 11.10. − z

I. INTRODUCTION

During the last several years there has been considerable effort to include fermions in a formulation of lattice (gauge) field theory and to investigate its nonperturbative effects. Various descriptions of fermions on a lattice[1–3] have been proposed, but all met with certain difficulties already at the level of free fermions. These difficulties have their origin in the periodicity of the Hamiltonian on the reciprocal lattice, the momentum space, and consist of the fact that one has too may modes (particle degeneracy) or loses invariance properties of the theory if one reduces these modes. The precise number of modes depends on the lattice symmetry and the structure of the Hamiltonian. Periodicity alone, without symmetry arguments, leads, e.g., to the result that a lattice cannot carry only one massless (Weyl) particle if the Hamiltonian is not too nonlocal.[4,5] These particles have to be at least two in number, and this minimal number is reached[4,5] by Hamiltonians with very low lattice symmetry. One can easily see that a minimum of symmetry of the Hamiltonian, as, e.g., invariance with respect to space reflection already increases this number in d dimensions to 2^d particles.

Symmetry considerations have already played a role in the description of degeneracy of particular lattice fermion Hamiltonians[3,6,7] and in its partial reduction. In this note we investigate the effect of symmetry on the particle degeneracy of the Hamiltonian on a lattice from a general point of view, which is similar in spirit to work performed in solid state physics[8,9] on critical points in spectra of lattice vibrations.

We adopt the usual (group theoretic) definition of spin-$\frac{1}{2}$ fermions in terms of a representation of the Poincaré group

$$(U(\Lambda,a)\psi)(x) = \mathscr{D}(\Lambda)\psi(\Lambda^{-1}x - \Lambda^{-1}a) \tag{1.1}$$

and the spectrum condition (in the Hamiltonian form)

$$i\frac{\partial\psi}{\partial t} = H\psi, \quad H = \alpha\cdot P + \beta m. \tag{1.2}$$

On a lattice we have instead of the Poincaré group a discrete subgroup of the Euclidean group in three dimensions, which contains the lattice translations, i.e., a crystallographic space group. If we would use the Lagrangian form in (Euclidean) four dimensions we would arrive at the four-dimensional space groups.

The number of particles (,one,) present in the continuum is specified by the propagator of the Dirac Hamiltonian (1.2): It is equal to the number of points where the vector $p\equiv(p_1, p_2, p_3)$ vanishes, which is also (here) equal to the number of nondegenerate, $\det(\partial^2 E^2/\partial p_i \partial p_j)\neq 0$, minima of $E^2 = p^2 + m^2$.

On the lattice the partial derivatives $P_k = -i\partial/\partial x_k$ in (1.2) are substituted by approximating finite-difference expressions. In momentum space, the reciprocal lattice, there will appear then, as multiplication factors of the matrices α_i, instead of the components of p three real periodic functions $V_i(p)$ and in addition a real periodic mass function as factor of β. Their precise form depends on the way one chooses to approximate: by nearest neighbors only or by including also higher order differences, possibly in various directions. The particle content of the lattice Hamiltonian is then given by the structure of the propagator of $\alpha\cdot V(p) + \beta m(p)$, i.e., by the number of (nondegenerate) minima of $E^2 = \Sigma_k V_k^2(p) + m^2(p)$ in the unit cell of the reciprocal lattice. This number is related, because of

$$\frac{\partial E^2}{\partial p_i} = 2\left(\sum_k V_k(p)\frac{\partial V_k}{\partial p_i} + m(p)\frac{\partial m}{\partial p_i}\right),$$

to the number of zeros of the vector function $V(p) = (V_1(p), V_2(p), V_3(p))$ and critical points of the scalar function $m(p)$ [zeros of grad $m(p)$]. Generally, the number and position of points where $V(p)$ or grad $m(p)$ vanishes depends on the forms of the Hamiltonian, i.e., of $V(p)$ and $m(p)$, but it turns out that for any symmetry (crystallographic space group) there is a minimal set of points, determined entirely by it, where any function $V(p)$ and grad $m(p)$ respecting this symmetry has to vanish. The determination of these sets, to which the degeneracies due to symmetry are connected, is the aim of this paper. As long as $m(p) = \text{const}$, all zeros of $V(p)$ are minima of E^2 (particles) and the minimal particle degeneracy is given by this minimal set of zeros of $V(p)$, prescribed by symmetry. The connection between this set of zeros and the particle degeneracy may, however, be weakened by momentum dependent mass terms $m(p)$. Then, in a symmetric zero point of $V(p)$ and grad $m(p)$, $\partial^2 E/\partial p_i \partial p_j$ becomes

 0022-2488/83/030642-09$02.50

$$2\left(\sum_k \frac{\partial V_k}{\partial p_i}\frac{\partial V_k}{\partial p_j} + m\frac{\partial^2 m}{\partial p_i \partial p_j}\right)$$

and only the minimum of $m(p)$ has to remain a minimum of E^2. This is, e.g., what happens with $V_i(p) = \sin 2\pi p_i$ and $m(p) = (m^2 + \Sigma_i(1 - \cos 2\pi p_i)^2)^{1/2}$ in $E^2 = 4\Sigma_i \sin^2\pi p_i + m^2$.

The paper is organized as follows: We first summarize (Sec. II), with emphasis on symmetry considerations, the description of a fermion on a lattice in coordinate space and in momentum space (the unit cell of the reciprocal lattice or, equivalently, the torus). Then, we apply, in Sec. III, to our situation a theorem on zeros of continuous real vector functions (vector fields), required by symmetry, and consider (Sec. IV) two examples from this point of view. In Sec. V we consider such finite-difference approximations to the continuum Hamiltonian that the vector $V(p)$ is a gradient. Then it turns out that the degree of degeneracy is at least 2^d in d dimensions (8 for $d = 3$), independently of any symmetry. When, however, symmetry already leads to a degeneracy of at least 8, then gradients are compatible with it and need not increase it further. The zeros of gradients are also relevant for the location of the critical points of the mass term $m(p)$. Finally we comment (Sec. VI) on the possibility of (local) reduction of the degeneracy.

The information on crystallographic space groups, needed in order to determine the particle degeneracy required by symmetry and the location of particles in momentum space, can be taken over from the so-called international tables for these groups. For convenience of the reader we have compiled this information in Tables I–VII, in a form appropriate to our purpose, for the 73 relevant (out of the total of 230) symmetry groups.

II. DIRAC FERMIONS ON A LATTICE

The Dirac Hamiltonian (1.2) acts as a multiplication operator $\alpha\cdot p + \beta m$ in momentum space. The only zero of the vector field p associatd with it, $p = 0$, has the remarkable property, of interest to us, that it is isolated: It represents by itself an orbit of the rotation group, whereas all its neighboring points belong to larger orbits (spheres). The concept of orbit will play an important role in the following: the orbit of a point under a group is the whole set of points into which it is transformed by the group operations.[10]

Invariance of H under three-dimensional Euclidean transformations (R,a), composed of a rotation R and a translation a, $(R,a)x = Rx + a$, in coordinate space $x \equiv (x_1, x_2, x_3)$,

$$U(R,a)HU^{-1}(R,a) = H \tag{2.1}$$

is obeyed due to the relations

$$\mathscr{D}^{-1}(R)\alpha_i\mathscr{D}(R) = R_{ij}\alpha_j,$$
$$\mathscr{D}^{-1}(R)\beta\mathscr{D}(R) = \beta. \tag{2.2}$$

The lattice L is defined in terms of (primitive) vectors e_1,e_2,e_3 $[e_1(e_2\times e_3)\neq 0]$,

$$L = \{v = n_1e_1 + n_2e_2 + n_3e_3;\ n_i = \text{integers}\}. \tag{2.3}$$

With it we associate a crystallographic space group G_s [10,11]; it is a discrete subgroup of the three-dimensional Euclidean group, which contains the lattice translations as an invariant subgroup. The 230 crystallographic space groups are of two types,[10,11] and between them there is an essential difference. The so-called symmorphic groups (there are 73 of them) have only elements (R, t) which can be written as compositions of pure translations $(0,t)$ and pure (proper or improper) rotations $(R,0)$ belonging to the group. Since by operations of such a group a lattice point is always transformed again into a lattice point, the lattice wavefunction ψ is defined precisely on the lattice (2.3). The other type, the nonsymmorphic space groups, contain besides elements of the kind just described, also elements (R,τ) consisting of combined rotations and translations, such that the translation $(0,\tau)$ is smaller than a primitive translation and neither $(0,\tau)$ nor the rotation $(R,0)$, taken separately, are symmetry operations. By operatins of this kind a lattice is not transformed into itself, but into another lattice, superposed on it. So the wavefunction ψ has to be defined in these cases on a number of superposed lattices, which exceeds by one the number of nonprimitive translations $(0,\tau)$ appearing in operations (R,τ) of such a group. Because of the difference between these two types of symmetry groups we consider them separately.

(*a*) *Symmorphic case*: Since we may specify all vectors (2.3) by their integer components,

$$v \Leftrightarrow (n_1,n_2,n_3) \tag{2.4}$$

the wavefunction $\psi(v)$ may be considered as defined on the set $\mathbb{Z}^3$ of triples of integer numbers: $\psi(n) = \psi(n_1,n_2,n_3)$. Its transformation properties

TABLES I–VII. Isolated orbits of the symmorphic space groups (critical points required by symmetry) on the unit cell of the reciprocal lattice (torus T^3). The coordinates are given in units of π in the range $0<\theta_i<2\pi$. The space groups and the little (stability) groups of the orbit points are given in the Schoenflies notation. The lattices are primitive (P), one-face-centered (C), all-face-centered (F), body-centered (I), or rhomohedral (R). Pairs of space groups differing in the way their point group acts on the lattice are distinguished by a prime (′) attached to the lattice. The reciprocal of the hexagonal lattice (P) is denoted by P^x. Points in an orbit are separated by a comma (,); orbits by semicolons (;).

TABLE I. Triclinic system.

Space group	Lattice	Reciprocal lattice	Critical points required by symmetry	Little group of the critical points	Number of these critical points Of given little group	Total
C_1^1	P	P				—
C_i^1	P	P	000; 111; 100; 010; 001; 110; 101; 011;	C_i	8	8

TABLE II. Monoclinic system.

Space group	Lattice	Reciprocal lattice	Critical points required by symmetry	Little group of the critical points	Number of these critical points Of given little group	Total
C_2^1	P	P				—
C_2^3	C	C				—
C_s^1	P	P				—
C_s^3	C	C				—
C_{2h}^1	P	P	000; 111;			
			100; 010; 001;			
			110; 101; 011;	C_{2h}	8	8
C_{2h}^3	C	C	000; 110; 001; 111;	C_{2h}	4	
			100, 010; 101, 011;	C_i	4	8

$$(U(R,t)\psi)(v) = \mathscr{D}(R)\psi(R^{-1}v - R^{-1}t) \tag{2.5}$$

are obtained by restriction of (1.1) to the space group elements. Through Fourier series $\psi(v)$ may be transformed to the momentum space, i.e., as a periodic function

$$\psi(p) = \sum_v e^{iv\cdot p}\psi(v) \tag{2.6}$$

on the reciprocal lattice defined by the primitive vectors f_i:

$$2\pi f_i\cdot e_j = \delta_{ij}, \quad p \equiv (p_1,p_2,p_3)\sum_i p_i f_i,$$
$$= -\tfrac{1}{2} < p_i \leqslant \tfrac{1}{2}, \tag{2.7}$$

or, equivalently, as a function

$$\psi(\theta) \equiv \psi(\theta_1,\theta_2,\theta_3) = \sum_n e^{in\cdot\theta}\psi(n),$$
$$\theta_i = 2\pi p_i, \quad -\pi < \theta_i \leqslant \pi, \tag{2.6'}$$

on the three-dimensional torus T^3. Then $\psi(p)$ obeys

$$(U(R,t)\psi)(p) = \mathscr{D}(R)e^{it\cdot p}\psi(R^{-1}p). \tag{2.8}$$

We now define the lattice Hamiltonian. As already mentioned in the Introduction, by a finite-difference approximation of the partial derivatives we are led to a periodic Hamiltonian of the reciprocal lattice. Its form will depend on the way we approximate. The simplest approximation, by nearest neighbors on a cubic lattice (with $e_i^2 = 1$): $-i\partial\psi/\partial x_j \to (1/2i)[\psi(x+e_j) - \psi(x-e_j)]$ would give a Hamiltonian of the form

$$H = \alpha\cdot V(p) + \beta m(p), \tag{2.9}$$

with a vector field $V_i(p) = \sin 2\pi p_i$ and a constant mass term $m(p) = m$. Inclusion of higher order differences would change $V_i(p)$ to more complicated periodic functions, which may depend not only on a single variable each, if one includes differences not only in the direction of the partial derivative to be approximated. The presences of higher-order difference terms reflect themselves in terms of higher powers of $\sin\theta_i$ or $\cos\theta_i$ in $V(p)$.

Invariance of the Hamiltonian (2.9) under operations of a symmorphic space group leads, due to (2.2), to the relations

TABLE III. Orthorhombic system.

Space group	Lattice	Reciprocal lattice	Critical points required by symmetry	Little group of the critical points	Number of these critical points Of given little group	Total
D_2^1	P	P	000; 001; 010; 100;			
			011; 101; 110; 111;	D_2	8	8
D_2^6	C	C	000; 110; 111; 001;	D_2	4	4
D_2^7	F	I	000; 011; 101; 110;	D_2	4	4
D_2^8	I	F	$000; \frac{111}{222}; 111; \frac{333}{222};$	D_2	4	4
C_{2v}^1	P	P				—
C_{2v}^{11}	C	C				—
C_{2v}^{14}	C'	C'				—
C_{2v}^{18}	F	I				—
C_{2v}^{21}	I	F				—
D_{2h}^1	P	P	000; 100; 010; 001;			
			110; 011; 101; 111;	D_{2h}	8	8
D_{2h}^{19}	C	C	000; 110; 001; 111;	D_{2h}	4	
			100, 010; 101, 011;	C_{2h}	4	8
D_{2h}^{23}	F	I	000; 100; 010; 110;	D_{2h}	4	
			001, 011, 101, 111;	C_i	4	8
D_{2h}^{25}	I	F	000; 111;	D_{2h}	2	
			100, 011; 010, 101;			
			001, 110;	C_{2h}	6	
			$\frac{111}{222}, \frac{333}{222};$	D_2	2	10

TABLE IV. Tetragonal system.

Space group	Lattice	Reciprocal lattice	Critical points required by symmetry	Little group of the critical points	Number of these critical points Of given little group	Total
C_4^1	P	P				—
C_4^5	F	I				—
S_4^1	P	P	000; 001; 110; 111;	S_4	4	4
S_4^2	F	I	000; 110; $\frac{31}{22}1$; $\frac{13}{22}1$;	S_4	4	4
C_{4h}^1	P	P	000; 001; 110; 111;	C_{4h}	4	
			010, 100; 011, 101;	C_{2h}	4	8
C_{4h}^5	F	I	000; 110;	C_{4h}	2	
			100, 010;	C_{2h}	2	
			$\frac{31}{22}1, \frac{13}{22}1$;	S_4	2	
			001, 011, 101, 111;	C_i	4	10
D_4^1	P	P	000; 001; 110; 111;	D_4	4	
			100, 010; 101, 011;	D_2	4	8
D_4^9	F	I	000; 110;	D_4	2	
			010, 100; $\frac{31}{22}1, \frac{13}{22}1$;	D_2	4	6
C_{4v}^1	P	P				—
C_{4v}^9	F	I				—
D_{2d}^1	P	P	000; 111; 001; 110;	D_{2d}	4	
			100, 010; 101, 011;	D_2	4	8
D_{2d}^5	P'	P'	000; 110; 001; 111;	D_{2d}	4	4
D_{2d}^9	F	I	000; 110; $\frac{31}{22}1$; $\frac{13}{22}1$;	D_{2d}	4	4
D_{2d}^{11}	F'	I'	000; 110;	D_{2d}	2	
			010, 100; $\frac{31}{22}1, \frac{13}{22}1$;	D_2	4	6
D_{4h}^1	P	P	000; 001; 110; 111;	D_{4h}	4	
			011, 101; 010, 100;	D_{2h}	4	8
D_{4h}^{17}	F	I	000; 110;	D_{4h}	2	
			010, 100;	D_{2h}	2	
			$\frac{31}{22}1, \frac{13}{22}1$;	D_{2d}	2	
			001, 111, 011, 101;	C_{2h}	4	10

$$V_i(Rp) = R_{ij}V_j(p),$$
$$m(Rp) = m(p), \tag{2.10}$$

which have to be obeyed for any element $(R,0)$ of this group. The elementary consequence of these relations are the facts that the zeros of $V(p)$, to which we have related the particles described by the Hamiltonian (2.9), represent orbits of the space group in the unit cell of the reciprocal lattice, and that $m(p)$ is constant on each of these orbits.

We have thus connected the degeneracy in the lattice Hamiltonian (2.9) with the orbit structure induced by the space group in the unit cell of the reciprocal lattice (Brillouin zone).

(*b*) *Nonsymmorphic case*: Here the lattice is given by the subgroup of translations, but the point set on which the wavefunction has to be defined, the orbit of a lattice point under the action of the space group, is a superposition of several lattices. The vectors specifying the additional lattices are of the form

$$\tau_j = \alpha_j^i e_i, \quad 0 < \alpha_j^i < 1, \quad j = 1,...,k. \tag{2.11}$$

To each lattice there corresponds, according to (2.6), a unit cell (a torus) of its own reciprocal lattice. If k is the number of vectors (2.11), then we have as the whole momentum space the sum of unit cells on $k+1$ reciprocal lattices. Although the transformation (2.5) now mixes the lattices, the Hamiltonian constructed out of pure translations does not mix them: On each lattice there acts a separate Hamiltonian. The form of these Hamiltonians will be related by the symmetry operations containing nonprimitive translations which transform one lattice into another. This will, however, not influence in any important way the orbit structure in the unit cells of the reciprocal lattices, because all orbits generated in one unit cell by the symmetry operations of the (symmorphic) subgroup, which does not mix the lattices, appear multiplied by the same factor, the number of lattices.

The orbits are thus completely characterized for any nonsymmorphic group by those of its (maximal) symmorphic subgroup, which is one of the groups considered before, and by the number of lattices.

III. INVARIANT SMOOTH VECTOR FIELDS

By (2.10) we are given a periodic real vector function on the reciprocal lattice i.e., a real vector field on the torus T^3, which is invariant with respect to a space group. We shall now assume that it is smooth, i.e., continuous together with its partial derivatives up to a certain order. This assumption is a condition of a certain degree of locality: The nonlocality

TABLE V. Trigonal system.

Space group	Lattice	Reciprocal lattice	Critical points required by symmetry	Little group of the critical points	Number of these critical points Of given little group	Total
C_3^1	P	P^x				—
C_3^4	R	R				—
C_{3i}^1	P	P^x	000; 001;	C_{3i}	2	
			100, 010, 110; 101, 011, 111;	C_i	6	8
C_{3i}^2	R	R	000; 111;	C_{3i}	2	
			011, 101, 110; 100, 010, 001;	C_i	6	8
D_3^1	P	P^x	000; 001;			
			$\frac{24}{33}0; \frac{24}{33}1; \frac{42}{33}0; \frac{42}{33}1;$	D_3	6	6
D_3^2	P'	P'^x	000; 001;	D_3	2	2
D_3^7	R	R	000; 111;	D_3	2	2
C_{3v}^1	P	P^x				—
C_{3v}^2	P'	P'^x				—
C_{3v}^5	R	R				—
D_{3d}^1	P	P^x	000; 001;	D_{3d}	2	
			$\frac{24}{33}0, \frac{42}{33}0; \frac{24}{33}1, \frac{42}{33}1;$	D_3	4	
			100, 010, 110; 101, 011, 111;	C_{2h}	6	12
D_{3d}^3	P'	P'^x	000; 001;	D_{3d}	2	
			100, 010, 110; 101, 011, 111;	C_{2h}	6	8
D_{3d}^5	R	R	000; 111;	D_{3d}	2	
			100, 010, 001; 011, 101, 110;	C_{2h}	6	8

TABLE VI. Hexagonal system.

Space group	Lattice	Reciprocal lattice	Critical points required by symmetry	Little group of the critical points	Number of these critical points Of given little group	Total
C_6^1	P	P^x				—
C_{3h}^1	P	P^x	000; 001;			
			$\frac{22}{33}0; \frac{44}{33}0; \frac{22}{33}1; \frac{44}{33}1;$	C_{3h}	6	6
C_{6h}^1	P	P^x	000; 001;	C_{6h}	2	
			$\frac{22}{33}0, \frac{44}{33}0; \frac{22}{33}1, \frac{44}{33}1;$	C_{3h}	4	
			100, 010, 110; 101, 011, 111;	C_{2h}	6	12
D_6^1	P	P^x	000; 001;	D_6	2	
			$\frac{22}{33}0, \frac{44}{33}0; \frac{22}{33}1, \frac{44}{33};$	D_3	4	
			100, 010, 110; 101, 011, 111;	D_2	6	12
C_{6v}^1	P	P^x				—
D_{3h}^1	P	P^x	000; 001;			
			$\frac{22}{33}0; \frac{44}{33}0; \frac{22}{33}1; \frac{44}{33}1;$	D_{3h}	6	6
D_{3h}^3	P'	P'^x	000; 001;	D_{3h}	2	
			$\frac{22}{33}0, \frac{44}{33}0; \frac{22}{33}1, \frac{44}{33}1;$	C_{3h}	4	6
D_{6h}^1	P	P^x	000; 001;	D_{6h}	2	
			$\frac{22}{33}0, \frac{44}{33}0; \frac{22}{33}1, \frac{44}{33}1;$	D_{3h}	4	
			100, 010, 110; 101, 011, 111;	D_{2h}	6	12

TABLE VII. Cubic system.

Space group	Lattice	Reciprocal lattice	Critical points required by symmetry	Little group of the critical points	Number of these critical points Of given little group	Total
T^1	P	P	000; 111;	T	2	
			011, 101, 110;			
			100, 010, 001;	D_2	6	8
T^2	F	I	000;	T	1	
			011, 101, 110;	D_2	3	4
T^3	I	F	000; 111; $\frac{111}{222}; \frac{333}{222}$;	T	4	4
T_h^1	P	P	000; 111;	T_h	2	
			011, 101, 110;			
			100, 010, 001;	D_{2h}	6	8
T_h^3	F	I	000;	T_h	1	
			011, 101, 110;	D_{2h}	3	
			111, 100, 010, 001;	C_{3i}	4	8
T_h^5	I	F	000; 111;	T_h	2	
			$\frac{111}{222}, \frac{333}{222}$;	T	2	
			100, 010, 001,			
			110, 101, 011;	C_{2h}	6	10
O^1	P	P	000; 111;	O	2	
			011, 101, 110;			
			100, 010, 001;	D_4	6	8
O^3	F	I	000;	O	1	
			011, 101, 110;	D_4	3	
			111, 100, 010, 001;	D_3	4	
			$1\frac{13}{22}, \frac{1}{2}\,1\frac{3}{2}, \frac{13}{22}\,1$,			
			$1\frac{31}{22}, \frac{3}{2}\,1\frac{1}{2}, \frac{31}{22}\,1$;	D_2	6	14
O^5	I	F	000; 111;	O	2	
			$\frac{111}{222}, \frac{333}{222}$;	T	2	
			100, 010, 001,			
			110, 101, 011;	D_2	6	10
T_d^1	P	P	000; 111;	T_d	2	
			011, 101, 110;			
			100, 010, 001;	D_{2d}	6	8
T_d^2	F	I	000;	T_d	1	
			011, 101, 110;	D_{2d}	3	
			$\frac{113}{244}, \frac{311}{424}, \frac{131}{442}$,			
			$\frac{113}{424}, \frac{311}{442}, \frac{131}{244}$;	S_4	6	10
T_d^3	I	F	000; 111; $\frac{111}{222}; \frac{333}{222}$;	T_d	4	4
O_h^1	P	P	000; 111;	O_h	2	
			011, 101, 110;			
			100, 010, 001;	D_{4h}	6	8
O_h^5	F	I	000;	O_h	1	
			011, 101, 110;	D_{4h}	3	
			111, 100, 010, 001;	D_{3d}	4	
			$1\frac{13}{22}, \frac{1}{2}\,1\frac{3}{2}, \frac{13}{22}\,1$,			
			$1\frac{31}{22}, \frac{3}{2}\,1\frac{1}{2}, \frac{31}{22}\,1$;	D_{2d}	6	14
O_h^9	I	F	000; 111;	O_h	2	
			$\frac{111}{222}, \frac{333}{222}$;	T_d	2	
			100, 010, 001;			
			011, 101, 110;	D_{2h}	6	10

of the finite-difference approximation to the continuum Hamiltonian increases with the inclusion of higher order differences. If there are only terms up to a certain order, then $V_i(p)$ are trigonometric polynomials, which are smooth. But we can admit even trigonometric series (difference terms of any order) as long as the coefficients of the higher-order terms decrease fast enough, so that the series represent smooth periodic functions.

If the vector field $V(p)$ is smooth, the following statement is true[12]: It has to vanish on the orbits which are isolat-

ed in their strata.

We comment on the terminology of this statement: The unit cell of the reciprocal lattice (the torus) may be divided into orbits. Any point of a orbit has essentially (up to conjugation) the same stability (little) group. But even more orbits may have essentially the same stability group. These orbits are then called of the same type. A stratum is the collection of all orbits of the same type. An orbit is called isolated if in its neighborhood there is no element of an orbit of the same type with its own.

We thus need to determine, for all crystallographic space groups, those orbits in the unit cell of the reciprocal lattice (or lattices), which are isolated in their strata. These orbits represent the points where for any lattice Hamiltonian with the symmetry of the space group and constant mass term $m(p) = m$ the energy squared $E^2 = \Sigma_k V_k^2(p) + m^2$ has to have a minimum. There may be, of course, in addition, other minima depending on the concrete form of $V(p)$. These isolated orbits can be identified from the international crystallographic tables.[13] We give them here for the symmorphic space groups in Tables I–VII; for the nonsymmorphic groups they are, up to the multiplicity factor, the same as those of their maximal symmorphic subgroups.

In order to identify zeros of $V(p)$ with particles in a Hamiltonian with $m(p) = \text{const}$, we have to ask that they be nondegenerate [$\det(\partial V_i/\partial p_j) \neq 0$]. For all zeros (critical points) in the same orbit $\det(\partial V_i/\partial p_j)$ has the same value; in particular, the sign of it, $\kappa = \text{sgn}\det(\partial V_i/\partial p_j)$, the index of the vector field in the critical point, depends only on the orbit. By the theorem of Poincaré and Hopf,[5,14] the sum of indices in all (nondegenerate) zeros on the torus has to be zero. Therefore, the possibility that the minimal set of zeros of a vector field $V(p)$ consists only of nondegenerate zeros depends on the orbit structure of the symmetry group. There are groups where this indeed cannot happen—T^2, T_h^5, O^5, T_d^2, and O_h^9—because one of their orbits has more points than all the others together, and so the Poincaré–Hopf theorem cannot be obeyed. Hamiltonians with these symmetries must then have an even higher particle degeneracy than given by the isolated orbits: 14, 16, 16, 14, and 16, respectively, since the smallest (nonisolated) orbits that can be added in order to fulfill this theorem are of order 4 + 6, 6, 6, 4, and 6, respectively.

IV. EXAMPLES

As a first example we consider the space group O_h^1 of the primitive cubic lattice. The vector field $V(p)$ with this symmetry has to have the zeros $(0,0,0);(\frac{1}{2},\frac{1}{2},\frac{1}{2});(0,0,\frac{1}{2}),(0,\frac{1}{2},0),(\frac{1}{2},0,0);(0,\frac{1}{2},\frac{1}{2}),(\frac{1}{2},0,\frac{1}{2}),(\frac{1}{2},\frac{1}{2},0)$. These are the points also reproduced by $V(p) = \sin 2\pi p_i$,[3] which has this symmetry.

As a second example we consider the space group of the two-dimensional hexagonal lattice, determined by the group D_{6h}^1. Its isolated orbits (in two dimensions) are: $(0,0);(0,\frac{1}{2}),(\frac{1}{2},0),(\frac{1}{2},\frac{1}{2});(\frac{1}{3},\frac{1}{3}),(-\frac{1}{3},-\frac{1}{3})$. These are precisely the zeros found in Ref. 6 with a Hamiltonian of this symmetry. We identify $\theta_1 = -k_3$, $\theta_2 = -k_2$, $-\pi < \theta_i \leqslant \pi$, and get the vector field in coordinates $\theta_i = 2\pi p_i$ as

$$V_1 = (1/\sqrt{3})[\sin(\theta_1 + \theta_2) + \sin\theta_1],$$
$$V_2 = \tfrac{1}{3}[\sin(\theta_1 + \theta_2) + 2\sin\theta_2 - \sin\theta_1].$$

V. GRADIENT FIELDS

We have connected the particle content of a Hamiltonian (2.9) with the number of minima of $E^2 = \Sigma_i V_i^2(p) + m^2(p)$, which we required to be nondegenerate. So, we are, in fact, interested in a special type of critical points of the function E^2 on the unit cell of the reciprocal lattice (torus T^3). This function is symmetric, $E^2(Rp) = E^2(p)$, because of (2.10) and its critical points (zeros of grad E^2) have to include the isolated orbits determined in Sec. III. What kind of critical points the points of various orbits can be has to be determined by consistency of the orbit structure with the constraints coming from Morse theory. We may assume that E^2 is a Morse function (a smooth function which has only isolated and nondegenerate critical points), since we already had demanded that its minima be nondegenerate. Then Morse theory allows to apply to it the following statement[15]: The number ν_k of critical points ($\text{grad} f = 0$) of (Morse) index (Footnote 16) $k = 0,\ldots,n$ of a Morse function f on a compact manifold of dimension n, characterized by the Betti numbers b_k ($\equiv b_{n-k}$), is limited by the inequalities

$$\nu_k - \nu_{k-1} + \cdots + (-1)^k \nu_0 \geqslant b_k - b_{k-1} + \cdots + (-1)^k b_0 \tag{5.1}$$

and by the equality

$$\sum_0^n (-1)^k \nu_k = \sum (-1)^k b_k \equiv \chi \tag{5.2}$$

expressed in terms of the Euler characteristics χ. For the torus T^n the Betti numbers are $b_k = \binom{n}{k}$ and therefore the total number of critical points of a gradient is at least 2^n:

$$\sum_0^n \nu_k \geqslant \sum_0^n b_k = 2^n. \tag{5.3}$$

For the dimensions we are considering ($n = 2,3$) one gets from (5.1) and (5.2):

$$\begin{aligned} n = 2:\quad & \nu_0 \geqslant 1, \\ & \nu_1 \geqslant \nu_0 + 1, \\ & \nu_2 = \nu_1 - \nu_0; \end{aligned} \tag{5.4}$$

$$\begin{aligned} n = 3:\quad & \nu_0 \geqslant 1, \\ & \nu_1 \geqslant \nu_0 + 2, \\ & \nu_2 \geqslant \nu_1 - \nu_0 + 1, \\ & \nu_3 = \nu_2 - \nu_1 + \nu_0. \end{aligned} \tag{5.5}$$

Comparing the orbit structures from Tables I–VII with the inequalities (5.5) and taking into account that points of one orbit have the same Morse index, one gets that the set of points given by the isolated orbits, supplemented eventually by additional orbits as described in Sec. III, can be minimal sets of critical points of E^2. Since (0,0,0) is always an orbit, Morse theory as such does not demand more than one minimum for E^2. A possible obstruction to this only lies in the structure of E^2 itself. Namely, as long as $m(p) = \text{const}$,

$$\left(\frac{\partial^2 E^2}{\partial p_i \partial p_j}\right) = \left(2\sum_k \frac{\partial V_k}{\partial p_i}\frac{\partial V_k}{\partial p_j}\right) > 0$$

in the critical points of $V(p)$, and each one is a minimum of E^2, as already discussed, and E^2 cannot saturate the inequalities (5.5). In order to reduce this high number of minima, one has to allow for momentum dependent mass terms $m(p) \neq \text{const.}$[1] Then although

$$\left(\frac{\partial^2 E^2}{\partial p_i \partial p_j}\right) = \left(2\sum_k \frac{\partial V_k}{\partial p_i}\frac{\partial V_k}{\partial p_j} + \frac{\partial^2 m^2}{\partial p_i \partial p_j}\right) > \left(\frac{\partial^2 m^2}{\partial p_i \partial p_j}\right),$$

the nature (index k) of the critical points may be determined by $m^2(p)$, and not by $V^2(p)$.

The fact that the minimal sets of critical points determined in Sec. III by the isolated orbits and the Poincaré–Hopf theorem are compatible with the (stronger) restrictions of Morse theory [equality (5.2) contains the information of this theorem] also means that, for the groups which have isolated orbits, the general vector fields $V(p)$ cannot reach smaller sets of nondegenerate critical points than the gradients do.

The examples considered in Sec. IV are, in fact, gradients. In the first example

$$V_i = \frac{\partial f}{\partial \xi_i}, \quad \theta_i = \xi_i, \tag{5.6}$$

$$f(\theta) = -\sum_i \cos\theta_i,$$

and in the second

$$V_i = \frac{\partial f}{\partial \xi_i}, \quad \theta_1 = \xi_1 - (1/\sqrt{3})\xi_2, \quad \theta_2 = (2/\sqrt{3})\xi_2, \tag{5.7}$$

$$f(\theta) = -(1/\sqrt{3})[\cos(\theta_1 + \theta_2) + \cos\theta_1 + \cos\theta_2].$$

In these examples the numbers ν_k are $\nu_0 = 1, \nu_1 = 3, \nu_2 = 3, \nu_3 = 1$ and $\nu_0 = 1, \nu_1 = 3, \nu_2 = 2$, respectively.

VI. REDUCTION OF DEGENERACY

One (obvious) way to reduce completely the particle degeneracy on the lattice is by projection on the corresponding subspaces defined by subsets, the so-called fundamental domains, of the unit cell of the reciprocal lattice. This is nonlocal on the lattice of coordinate space, essentially as nonlocal as the (discontinuous) vector $V_i(p) = p_i$.[2]

If one is prepared to accept momentum dependent mass terms,[1] then the reduction of particle degeneracy from the minimum number given by $V(p)$ is not difficult. One example for this has already been mentioned in the Introduction, with $V_i(p) = \sin 2\pi p_i$ and $m(p) = [m^2 + \Sigma_1^3(1 - \cos 2\pi p_i)^2]^{1/2}$, of symmetry O_h^1. The only critical points of $E^2 = \Sigma_i V_i^2(p) + m^2(p) = m^2 + 4\Sigma_i \sin^2\pi p_i$ are those required by O_h^1 symmetry. The minimum of E^2 is given by that of $m^2(p)$. It would also do to take[1] $m(p) = m + m_1\Sigma_i(1 - \cos 2\pi p_i)$, $m_1 > 1$.

A local reduction of the minimal degeneracy, which does not go through $m(p)$, has to make use of some nonprimitive translation invariance of the Hamiltonian on the reciprocal lattice. In order to allow this, the critical points should arrange themselves on a lattice. In several cases this turns out to be so, in many others (Table I–VII) not. If it happens to be the case, as it is for the space group O_h^1 of the primitive cubic lattice, then one has to pay the prize for the construction of a reduced reciprocal lattice by giving up partly symmetry. The point symmetry of the smaller lattice can only be the common stability group of the critical points on it. For O_h^1 this is, as one can see from Table VII, the group O_h, if in addition to (0,0,0) one also takes $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$ as a lattice point, D_{4h} if one takes $(0,0,\frac{1}{2}),(\frac{1}{2},\frac{1}{2},0)$ and $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$ as additional lattice points, or D_{2h} (the intersection of the three D_{4h} in O_h) if all eight critical points are taken as lattice points.

This argument only refers to point group operations and determines the point symmetry which may survive if one reduces the unit cell of the reciprocal lattice. Translation on the reduced reciprocal lattice has to be defined for the wavefunction in such a way that it is consistent with periodicity on the initial lattice and, as much as possible, with the point symmetry of the reduced lattice. It cannot be done by simply asking periodicity on the reduced lattice for the reason that it would relate the index $\kappa = \text{sgn det}(\partial V_i/\partial p_j)$ of the vector field $V(p)$ in the critical points in such a way that the Poincaré–Hopf theorem is contradicted. One has to allow for unitary (and Hermitian) transformations in the space of wavefunctions together with the translations on the reduced reciprocal lattice.

These transformations are obtained by considering the transformation properties induced through (2.2) by a point group on the algebra generated by the Dirac matrices α_i and β. One has to consider the common stability (little) group of the elements of its commuting subalgebras. The maximal commuting subalgebras are of order four and are generated, up to permutations of α_i's, by $1,\beta, -i\alpha_1\alpha_2, -i\alpha_1\alpha_2\beta$; $1,\alpha_3, -i\alpha_1\alpha_2, -i\alpha_1\alpha_2\alpha_3$; $1,\alpha_1, -i\alpha_3\beta, -i\alpha_1\alpha_3\beta$. Any point group is a subgroup of O_h or D_{6h}. In O_h the stability group of 1 and β is O_h, that of $-i\alpha_1\alpha_2, -i\alpha_1\alpha_2\beta$ only C_{4h}. Further, for α_3 and $-i\alpha_3\beta$ the group is C_{4v} and for $-i\alpha_1\alpha_2\alpha_3$ it is O. The algebra generated by $1, \beta$ is invariant with respect to O_h, but $1, \beta, -i\alpha_1\alpha_2, -i\alpha_1\alpha_2\beta$ conserves C_{4h} (C_{2h}), not D_{4h} (D_{2h}). The stability groups in D_{6h} are as follows: for 1 and β it is D_{6h}, for $-i\alpha_1\alpha_2$ and $-i\alpha_1\alpha_2\beta$ it is C_{2h}; further, for α_3 and $-i\alpha_3\beta$ it is C_{6v}, and for $-i\alpha_1\alpha_3, -i\alpha_1\alpha_3\beta$ also an (isomorphic) C_{2h}. Finally, the stability group of $-i\alpha_1\alpha_2\alpha_3$ is D_2.

A maximal commuting subalgebra of the Dirac algebra generates an abelian group. Its product with the group of translations on a reduced reciprocal lattice determined by critical points is generally not an invariance group of the Hamiltonian. Its maximal invariance subgroups perform the reduction to subspaces. In order that such an invariance group exist at all, the masses have to be degenerate: $m(p)$ has to be periodic on the reduced lattice.

In the example $V_i = \sin\theta_i, m = \text{const}$ the reduction is performed[3] by the projections

$$(P_\pm \psi)(\theta) = \tfrac{1}{2}[\psi(\theta) \pm i\alpha_1\alpha_2\beta\psi(\theta - \tau)],$$
$$(Q_\pm \psi)(\theta) = \tfrac{1}{2}[\psi(\theta) \pm i\alpha_1\alpha_2\psi(\theta - \tau')], \tag{6.1}$$
$$\tau = (0,0,\pi), \quad \tau' = (\pi,\pi,0),$$

$$(P_{\pm}\psi)(n) = \tfrac{1}{2}\,[1 + i\alpha_1\alpha_2\beta(-1)^{(n\cdot\tau)/\pi}]\psi(n), \tag{6.1'}$$

$$(Q_{\pm}\psi)(n) = \tfrac{1}{2}\,[1 + i\alpha_1\alpha_2(-1)^{(n\cdot\tau)/\pi}]\psi(n).$$

The symmetry left after projection is the tetragonal space group C^1_{4h} of the lattice $(2\mathbb{Z})^3$ of doubled size in coordinate space.

Since $V_i = \sin\theta_i$ has exactly the minimal number of particles for O^1_h, one cannot do better than this also in the general case of O^1_h symmetry with constant mass term: O^1_h symmetry is reduced by reduction of particles (from eight to two) to tetragonal symmetry C^1_{4h}.

VII. DISCUSSION

We have shown that a symmetry of the finite-difference Dirac Hamiltonian generates, if the lattice admits isolated points of special symmetry, a minimal number of particles which depends only on the crystallographic symmetry group. We have listed in detail the results for the symmorphic groups (Tables I–VII), the (semidirect) product of a point group and and the group of lattice translations. The nonsymmorphic groups lead, since they contain nonprimitive translations, to an increase of the sets (coordinate and momentum spaces) where the wavefunctions are defined (i.e., to a lattice multiplication). This has as a consequence that the minimal particle degeneracy is determined by the minimal number required by the symmorphic group of one lattice copy times the total number of lattice copies, equal to the number of nonprimitive translations.

There is a number of crystallographic space groups which do not generate isolated orbits (Tables I–VII). Since any of these groups leads to (nonisolated) orbits consisting of one single point each, its minimal partical content is the minimum required by the Poincaré–Hopf theorem: two.

As to the reduction of degeneracy it can be performed by momentum-dependent mass terms or by (local) projection on subspaces. In the first case it conserves symmetry, whereas in the second it is accompanied by a reduction of symmetry. In the example we have considered[3] the reduction is from O^1_h to C^1_{4h}.

We have to state, finally, that the degeneracy described is connected to the form (2.9) of the lattice Hamiltonian, which is expressed in terms of a symmetric vector field $V(p)$ and a symmetric scalar function $m(p)$. All results derive from the fact that both these types of functions have minimal sets of critical points, imposed by the symmetry group and depending on it. This point of view could be relaxed[17] (see also Ref. 18) in the sense that one could start from a reducible (eight-component) continuum equation, which already describes two fermions (species doubling), and make on it a lattice approximation. As far as the particle spectrum is concerned, the result on gets in the simplest case[17] is equivalent to $E^2 = \Sigma_i V_i(p) + m^2$ with $V_i = \sin\theta_i$ and with a mass function $m(p) = [m^2 + \Sigma_i(1 - \cos\theta_i)^2]^{1/2}$. Beyond this, it can be considered as the interpretation of the square root in the manner of Dirac: $(m^2 + \Sigma_i(1 - \cos\theta_i)^2)^{1/2} \to \Sigma\alpha_i(1 - \cos\theta_i) + \beta m$ to give $H = \Sigma_i(\alpha_i \times 1)\sin\theta_i + \Sigma_i(\beta \times \alpha_i)(1 - \cos\theta_i) + (\beta \times \beta)m$.

ACKNOWLEDGMENTS

I would like to thank G. V. Gehlen and V. Rittenberg for useful comments.

[1] K. G. Wilson, in *Phenomena in Subnuclear Physics*, edited by A. Zichichi (Plenum, New York, 1977) (Erice, 1975).
[2] S. D. Drell, M. Weinstein, and S. Yankielowicz, Phys. Rev. D **14**, 487 (1976).
[3] L. Susskind, Phys. Rev. D **16**, 3031 (1977).
[4] H. B. Nielsen and M. Ninomiya, Nucl. Phys. B **185**, 20 (1981).
[5] L. H. Karsten, Phys. Lett. B **104**, 315 (1981).
[6] A. Chodos and J. B. Healy, Nucl. Phys. B **127**, 426 (1977).
[7] H. S. Sarachandra, H. J. Thun, and P. Weisz, Nucl. Phys. B **192**, 205 (1981).
[8] L. Van Hove, Phys. Rev. **89**, 1189 (1953).
[9] J. C. Phillips, Rhys. Rev. **104**, 1263 (1956).
[10] W. Miller, Jr., *Symmetry Groups and Their Applications* (Academic, New York, 1972).
[11] G. Burns and A. M. Glazer, *Space Groups for Solid State Scientists* (Academic, New York, 1978).
[12] L. Michel, Rev. Mod. Phys. **52**, 617 (1980).
[13] N. F. M. Henry and K. Lonsdale, *International Tables for X-Ray Crystallography* (Kynoch, Birmingham, 1969), Vol. I; *Internationale Tabellen zur Bestimmung von Kristallstrukturen, Bd. I: Gruppentheoretische Tafeln* (Borntraeger, Berlin, 1935).
[14] J. Milnor, *Topology from the Differentiable Point of View* (University Press of Virginia, Charlottesville, 1965).
[15] J. Milnor, *Morse Theory* (Princeton U. P., Princeton, N.J., 1973).
[16] The Morse index k of an isolated nondegenerate critical point is the number of negative eigenvalues of the Hesse matrix of f in this point; in its neighborhood f can be written as $f = -x_1^2 - \cdots - x_k^2 + x_{k+1}^2 + \cdots + x_n^2$ in suitable local coordinates. The index κ of the gradient is related to it by $\kappa_k = (-1)^k$.
[17] P. Becher, Phys. Lett. B **104**, 221 (1981).
[18] J. M. Rabin, "Homology Theory of Lattice Fermion Doubling," Yale University Preprint YTP81-29, December 1981.

Mechanisms of the "fermion–boson reciprocity"

Piotr Garbaczewski
Institute of Theoretical Physics, University of Wrocław, 50-205 Wrocław, Poland

(Received 8 December 1981; accepted for publication 22 January 1982)

We analyze two completely integrable quantized systems: the nonlinear Schrödinger and sine–Gordon ones, with the aim explicitly to recover the (spin-$\frac{1}{2}$ approximation) mechanisms responsible for their being allowed in the continuum limit equivalence with Fermi systems.

PACS numbers: 11.10.Lm, 11.30.Pb

I. MOTIVATION

Since Refs. 1–3, a relationship between the (lattice) spin *xyz* Heisenberg and the (continuum) massive Thirring models has become well known. On the other hand, there is a correspondence[4] between the massive Thirring model and the sine–Gordon model, in which, upon adjustments of suitable parameters, the Schwinger functions of both models are identical, and hence the spectra are the same. In fact, the latter coincide with the semiclassical spectrum of the sine–Gordon system. In an *indirect* way we thus have a relationship between the spin-$\frac{1}{2}$ *xyz* and the sine–Gordon model.

There is an isolated attempt[5] to establish the correspondence *directly*, via the so-called spin-$\frac{1}{2}$ approximation of the sine–Gordon system on a lattice. One finds that in a Hilbert space (state space) of the sine–Gordon model there exists a proper subspace for which the reduction of the sine–Gordon Hamiltonian to this subspace makes it identical to the appropriate spin-$\frac{1}{2}$ *xyz* type Hamiltonian. The spin-$\frac{1}{2}$ approximation idea is then to recover these properties of the original Bose system which allow it to approximate, at least weakly, the related spin-$\frac{1}{2}$ or Fermi system (see also Ref. 6). One knows that if the spectra of both are identical, we can expect that then H_{sG} commutes with the underlying projection P: $[H_{sG},P]=0$ and then either $H_{sG}=PH_{sG}P$ or $H_{sG}=PH_{sG}P+(1-P)H_{sG}(1-P)$. *However this simple property does not hold true while on a lattice, in general, but it can be recovered in the continuum limit.*

In the present paper we exploit the results of the quantum inverse scattering analysis for the nonlinear Schrödinger and sine–Gordon models, to demonstrate explicitly the mechanisms of the "fermion–boson reciprocity" known to occur for both in the continuum limit, with the emphasis on the involved spin-$\frac{1}{2}$ approximation formalism.

In case of the nonlinear Schrödinger model, it is important to have available explicit formulas for Bethe-type eigenfunctions of the Hamiltonian. In the case of the sine–Gordon system such functions are not explicitly known, but nevertheless we are able to recover the spin-$\frac{1}{2}$ approximation scheme by analyzing the algebraized form of the Bethe ansatz and relating it to this of the spin-$\frac{1}{2}$ *xyz* Heisenberg model.

II. NONLINEAR SCHRÖDINGER MODEL IN THE REPULSIVE CASE

(A) Let us consider the famous Hamiltonian Bose system in 1 + 1 dimensions:

$$H=\frac{1}{2}\int\nabla\phi^{*}\nabla\phi\,dx+\frac{1}{2}\int\int\phi^{*}(x)\phi^{*}(y)V(x-y)\phi(x)\phi(y)\,dxdy,$$

$$[\phi(x),\phi^{*}(y)]_{-}=\delta(x-y),\quad [\phi(x),\phi(y)]_{-}=0, \tag{2.1}$$

$$i\dot{\phi}=[\phi,H]_{-}=-\frac{1}{2}\nabla\phi^{2}+\int V(x-y)\phi^{*}(y)\phi(y)\,dy\,\phi(x). \tag{2.2}$$

The latter equation of motion, in case of $V(x-y)=c\delta(x-y)$ is known as the nonlinear Schrödinger equation.

Let us choose a countable set of square integrable functions in $\mathcal{L}^{2}(R^{1})$ subject to the restrictions:

$\{\varphi_s\}$, $\mathrm{supp}\varphi_s=\Delta_s$, $\Delta_s\cap\Delta_{s+1}\neq\emptyset$, $\Delta_s\cap\Delta_t=\emptyset$ otherwise, Δ_s being a closed interval in R^{1}. Now let us approximate $\phi(x)$ by

$$\phi(x)\cong\sum_s a_s\varphi_s(x),\quad [a_s a_t^{*}]_{-}=\delta_{st},\quad [a_s,a_t]_{-}=0. \tag{2.3}$$

Then (2.1) reads

$$H\cong\sum T_{mn}a_m^{*}a_n+\sum G_{mnjk}a_m^{*}a_n^{*}a_j a_k, \tag{2.4}$$

where

$$T_{mn}=\frac{1}{2}\int\nabla\bar{\varphi}_n\nabla\varphi_m\,dx,$$

$$G_{mnjk}=\frac{1}{2}\int\int\bar{\varphi}_m(x)\bar{\varphi}_n(y)V(x-y)\varphi_j(x)\varphi_k(y)\,dxdy \tag{2.5}$$

and because of (2.2) only the nearest neighbor exchange integrals remain (provided they exist):

$$T_{mm}=\frac{1}{2}\int|\nabla\varphi_n|^{2}dx=\alpha_m,$$

$$G_{mmmm}=\frac{c}{2}\int|\varphi_m|^{4}dx=\beta_m,$$

$$T_{mn}=\delta_{n\,m-1}\alpha_{m-1}+\delta_{n\,m+1}\alpha_{m+1}, \tag{2.6}$$

$$G_{mnjk}=\delta_{n\,m-1}\delta_{jm}\delta_{k\,m-1}\beta_{m-1}+\delta_{n\,m+1}\delta_{jm}\delta_{k\,m+1}\beta_{m+1}$$

with

$$\alpha_{m\pm1}=\frac{1}{2}\int\nabla\bar{\varphi}_m\nabla\varphi_{m\pm1}dx,$$

$$\beta_{m\pm1}=\frac{c}{2}\int|\varphi_m|^{2}|\varphi_{m\pm1}|^{2}dx. \tag{2.7}$$

 0022-2488/83/030651-08$02.50

Hence, using (2.6), the formula (2.4) becomes

$$H \cong \sum_m [a_m^* (\alpha_{m+1} a_{m+1} + \alpha_{m-1} a_{m-1} + \alpha_m a_m) + a_m^{*2} (\beta_{m+1} a_{m+1}^2 + \beta_{m-1} a_{m-1}^2 + \beta_m a_m^2)]. \quad (2.8)$$

(B) Let $|0\rangle$ be a Fock vacuum for the representation of the CCR algebra generated by (2.3): $a_k|0\rangle = 0\ \forall k$. We denote

$$p_k = :\exp(-a_k^* a_k): + a_k^* :\exp(-a_k^* a_k): a_k \quad (2.9)$$

and observe that

$$p_k^* = p_k, \quad p_k^2 = p_k, \quad [p_k, p_l] = 0, \quad k \neq l; \quad (2.10)$$

and, moreover,

$$p_k a_k^* p_k \equiv \sigma_k^+, \quad p_k a_k p_k \equiv \sigma_k^-, \quad (2.11)$$

i.e.,

$$\begin{aligned} (\sigma_k^+)^2 &= 0 \\ &= (\sigma_k^-)^2 \\ &= [\sigma_k^-, \sigma_l^+]_- \\ &= [\sigma_k^-, \sigma_l^-]_- \\ &= [\sigma_k^+, \sigma_l^+]_-, \quad k \neq l, \\ & [\sigma_k^-, \sigma_k^+]_+ = p_k, \end{aligned} \quad (2.12)$$

thus giving rise to spin-$\frac{1}{2}$ Pauli operators for the linear chain. An operator

$$P = \prod_k p_K \quad (2.13)$$

is a well-defined projection operator in the Fock space of our Bose system,[7] and allows us to consider a projected Hamiltonian $H_F = PHP$ of the (spin-$\frac{1}{2}$ approximation) type of Ref. 6:

$$H = H_F + (1-P)HP + PH(1-P) + (1-P)H(1-P). \quad (2.14)$$

Obviously (2.8) reads

$$H_F = PHP = \sum_m [\sigma_m^+ (\alpha_{m+1} \sigma_{m+1}^- + \alpha_{m-1} \sigma_{m-1}^- + \alpha_m \sigma_m^-)]. \quad (2.15)$$

If, for any vector $|\lambda\rangle$, we have $P|\lambda\rangle = |\lambda\rangle$, then

$$H|\lambda\rangle = H_F|\lambda\rangle + (1-P)H|\lambda\rangle,$$

i.e.,

$$(1-P)H|\lambda\rangle = 0 \Rightarrow H|\lambda\rangle = H_F|\lambda\rangle. \quad (2.16)$$

Suppose that H is diagonalized, and $|\lambda\rangle$ is an eigenvector of H: $H|\lambda\rangle = E|\lambda\rangle$. Then,

$$P|\lambda\rangle = |\lambda\rangle \Rightarrow H|\lambda\rangle = H_F|\lambda\rangle = E|\lambda\rangle, \quad (2.17)$$

and because of $[P, H] = 0$ we can equivalently write

$$H = H_F + (1-P)H(1-P)]. \quad (2.18)$$

(C) Notice that (2.15) is a familiar form of Schultz's Hamiltonian[8] for the impenetrable Bose lattice gas in one space dimension, provided one adopts

$$\alpha_{m+1} = \alpha_{m-1} = -1/2M\delta^2, \quad \alpha_m = 1/M\delta^2 \quad (2.19)$$

with δ being the lattice spacing, M the single-particle mass, and the finite volume restriction (to N sites) being conventionally imposed. One knows that after making the Jordan Wigner transformation from spins-$\frac{1}{2}$ to anticommuting (CAR) variables, followed by the appropriate canonical transformation, a resulting Hamiltonian describes the free Fermi gas. In the continuum limit the Girardeau eigenvalues and wavefunctions arise.[9] The free Bose gas Hamiltonian should then be used instead of H:

$$H = H_0 = -\frac{1}{2M} \int_0^L \phi^*(x) \frac{\partial^2}{\partial x^2} \phi(x)\, dx \quad (2.20)$$

provided one imposes the domain restriction:

$$\left[\int dx' \phi^*(x) \phi^*(x') \delta(x-x') \phi(x') \phi(x) \right] |\phi\rangle = 0, \quad (2.21)$$

which eliminates the coupling term

$$\tfrac{1}{2} c \int dx \int dx' \phi^*(x) \phi^*(x') \delta(x-x') \phi(x') \phi(x)$$

of the original nonlinear Schrödinger problem.

In Girardeau's model it is obvious that $[H, P]_- = 0$ and hence all eigenvectors of H satisfying $P|\lambda\rangle = |\lambda\rangle$ are just those obseying (2.21). Here P projects on a subspace of the (Hilbert) eigenspace of H, on which (2.21) holds true. Let us, however, mention that the role of the coupling constant c, by virtue of (2.21) is missing, while from Refs. 10 and 11 one knows that at $c = 0$ the Bose (free) gas appears while at $c = \infty$ H converts into $H_F = PHP$, i.e., the equivalent free Fermi gas appears.

(D) Let H be a nonlinear Schrödinger model Hamiltonian. For any value of $c > 0$, the normalized in-eigenstates of H read as follows[12,13]:

$$\begin{aligned} |\phi(k_1, \cdots, k_n)\rangle_{\text{in}} &= R(k_1) \cdots R(k_n) |0\rangle \\ &= \int \left[\prod_{i=1}^n dx_i \exp(ik_i x_i) \right] \\ &\times \left\{ \prod_{1 \leq j < i \leq n} \left[\Theta(x_j - x_i) + \Theta(x_i - x_j) \frac{k_i - k_j - ic}{k_i - k_j + ic} \right] \right\} \\ &\times \phi^*(x_1) \cdots \phi^*(x_n) |0\rangle, \quad k_1 < k_2 < \cdots < k_n. \end{aligned} \quad (2.22)$$

Here the Zamolodchikov operators $R(k), R^*(k)$ can be introduced to algebraize the Bethe ansatz for the eigenvectors of H, and one has

$$R(k_i) R(k_j) = \frac{k_i - k_j - ic}{k_i + k_j + ic} R(k_j) R(k_i) \quad (2.23)$$

and

$$R(k_i) R^*(k_j) = \frac{k_i - k_j - ic}{k_i - k_j + ic} R^*(k_j) R(k_i) + 2\pi \delta(k_i - k_j) \quad (2.24)$$

together with

$$[H, R^*(k)]_- = k^2 R^*(k), \quad (2.25)$$

which implies that $|\phi(k_1, \cdots, k_n)\rangle$ is an eigenvector of H. By making use of (2.22), one finds easily that at $c = 0$

$$|\phi(k_1,\cdots,k_n)_B$$
$$= \int\left\{\prod_{i=1}^{n} dx_i\left[\exp\left(i\sum_i k_i x_i\right)\right]\right\}\phi^*(x_1|\cdots\phi^*(x_n|0)$$
$$= a^*(k_1)\cdots a^*(k_n)|0), \tag{2.26}$$

which corresponds to the free Bose gas problem, while the free Fermi gas problem appears at $c=\infty$, when

$$|\phi(k_1,\cdots,k_n))_F = \int dx_1\cdots\int dx_n\left[\exp\left(i\sum_i k_i x_i\right)\right]$$
$$\times\sigma(x_1,\cdots,x_n]\phi^*(x_1)\cdots\phi^*(x_n)|0) \tag{2.27}$$

and

$$\sigma(x_1,\cdots,x_n) = \prod_{1\le j<i\le n}[\Theta(x_i-x_j)-\Theta(x_j-x_i)] \tag{2.28}$$

satisfies

$$\sigma_n^3=\sigma_n,\quad \sigma_n^2(1-\sigma_n^2)=0,$$
$$\sigma(\cdots x_i\cdots x_j\cdots) = -\sigma(\cdots x_j\cdots x_i\cdots). \tag{2.29}$$

Obviously $R=R(\phi^*,\phi)$ follows from (2.22); see Refs 12, 13. Girardeau's version of the free Bose gas wavefunctions arriving at the free Fermi gas is realized here by means of σ_n. This is a special case of the general isomorphisms between linear spaces of symmetric functions and antisymmetric functions[14] see also Ref. 7.

Let us emphasize that the general theory of Ref. 14 allows us to recover Fermi states of Bose systems in more than one space dimension. Obviously the simple multiplicative alternation σ_n is insufficient for such a purpose. Because of (2.29) for an n-point function f_n, one has

$$f_n = \sigma_n^2 f_n + (1-\sigma_n^2)f_n = \overset{1}{f}_n + \overset{2}{f}_n \tag{2.30}$$

so that if $f_n = f(x_1,\ldots,x_n)$ is a symmetric function, its antisymmetric image á la Girardeau[9] is

$$\sigma_n \overset{1}{f}_n = \sigma(x_1,\ldots,x_n)\overset{1}{f}(x_1,\ldots,x_n).$$

In our case, the vector

$$\sigma(k_1,\ldots,k_n)|\phi(k_1,\ldots,k_n)\rangle_F$$
$$= \int dx_1\cdots\int dx_n\left[\exp\left(i\sum_j k_j x_j\right)\right]$$
$$\times\sigma^2(x_1,\ldots,x_n)\phi^*(x_1)\cdots\phi^*(x_n|0) \tag{2.31}$$

is a typical element of the range of a projection operator $P=1_F$ of Ref. 7, i.e., $P|\overset{1}{\phi}\rangle_B = |\overset{1}{\phi}\rangle_B$. If adopted to our notation, 1_F reads

$$1_F = \sum_n \frac{1}{n!}\int dx_1\cdots\int dx_n\sigma^2(x_1,\ldots,x_n)\phi^*(x_1)\cdots\phi^*(x_n)$$
$$\times{:}\exp\left[-\int dy\,\phi^*(y)\phi(y)\right]{:}\phi(x_1)\cdots\phi(x_n). \tag{2.32}$$

It is an operator unit of the CAR algebra (a Fock representation):

$$[b(x),b^*(y)]_+ = \delta(x-y)1_F,\quad [b(x),b(y)]_+ = 0$$
$$b(x) = \sum_m (1+n)^{1/2}\int dy_1\cdots\int dy_n$$
$$\times\sigma(y_1,\ldots,y_n)\sigma(x,y_1,\ldots,y_n) \tag{2.33}$$
$$\times\phi^*(y_1)\cdots\phi^*(y_n){:}\exp\left[-\int dz\,\phi^*(z)\phi(z)\right]{:}$$
$$\times\phi(x)\phi(y_1)\cdots\phi(y_n).$$

By inspection one easily verifies that

$$\overset{1}{|\phi)}_B = \sigma(k_1,\ldots,k_n)|\phi(k_1,\ldots,k_n))_F$$
$$= \sigma(k_1,\ldots,k_n)\int dx_1\cdots\int dx_m\left[\exp\left(i\sum_j k_j x_j\right)\right]$$
$$\times b^*(x_1|\cdots b^*(x_n)|0)$$
$$= \sigma(k_1,\ldots,k_n)b^*(k_1)\cdots b^*(k_n)|0). \tag{2.34}$$

Obviously at $c=0$, the $\{R,R^+\}$ algebra generates the CCR algebra, while at $c=\infty$ the CAR's arise; see (2.22)–(2.24).

By using the Jordan Wigner transformation, (2.34) can be replaced by

$$\overset{1}{|\phi)}_B = \sigma^+(k_1)\cdots\sigma^+(k_n|0\rangle, \tag{2.35}$$

where $\overset{1}{|\phi)}_B$ vanishes if any two k's coincide, and is permutation invariant. In this connection compare also Ref. 15, where the quantum Gel'fand–Levitan transform problem has been solved for the nonlinear Schrödinger field: $\phi(x)=\phi[R^*,R](x)$ in the limit of the infinite repulsion reduces to the Jordan Wigner mapping from the CAR generators $\{R,R^*\}$ to Pauli operators $\{\phi,\phi^*\}$.

(E) From Ref. 6 one knows that a family $\{|\phi(k_1,\ldots,k_n))\}_{n=0,1,2,\cdots}$ of vectors forms a complete eigenfunction system for the nonlinear Schrödinger model Hamiltonian if $c\geqslant 0$. However, one should realize that a continuous transition from $c=0$ to $c=\infty$ results in the contraction of the dynamically accessible state space for our Bose system from the whole of the Fock space $\mathcal{F}$ ($c=0$) to its proper subspace $P\mathcal{F}=\mathcal{F}_F$.

This peculiar property is precisely a (very special) realization of the Bose→Fermi metamorphosis phenomenon via the spin-½ approximation procedure, as described in Refs. 16, 6.

Recall that $P=1_F$, (2.32), is a continuous version of the lattice projection P, which replaces each single Bose degree of freedom by the two-level (spin-½) degree in the system. The formulas (2.14)–(2.18) apply both to the continuous and lattice cases.

Generally upon the parametric dependence $H=H(\lambda)$ we expect to arrive at $[H(\infty),P]_-=0$ so that $H(\infty)=PH(\infty)P+(1-P)H(\infty)(1-P)$ may hold true either on the lattice level or, if impossible, on the continuum level. If all eigenstates of $H(\infty)$ belong to the range of P, then $H(\infty)=H_F=PH(\infty)P$, which is the case for the nonlinear Schrödinger model at $c=\infty$. Then a "fermion–boson reciprocity" idea apparently applies. Otherwise, the diagonalization of H_F does not resolve the spectral problem for H, and one is forced to diagonalize H itself to recover a complete set of eigenvectors. Obviously, it is frequently much more favor-

able to diagonalize the Bose Hamiltonian than the related Fermi one. If one succeeds with H, then a solution for H_F is immediate, provided $[H,P]=0$. This idea lies at the foundations of the boson expansion methods as applied in the many-body (Fermi or finite spin) systems in the diagonalization-before-the-projection procedure of Refs. 17,18.

III. THE QUANTUM SINE–GORDON SYSTEM VS SPIN-½ *xyz* HEISENBERG MODEL: PRELIMINARIES

(A) As far as the nonlinear Schrödinger model is concerned, we are fortunate to have available an exact spectrum of the problem with an explicit form of the eigenvectors. This fact enables one to control *how the Bose system is becoming equivalent to Fermi one as c varies from 0 to* ∞, and to notice that $[H,P]\neq 0$ except for $c=0$ and $c=\infty$.

For the sine–Gordon system, despite the spectacular equivalence with the massive Thirring model[4] and the demonstration of complete integrability via the quantum inverse scattering transform method,[19] even the knowledge of the Bethe ansatz states for the massive Thirring model did not enable one to construct explicitly Bethe wavefunctions. On the other hand, we know from Ref. 5 that the sine–Gordon Hamiltonian, while put on a lattice with the nearest-neighbor coupling gradient term, in the spin-½ approximation (at $T\geqslant 0$) reduces to the spin-½ *xyz* problem. The latter under the weak anisotropy assumption in the continuum limit is known to reproduce the WKB spectrum of the sine–Gordon system.[1]

In addition, our lattice analysis of Ref. 5 suggests that in the continuum limit the ("fermion–boson reciprocity") statement:

$$H_{sG}=PH_{sG}P\equiv H_{xyz} \tag{3.1}$$

is not realized if at all, but rather

$$H_{sG}\equiv H_{xyz}+(1-P)H_{sG}(1-P). \tag{3.2}$$

Consequently, a selection of the appropriate subspace of the sine–Gordon state space is necessary to arrive at the Coleman's equivalence on the level of irreducible fields. To support this conjecture, we shall make an analysis of the available inverse scattering results for the sine–Gordon[19] and the spin-½ *xyz*[20] models with the emphasis on some limiting properties of both. Both systems are considered on the finite lattice of length L and spacing δ. This restriction is essential because the continuum limit of the sine–Gordon model does not exist for all coupling constant values.

Within the inverse scattering formalism, the basic object is the one-parameter family of the local transition matrices $L_n(\lambda)$, $n=1,\ldots,N=L/\delta$ which give rise to the matrix:

$$T(\lambda)=L_N(\lambda)\cdots L_1(\lambda)=\begin{pmatrix}A(\lambda) & B(\lambda)\\ C(\lambda) & D(\lambda)\end{pmatrix}, \tag{3.3}$$

where the following commutation relations hold true:

$$\begin{aligned}
&[A(\lambda),A(\mu)]_-=[B(\lambda),B(\mu)]_-=0,\\
&B(\lambda)A(\mu)=b(\lambda,\mu)B(\mu)A(\lambda)+c(\lambda,\mu)A(\mu)B(\lambda),\\
&B(\mu)D(\lambda)=b(\lambda,\mu)B(\lambda)D(\mu)+c(\lambda,\mu)D(\lambda)B(\mu),\\
&c(\lambda,\mu)[C(\lambda),B(\mu)]_-=b(\lambda,\mu)[A(\mu)D(\lambda)-A(\lambda)D(\mu)],\\
&c(\lambda,\mu)[D(x),A(\mu)]_-=b(\lambda,\mu)[B(\mu)C(\lambda)-B(\lambda)C(\mu)].
\end{aligned} \tag{3.4}$$

As a consequence of (3.4), one has, for example,

$$[\mathcal{T}(\lambda),\mathcal{T}(\mu)]_-=0,\mathcal{T}(\lambda)=A(\lambda)+D(\lambda). \tag{3.5}$$

We wish to have represented relations (3.4) in terms of the operator algebra in the N-particle Hilbert space

$$\mathcal{H}=\prod_{i=1}^{N}{}^{\otimes} h_i,\quad h_i=h\quad \forall i,$$

including a reference state Ω such that

$$\begin{aligned}
&C(\lambda)\Omega=0,\quad A(\lambda)\Omega=\exp[a(\lambda)N]\Omega,\\
&D(\lambda)\Omega=\exp[d(\lambda)N]\Omega.
\end{aligned} \tag{3.6}$$

The particular choice of functions $b(\lambda,\mu),c(\lambda,\mu),a(\lambda),d(\lambda)$ determines the model of interest.

The eigenstates of $\mathcal{T}(\lambda)$ are constructed from Ω as follows:

$$|\lambda_1,\ldots,\lambda_n)=\prod_{i=1}^{n}B(\lambda_i)\Omega \tag{3.7}$$

provided the following (periodicity) condition holds true:

$$\exp\{[a(\lambda_k)-d(\lambda_k)]N\}=\prod_{\substack{i=1\\ j\neq k}}^{n}\frac{c(\lambda_j,\lambda_k)}{c(\lambda_k,\lambda_j)},\quad k=1,2,\ldots,n. \tag{3.8}$$

Then (3.7) is an eigenvector of the operator $\mathcal{T}(\lambda)$ with the eigenvalue

$$\Lambda(\lambda,\lambda_1,\ldots,\lambda_n)=e^{a(\lambda)N}\prod_{j=1}^{n}\frac{1}{c(\lambda_j,\lambda)}+e^{d(\lambda)N}\prod_{j=1}^{n}\frac{1}{c(\lambda,\lambda_j)}. \tag{3.9}$$

(B) The sine–Gordon model in the above framework is defined[19] as follows,

$$\begin{aligned}
&b(\lambda,\mu)=\frac{\sinh(v-v')}{\sinh(v-v'+i\gamma)},\quad c(\lambda,\mu)=\frac{i\sin\gamma}{\sinh(v-v'+i\gamma)},\\
&v=\ln\lambda,\quad v'=\ln\mu,\\
&a(\lambda)=d(\bar{\lambda})=\tfrac{1}{8}m^2\delta^2\cosh(2v-i\delta)
\end{aligned} \tag{3.10}$$

provided one starts from the field equation

$$\varphi_{tt}-\varphi_{xx}=(m^2/\beta)\sin\beta\varphi,\quad \gamma=\tfrac{1}{8}\beta^2, \tag{3.11}$$

and then discretizes the problem: $L=N\delta$, δ being the lattice spacing. The local transition operator $L_n(\lambda)$ appears then in the form

$$L_n(\lambda)=\begin{pmatrix}\mu_n, & (m/4)[\lambda v_n^*-(1/\lambda)v_n]\\ (m/4)[(1/\lambda)v_n^*-\lambda v_n], & \mu_n^*\end{pmatrix}, \tag{3.12}$$

where the Weyl commutation relations

$$\begin{aligned}
&v_nv_m=v_mv_n,\quad u_nu_m=u_mu_n,\quad v_n^*v_m=v_mv_n^*,\\
&u_n^*u_m=u_mu_n^*,\quad u_nv_n=e^{-i\gamma}v_nu_n,\quad u_nv_n^*=e^{i\gamma}v_n^*u_n
\end{aligned} \tag{3.13}$$

are satisfied by operators

$$\begin{aligned}
&v_n=\delta\exp\left[(i\beta/2\delta)\int_{x_n}^{(x_n+\delta)}\varphi(x)dx\right],\\
&p_n=\tfrac{1}{4}\beta\int_{x_n}^{x_n+\delta}\dot{\varphi}(x)dx,\quad u_n=\exp(-ip_n),\\
&[\varphi(x),\varphi(y)]_-=0=[\dot{\varphi}(x),\dot{\varphi}(y)]_-,\\
&[\varphi(x),\dot{\varphi}(y)]_-=i\delta(x-y).
\end{aligned} \tag{3.14}$$

Let us emphasize that the defining commutation relations (3.9) can be rewritten in the compact tensor product form:

$$R(\lambda,\mu)[T(x)\otimes T(\mu)] = [T(\mu)\otimes T(\lambda)]R(\lambda,\mu) \quad (3.15)$$

where $R = R(\lambda,\mu)$ is a 4×4 matrix with c-valued elements:

$$R_{sG} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & b & c & 0 \\ 0 & c & b & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad b = b(\lambda,\mu),\ c = c(\lambda,\mu). \quad (3.16)$$

To specify a concrete model in the inverse scattering formalism, one must thus know R and solutions of the eigenvalue equations (3.6). This amount of knowledge suffices to reconstruct the whole model.

(C) Now we shall recall the basic results for the spin-$\frac{1}{2}$ xyz Heisenberg model, with a final goal of identifying these eigenvectors for the sine–Gordon system, which belong to the range of the projection P of Sec. I, in

$$\mathscr{H} = \prod_{i=1}^{N}{}^{\otimes} h_i, \quad h_i = \mathscr{L}^2(R^1) \quad \forall i,$$

after making a transition to the continuum. Without entering into the detailed (Baxter's) parametrization, let us notice that the form of the local transition matrix[20]

$$L_n(\lambda) = \begin{pmatrix} w_4\sigma_n^4 + w_3\sigma_n^3, & w_1\sigma_n^1 - iw_2\sigma_n^2 \\ w_1\sigma_n^1 + iw_2\sigma_n^2, & w_4\sigma_n^4 - w_3\sigma_n^4 \end{pmatrix}, \quad (3.17)$$

where $[\sigma_n^j,\sigma_m^j]_- = 0$, $n\neq m$, and operators $\{\sigma_n^j\}^{j=1,2,3,4}$ are equivalent to 2×2 Pauli matrices assigned to the nth site, allows a representation of $L_n(\lambda)$ in

$$P\mathscr{H} = \prod_{i=1}^{N}{}^{\otimes}(ph)_i \equiv \prod_{i=1}^{N}{}^{\otimes} C_i^2$$

provided we introduce

$$\begin{aligned} &\varphi_k = \delta^{-1/2}\int_{R^1}\varphi(x)\chi_k(x)\,dx = 2^{-1/2}(a_k^* + a_k), \\ &\dot{\varphi}_k = (1/i2^{1/2})(a_k^* - a_k), \\ &\chi_k(x) = 1 \quad \text{for } x\in(x_k, x_k+\delta) \\ &\qquad = 0 \quad \text{otherwise.} \end{aligned} \quad (3.18)$$

Then define $\{\sigma_n^j\}_{n=1,2,\cdots}^{i=1,2,3}$ and $\sigma_n^4 = p_n$ via formulas (1.10)–(1.13). However, we are still far from any relationship with the sine–Gordon system, especially because the R matrix reads

$$R_{xyz} = \begin{pmatrix} a & 0 & 0 & d \\ 0 & b & c & 0 \\ 0 & c & b & 0 \\ d & 0 & 0 & a \end{pmatrix}, \quad a\neq 1, \quad d\neq 0,$$

$$\begin{aligned} a &= \Theta(2\eta)\Theta(\lambda-\mu)H(\lambda-\mu+2\eta), \\ b &= H(2\eta)\Theta(\lambda-\mu)\Theta(\lambda-\mu+2\eta), \\ c &= \Theta(2\eta)H(\lambda-\mu)\Theta(\lambda-\mu+2\eta), \\ d &= H(2\eta)H(\lambda-\mu)H(\lambda-\mu+2\eta) \end{aligned} \quad (3.19)$$

with H, Θ being Jacobi's eta and theta elliptic functions.

(D) One knows that, for each family $\{L_n(\lambda)\}_{n=1,\cdots,N}$ of local transition matrices for the xyz model, there exists a sequence $\{M_n^l(\lambda) = M_n^l(\lambda,s,t)\}$ of matrices with complex-valued matrix elements, being functions of two integers l and n, $n, l = 0, \pm 1 \pm 2,\cdots$, and two arbitrary complex parameters s and t (to be omitted below for simplicity) such that the operators

$$\begin{aligned} L_n^l(\lambda) &= L_n^l(\lambda,s,t) \\ &= M_{n+l}^{-1} L_n(\lambda) M_{n+l-1} \\ &= \begin{pmatrix} \alpha_n^l(\lambda) & \beta_n^l(\lambda) \\ \gamma_n^l(\lambda) & \delta_n^l(\lambda) \end{pmatrix} \end{aligned} \quad (3.20)$$

with $M_n^l := M_{n+(l-1)}$ serve as the new local transition operators for the xyz model, with a resulting one-parameter family of transition operators:

$$T^l(\lambda) = \prod_{n=1}^{N} L_n^l(\lambda) = \begin{pmatrix} A^l(\lambda) & B^l(\lambda) \\ C^l(\lambda) & D^l(\lambda) \end{pmatrix}. \quad (3.21)$$

Let us consider a single-site Hilbert space $h = \mathscr{L}^2(R^1)$ and let $\{e_i\}_{i=0,1,\cdots}$ be a complete orthonormal system in it with $a^*e_0 = e_1$, $ae_0 = 0$, $e_n = (1/\sqrt{n!})a^{*n}e_0$. We denote $e_0 = e^-$, $e_1 = e^+$. In this notation, we introduce a one-parameter family of vectors[20]

$$\omega^l = H(s + 2(\eta+l)\eta - \eta)e^+ + \Theta(s + 2(\eta+l)\eta - \eta)e^-, \quad (3.22)$$

given as linear combinations of e^+, e^- in terms of Jacobi elliptic functions of the modulus k, η being one more (real) parameter. Each vector ω^l, (3.22) is annihilated by a respective γ^l of (3.20), and, moreover, we have

$$\begin{aligned} &\gamma^l(\lambda)\omega^l = 0, \\ &\alpha^l(\lambda)\omega^l = h(\lambda+\eta)\omega^{l-1}, \\ &\delta^l(\lambda)\omega^l = h(\lambda-\eta)\omega^{l-1}, \end{aligned} \quad (3.23)$$

with

$$h(u) = \Theta(0)H(u)\Theta(u). \quad (3.24)$$

The above local formula allows a construction of the one-parameter family of N-particle vectors in

$$\begin{aligned} &P\mathscr{H} = \prod_{i=1}^{N}{}^{\otimes}(p\mathscr{L}^2(R^1))_i: \\ &\Omega^l = \omega_1^l\otimes\omega_2^l\otimes\cdots\otimes\omega_N^l, \quad \omega_i^l = \omega^l \quad \forall i \end{aligned} \quad (3.25)$$

such that operators $A^l(\lambda), B^l(\lambda), C^l(\lambda), D^l(\lambda)$ of (3.21) satisfy

$$\begin{aligned} &A^l(\lambda)\Omega^l = h^N(\lambda+\eta)\Omega^{l-1}, \\ &D^l(\lambda)\Omega^l = h^N(\lambda-\eta)\Omega^{l-1}, \\ &C^l(\lambda)\Omega^l = 0 \end{aligned} \quad (3.26)$$

for each l. The formula (3.21) can be rewritten as follows:

$$T^l(\lambda) = M_{N+l}^{-1}(\lambda)T(\lambda)M_l(\lambda) := T_{N+l,l}(\lambda), \quad (3.27)$$

where the notation

$$T_{k,l}(\lambda) = M_k^{-1}(\lambda)T(\lambda)M_l(\lambda) \quad (3.28)$$

is introduced, and the corresponding (operator-valued) matrix elements read $A_{kl}(\lambda)$, $B_{kl}(\lambda)$, $C_{kl}(\lambda)$, $D_{kl}(\lambda)$. By using these operators, one constructs a $1+n$ operator family of vectors:

$$\begin{aligned} &\Psi_l(\lambda_1,\cdots,\lambda_n) \\ &\quad = B_{l+1,l-1}(\lambda_1)\cdots B_{l+n,l-n}(\lambda_n)\Omega^{l-n}, \quad n = N/2. \end{aligned} \quad (3.29)$$

The transfer operator of the model,[20]

$$\mathrm{Tr}[T(\lambda)] = \mathscr{T}(\lambda)$$
$$= A(\lambda) + D(\lambda) = \mathscr{T}_{II}(\lambda) = A_{II}(\lambda) + D_{II}(\lambda) \quad (3.30)$$

satisfies the following eigenvalue equation:

$$\mathscr{T}(\lambda)\Psi_\Theta(\lambda_1,...,\lambda_n) = \Lambda(\Theta,\lambda,\lambda_1,...,\lambda_n)\Psi_\Theta(\lambda_1,...,\lambda_n), \quad (3.31)$$

where

$$\Psi_\Theta(\lambda_1,...,\lambda_n) = \sum_{l=-\infty}^{+\infty} e^{2\pi i l\Theta}\Psi_l(\lambda_1,...,\lambda_n),$$
$$\Lambda(\Theta,\lambda,\lambda_1,...,\lambda_n) = e^{2\pi i\Theta}\Lambda_1 + e^{-2\pi i\Theta}\Lambda_2,$$
$$\Lambda_1(\lambda,\lambda_1,...,\lambda_n) = h^N(\lambda+\eta)\prod_{k=1}^{n}\alpha(\lambda,\lambda_k), \quad (3.32)$$
$$\Lambda_2(\lambda,\lambda_1,...,\lambda_n) = h^N(\lambda-\eta)\prod_{k=1}^{n}\alpha(\lambda_k,\lambda),$$
$$\alpha(\lambda,\mu) = h(\lambda-\mu-2\mu)/h(\lambda-\mu), \quad n = N/2,$$

and by construction $\Psi_\Theta(\lambda_1,...,\lambda_n)\in P\mathscr{H}$.

The periodicity condition reads

$$\frac{h^N(\lambda_j+\eta)}{h^N(\lambda_j-\eta)} = e^{-4\pi i\Theta}\sum_{\substack{k=1\\k\neq j}}^{n}\frac{\alpha(\lambda_k,\lambda_j)}{\alpha(\lambda_j,\lambda_k)},$$
$$j = 1,2,...,n = N/2. \quad (3.33)$$

By exploiting a property $H(u) = -H(-u)$, $\Theta(u) = \Theta(-u)$, i.e., $h(u) = -h(-u)$, we can write this condition as follows:

$$\left[\frac{h(\lambda_j+\eta)}{h(\lambda_j-\eta)}\right]^N e^{4\pi i\Theta} = \sum_{\substack{k=1\\k\neq j}}^{n}\frac{h(\lambda_j-\lambda_k+2\eta)}{h(\lambda_j-\lambda_k-2\eta)}. \quad (3.34)$$

IV. FROM THE *xyz* MODEL TO THE SINE–GORDON MODEL: SPIN-$\frac{1}{2}$ APPROXIMATION IDEA REVIVED

(A) In the above the elliptic Jacobi functions $H(u)$ and $\Theta(u)$ are of the modulus k with $k = (1-l'')/(1+l'')$ $0 < l'' < 1$. We are interested in the properties of the *xyz* Heisenberg system while approaching the limit $(1-k')/(1+k') = l' \to 0$, which implies $k = (1-k'^2)^{1/2} \to 0$. The transition from the modulus k to l', under an assumption $l' \to 0$, can be equivalently rewritten as

$$h(u) = H(u,k)\Theta(u,k)\Theta(0,k) = (\mathrm{const})H(u,l') \quad (4.1)$$

(see Ref. 3). It allows us to evaluate the limiting properties of (3.34) upon the following change of variables (formerly introduced in Ref. 3)

$$(\lambda_i - \lambda_j) \to -\tfrac{1}{2}i(\beta_i - \beta_j),$$
$$\eta \to \tfrac{1}{2}(\pi - \mu) \quad (4.2)$$

caused by

$$\lambda_i \to -\tfrac{1}{2}i\beta_i + iK_l, \quad \eta \to \tfrac{1}{2}(\pi-\mu),$$
$$K_l = \int_0^{\pi/2}\frac{d\varphi}{(1+l^2\sin^2\varphi)^{1/2}}, \quad K_l \underset{l'\to 0}{\to} \ln(4/l') \to \infty. \quad (4.3)$$

The following limits were investigated in Ref. 3:

$$\ln\left[\frac{h(\lambda+\eta)}{h(\lambda-\eta)}\right] \underset{l'\to 0}{\to} -i(\pi-\mu) - \tfrac{1}{16}il'^2\sin\mu\sinh\beta_i,$$
$$\ln\left[\frac{h(\lambda_i-\lambda_j+2\eta)}{h(\lambda_i-\lambda_j-2\mu)}\right] \underset{l'\to 0}{\to} \ln\left[\frac{\sinh(\frac{1}{2}(\beta_i-\beta_j)-2i\mu)}{\sinh(\frac{1}{2}(\beta_i-\beta_j+2i\mu)}\right]. \quad (4.4)$$

Consequently, we arrive at the following form of the periodicity condition,

$$\exp\{i[4\pi\Theta - (\pi-\mu)N]\}\cdot\exp(-ik_iL)$$
$$= \prod_{\substack{k=1\\k\neq i}}^{n}\frac{\sinh[\frac{1}{2}(\beta_i-\beta_j-2i\mu)]}{\sinh[\frac{1}{2}(\beta_i-\beta_j+2i\mu)]}$$
$$k_i = m_0\sinh\beta_i, \quad m_0 = \tfrac{1}{16}l'^2/\delta\sin\mu, \quad N = L/\delta, \quad (4.5)$$

and, upon the additional demands

$$\Theta = (\pi-\mu)N/4\pi, \quad \beta_i = 2v_i, \quad \mu \to -\mu, \quad (4.6)$$

we get finally

$$\exp(ik_iL) = \prod_{\substack{k=1\\k\neq l}}^{n}\frac{\sinh(v_i-v_j+i\mu)}{\sinh(v_i-v_j-i\mu)}, \quad i = 1,2,...,n = N/2. \quad (4.7)$$

On the other hand, by recalling (3.9) and (3.10), we realize that the original sine–Gordon periodicity condition reads

$$\exp\{[a(\lambda_K) - \overline{a(\lambda_k)}]N\} = \exp[\tfrac{1}{4}m^2\delta\sin\gamma\sinh 2v_k\cdot L]$$
$$= \prod_{\substack{j=1\\j\neq k}}^{n}\frac{\sinh(v_k-v_j+i\gamma)}{\sinh(v_k-v_j-i\gamma)}, \quad \gamma\to\mu, \quad (4.8)$$

i.e., upon an identification

$$\frac{l'^2}{16\delta} = \frac{m^2\delta}{4} \quad (4.9)$$

we can identify (3.34) with the sine–Gordon periodicity condition.

(B) We are interested in the $l' \to 0$ limit, upon the replacements (4.2), (4.6), and (4.9) of all the basic formulas for the *xyz* Heisenberg model. For this purpose let us make use of Ref. 21 and make transparent the l' dependence of the Jacobi functions. Namely we have

$$H(u,k) = \theta_1(u/2K_k,q), \quad \Theta(u,k) = \theta_4(u/2K_k,q),$$
$$k^2 + k'^2 = 1, \quad k' = (1-l')/(1+l'), \quad (4.10)$$

where

$$q = \exp(-\pi K_l/K_{l'}), \quad K_l = \int_0^{\pi/2} d\varphi/(1+l^2\sin^2\varphi)^{1/2},$$
$$K_l \underset{l'\to 0}{\to} \ln(4/l') \to \infty, \quad K_{l'} \underset{l'\to 0}{\to} \pi/2, \quad (4.11)$$
$$q \underset{l'\to 0}{\to} \exp(-2K_l) \to l'^2/16,$$

and the following q expansions hold true[21]:

$$\theta_1(v,q) = 2q^{1/4}(\sin\pi v - q^2\sin 3\pi v + q^6\sin 5\pi v - \cdots)$$
$$= i\sum_{n=-\infty}^{+\infty}(-1)^n q^{(n-1/2)^2}e^{(2n-1)\pi vi}, \quad (4.12)$$

$$\theta_4(v,q) = 1 - 2(q\cos 2\pi v - q^4\cos 4\pi v + q^5\cos 6\pi v - \cdots)$$
$$= \sum_{n=-\infty}^{+\infty} (-1)^n q^{n^2} e^{2n\pi i v}. \tag{4.13}$$

One should also realize that the elliptic modulus k is a function of q and for small l' reads

$$k^2 \cong 16q\,\frac{1+4q^2}{1+8q+24q^2}. \tag{4.14}$$

(C) Let us analyze the $l'\to 0$ behavior of the xyz model R matrix (3.19) upon replacements (4.2), (4.6), (4.9) and provided we demand that $\eta = 0$ is mapped into π. Then

$$\Theta(2\eta)\to 1, \quad \Theta(\lambda-\lambda')\to 1, \quad \Theta(\lambda-\lambda'+2\eta)\to 1, \tag{4.15}$$

but

$$H(\lambda-\lambda')\to 2q^{1/4}\sin[\tfrac{1}{2}i(\beta'-\beta)+\pi]$$
$$= 2q^{1/4}i\sinh[\tfrac{1}{2}(\beta-\beta')],$$
$$H(2\eta)\to 2q^{1/4}i\sin(\pi-\mu) = 2q^{1/4}\sin\mu, \tag{4.16}$$
$$H(\lambda-\lambda'+2\eta)\to 2q^{1/4}i\sinh[\tfrac{1}{2}(\beta-\beta')-i\mu].$$

Consequently,

$$a(\lambda,\lambda') = \Theta(2\eta)\Theta(\lambda-\lambda')H$$
$$\times(\lambda-\lambda'+2\eta)\to 2q^{1/4}i\sinh[\tfrac{1}{2}(\beta-\beta')-i\mu],$$
$$b(\lambda,\lambda') = H(2\eta)\Theta(\lambda-\lambda')$$
$$\times\Theta(\lambda-\lambda'+2\eta)\to 2q^{1/4}\sin\mu, \tag{4.17}$$
$$c(\lambda,\lambda') = \Theta(2\eta)H(\lambda-\lambda')$$
$$\times\Theta(\lambda-\lambda'+2\eta)\to 2q^{1/4}i\sinh[\tfrac{1}{2}(\beta-\beta')],$$
$$d(\lambda,\lambda') = H(2\eta)H(\lambda-\lambda')H(\lambda-\lambda'+2\eta)$$
$$\to -8q^{3/4}i\sin\mu\sinh[\tfrac{1}{2}(\beta-\beta')]\sinh\{\tfrac{1}{2}(\beta-\beta')-i\mu\}.$$

Now let us make a reflection $\mu\to-\mu$ in the above formulas [compare, e.g., also (4.6)] and divide all of them by $a(\lambda,\lambda',\mu\to-\mu)$. We get

$$a(\lambda,\lambda')\to 1,$$
$$b(\lambda,\lambda')\to \frac{i\sin\mu}{\sinh[\tfrac{1}{2}(\beta-\beta')+i\mu]}, \tag{4.18}$$
$$c(\lambda,\lambda')\to \frac{\sin[\tfrac{1}{2}(\beta-\beta')]}{\sinh[\tfrac{1}{2}(\beta-\beta')+i\mu]},$$
$$d(\lambda,\lambda')\to 4q^{1/2}\sin\mu\sinh[\tfrac{1}{2}(\beta-\beta')].$$

Now, it suffices to recall (4.6) with $\beta = 2v$ to noticing that, with an accuracy up to $q^{1/2}$ corrections, the following approximate relationship holds true for the R matrices of the xyz and sine–Gordon models:

$$R_{xyz}(\lambda,\lambda')\frac{1}{\alpha(\lambda,\lambda')} \cong R_{\mathrm{sG}}(\lambda,\lambda'), \tag{4.19}$$

where the right-hand side is l' independent. Though the approximation improves with $l'\to 0$, we cannot here make the limit $l' = 0$, like

$$\lim_{l'\to 0} R_{xyz}\frac{1}{\alpha(\lambda,\lambda')} = R_{\mathrm{sG}}(\lambda,\lambda'). \tag{4.20}$$

Namely, because of (4.3)–(4.7), the mass factor $m_0 = \frac{1}{16}(l'^2/\delta)\sin\mu$ appearing there as well as the identification $l'^2/16\delta = m^2\delta/4$ of (4.9) depends on l'. A nontrivial limit arises only if we simultaneously let $\delta\to 0$, so that $l'^2/\delta\to$const, i.e., if with $l'\to 0$, we recover the continuum limit of the lattice theory. But at the same time $m^2\delta$ should be kept constant, which means that we are forced to let m go to ∞ with $\delta\to 0$. The need for such an effect was explicitly recovered in Ref. 5; then only the spin-$\frac{1}{2}$ approximation of the sine-Gordon model in terms of the xyz model becomes reliable.

(D) The limiting properties of the transfer matrix eigenvalue are not straightforward if we use the form (3.32) of Λ. However,

$$e^{2\pi i\Theta}h^N(x-\eta)\Lambda_{xyz}(\Theta,\lambda,\lambda_1,\cdots,\lambda_n) = \Lambda'(\Theta,\lambda,\lambda_1,\cdots,\lambda_n)$$
$$= e^{4\pi i\Theta}\left[\frac{h(\lambda+\eta)}{h(\lambda-\eta)}\right]^N\prod_{k=1}^n\frac{h(\lambda-\lambda_k-2\eta)}{h(\lambda-\lambda_k)}$$
$$+\prod_{k=1}^n\frac{h(\lambda-\lambda_k+2\eta)}{h(\lambda-\lambda_k)} \tag{4.21}$$

exhibits quite reasonable limiting behavior: By virtue of (4.3)–(4.7) we arrive at

$$\Lambda' \underset{l'\to 0}{\to} \exp(ikL)\prod_{k=1}^n\frac{\sinh(\lambda-\lambda_k+i\mu)}{\sinh(\lambda-\lambda_k)}$$
$$+\prod_{k=1}^n\frac{\sinh(\lambda-\lambda_k+i\mu)}{\sinh(\lambda-\lambda_k)} \tag{4.22}$$

so that

$$e^{2\pi i\Theta}h^N(\lambda-\eta)\Lambda_{xyz}(\Theta,\lambda,\lambda_1,\ldots,\lambda_n)$$
$$\underset{l'\to 0}{\to} \exp[-d(\lambda)N]\cdot\Lambda_{\mathrm{sG}}(\lambda,\lambda_1,\ldots,\lambda_n). \tag{4.23}$$

In the above we can again approach $l' = 0$ while letting δ go to 0, (4.23) recovers a resolution of the eigenvalue problem $A(\lambda)\Omega = e^{a(\lambda)}\Omega$, $D(\lambda)\Omega = e^{d(\lambda)}\Omega$ in the $\{l'\to 0, \delta\to 0\}$ limit. Because tha data $\{R, a(\lambda), d(\lambda)\}$ are defining objects for the model, we find that indeed the simultaneous $\{l'\to 0, \delta\to 0\}$ limit of the xyz Heisenberg model defined in $P\mathscr{H}\subset\mathscr{H}$ gives rise to the (continuous) sine–Gordon model, in full agreement with Ref. 1. However, let us recall that the xyz model eigenvectors were explicitly constructed in terms of $\sigma_n^+ = Pa_n^*P$, $\sigma_n^- = Pa_nP$. The limiting procedure does not destroy the Pauli exclusion principle which is coded in the property $(\sigma_n^+)^2 = 0 = (\sigma_n^-)^2$ $\forall n$, and because of the finite volume (L is kept fixed) we have not destroyed the Fock-ness property

$$\sigma_n^-\Omega_0 = a_k\Omega_0 = 0 \quad \forall k, \quad \Omega_0 = \prod_{k=1}^N{}^{\otimes} e_k^-$$

[compare, e.g., (3.22)]. Consequently, in the continuum limit we still remain in the proper subspace $P\mathscr{H}$ of the Fock space, where P is a continuum generalization of the former discrete projection, in complete analogy with the case of the nonlinear Schrödinger model.

On the other hand, the representation (3.10) of the fundamental sine–Gordon algebra satisfies its continuum limit as well, being however defined in the whole of $\mathscr{H}$ through bounded in $\mathscr{H}$ Weyl operators. It means that in the continuum we can represent the fundamental sine–Gordon structure $\{R, a(\lambda), d(\lambda)\}$ in at least two equivalent ways: first in all of $\mathscr{H}$ and second in the proper subspace $P\mathscr{H}$ of $\mathscr{H}$. In the language of representation theory, it means that the contin-

uum limit of the fundamental representation (3.10) is reducible and has at least one nontrivial component, namely this on $P\mathscr{H}$. Hence though the spectral problem resolved in Ref. 1 gives an exact spectrum of the sine–Gordon model, it does not suffice to recover all of its eigenfunctions. We conjecture that the WKB/massive Thirring spectrum of the sine–Gordon system is exact and infinitely degenerate.

(E) The existence of a representation of the sine–Gordon algebra in $P\mathscr{H}$ implies the existence of a representation in the continuum limit of any among the proper subspaces of $\mathscr{H}$ received by composing projections on two arbitrary (neighboring) excitation levels at each single site. In this connection, let us notice that single-site operators,[22]

$$\sigma^+ = a^* \frac{\cos^2(\pi a^* a/2)}{(a^* a+1)^{1/2}}, \quad \sigma^- = \frac{\cos^2(\pi a^* a/2)}{(a^* a+1)^{1/2}} a \tag{4.24}$$

generate in $h = \mathscr{L}^2(R^1)$ a reducible representation of the CAR algebra:

$$\begin{aligned} &[\sigma^-,\sigma^+]_+ = 1 = \sum_n e_n \otimes e_n, \quad (e_n,e_k) = \delta_{nk}, \\ &\sigma^- e_{2n} = 0, \quad \sigma^+ e_{2n} = e_{2n+1} \quad \forall\ n = 0,1,\cdots, \\ &(\sigma^-)^2 = 0 = (\sigma^+)^2 \end{aligned} \tag{4.25}$$

provided $\{e_n\}$, is a complete basis system in $\mathscr{L}^2(R^1)$. For the sine–Gordon model $\mathscr{L}^2(R^1)$ is the Hilbert space of the quantum pendulum.[6] The representation (4.24) becomes reduced on each two-dimensional subspace of $h = \oplus_{n=0,1,\cdots} h(n)$, $h(n) = \mathscr{L}(e_{2n},e_{2n+1})$. Our construction allows to identify a representation of (4.24) related to the single-site projection $p(0) = :\exp(-a^*a): + a^*:\exp(-a^*a):a$, i.e., $P = P(0) = \Pi_{i=1}^N p_j(0)$. We can as well use a projection $P = P(k)$ with $p(k)h(k) = h(k)$ and the spin-$\frac{1}{2}$ algebra $\{P(k)\sigma_j^{\pm}P(k) = \sigma_j^{\pm}(k)\}_{k=0,1,\cdots}^{j=1,2,\cdots,N}$. In case of the nonlinear Schrödinger model we were able to give a detailed description of how fermions arise in the Bose system. In the present case, things are less straightforward, but once any operator quantity is given in terms of $\{\sigma_j^{\pm}\}$ then the Jordan Wigner transformation allows us to rewrite it in terms of pure fermionic variables. Let us exploit a reducible representation of the CAR given in Ref. 22:

$$\begin{aligned} &b_i = \exp\left(i\pi \sum_{k=1}^{i-1} N_k\right)\cdot(N_k+1)^{-1/2}\cos^2(\pi N_i/2)\cdot a_i, \\ &b_i^* = a_i^*(N_i+1)^{-1/2}\cos^2(\pi N_i/2)\exp\left[-i\pi \sum_{k=1}^{i-1} N_k\right], \\ &N_k = a_k^* a_k, \end{aligned} \tag{4.26}$$

where

$$[b_i,b_j^*]_+ = \delta_{ij}, \quad [b_i,b_j]_+ = 0. \tag{4.27}$$

The representation follows from the previously introduced (4.24), $\sigma^{\pm} \to \sigma_j^{\pm}$, via

$$\begin{aligned} &b_j = \left(\prod_{k=1}^{j-1} c_k\right)\sigma_j^-, \quad \sigma_j^* = \left(\prod_{k=1}^{j-1} c_k^*\right)\sigma_j^+, \\ &c_k = \exp(i\pi N_k) \Rightarrow [c_k,\sigma_j^{\pm}]_+ = 0, \quad c_k^2 = 1. \end{aligned} \tag{4.28}$$

Here $[N_k,P(j)]_- = 0 \forall k,j$; hence we can immediately reduce the representation of the CAR (4.26) according to $b_j \to b_j(k) = P(k)b_jP(k)$. In particular, due to $P(0) = \Pi_{j=1}^N P_j(0)$, we arrive at the well-known Jordan Wigner formulas:

$$\begin{aligned} &P(0)b_iP(0) = b_i(0) = \exp\left(i\pi \sum_{k=1}^{i=1} \sigma_k^+\sigma_k^-\right)\cdot\sigma_i^-, \\ &P(0)b_i^*P(0) = b_i^*(0), \end{aligned} \tag{4.29}$$

which are easily invertible. For the derivation of (4.29) one should notice that $(N_i+1)^{-1/2}\cos^2(\pi N_i/2) = 1$ on the allowed domain. If we supply (4.26) with

$$\tilde{b}_i^* = \left(\prod_{k=1}^{i-1} c_k^*\right)\sigma_i^-, \quad \tilde{b}_i = \left(\prod_{k=1}^{i-1} c_k\right)\sigma_i^+, \tag{4.30}$$

we arrive at the two-component Fermi system (the massive Thirring model demands it):

$$\begin{aligned} &[b_j(k),b_i^*(l)]_+ = \delta_{ji}\delta_{kl}, \quad [b_j(k),b_i(l)]_+ = 0, \\ &b_j(1) = b_j, \quad b_j(2) = \tilde{b}_j. \end{aligned} \tag{4.31}$$

ACKNOWLEDGMENT

A useful conversation with Professor J. Honerkamp is gratefully acknowledged.

[1] A Luther, Phys. Rev. B **14**, 2153 (1976).
[2] M. Luscher, Nucl. Phys. B **17**, 475 (1976).
[3] H. Bergknoff and H. B. Thacker, Phys. Rev. D **19**, 3666 (1979).
[4] S. Coleman, Phys. Rev. D **11**, 2088 (1975).
[5] P. Garbaczewski, J. Math. Phys. **22**, 574 (1981).
[6] P. Garbaczewski, J. Math. Phys. **21**, 2572 (1980).
[7] P. Garbaczewski, Commun. Math. Phys. **43**, 131 (1975).
[8] T. D. Schultz, J. Math. Phys. **4**, 666 (1963).
[9] M. Girardeau, J. Math. Phys. **1**, 516 (1960).
[10] E. H. Lieb and W. Liniger, Phys. Rev. **130**, 1605 (1963).
[11] E. H. Lieb, Phys. Rev. **130**, 1616 (1963).
[12] H. B. Thacker and D. Wilkinson, Phys. Rev. D **19**, 3660 (1979).
[13] H. Grosse, Phys. Lett. B **86**, 267 (1979).
[14] P. Garbaczewski and J. Rzewuski, Rep. Math. Phys. **6**, 431 (1974).
[15] D. B. Creamer, H. B. Thacker, and D. Wilkinson, Phys. Rev. D **21**, 1523 (1980).
[16] C. N. Yang and C. P. Yang, J. Math. Phys. **10**, 1115 (1969).
[17] P. Garbaczewski, Phys. Rep. C **36**, 65 (1978).
[18] J. Dobaczewski, Nucl. Phys. A **369**, 219 (1981).
[19] E. K. Sklyanin, L. A. Tahtajan, and L. D. Faddeev, Teor. Mat. Fiz. **40**, 194 (1979).
[20] L. A. Tahtajan and L. D. Faddeev, Usp. Mat. Nauk **34** (5), 13 (1979).
[21] E. Janke, F. Emde, and F. Lösch, *Tafeln höhere funktionen* (Teubner, Stuttgart, 1960).
[22] S. Naka, Progr. Theor. Phys. **59**, 2107 (1978).

Complex, self-dual Euclidean SU(2) gauge fields with finite real actions

A. Chakrabarti, A. Comtet,[a] and F. Koukiou
Groupe de Recherche du C.N.R.S. n° 48, Centre de Physique Théorique de l'Ecole Polytechnique, Plateau de Palaiseau, 91128 Palaiseau, Cedex, France

(Received 14 October 1981; accepted for publication 26 March 1982)

A class of complex, self-dual Euclidean SU(2) gauge field configurations are obtained. The action is finite and real. In a simple scaling limit, of infinite action, this class reduces to the complex, static monopoles of Manton. Various special features of these solutions are discussed.

PACS numbers: 11.15. – q

I. INTRODUCTION

In a previous paper[1] a conjugate pair of non-self-dual complex solutions for SU(2) gauge fields was presented. They had finite, complex Euclidean actions and zero topological charge. One motivation for constructing complex solutions is their possible role in systematic applications of saddle point methods.[2] One may conjecture that in trying to extract certain results by semiclassical methods it would be essential to supplement the real saddle points, corresponding to instantons, by complex ones. Only after having taken into account the effects not only of real but also of suitable classes of *complex* classical solutions should one decide whether the exploration of certain domains of interest is beyond the scope of such methods or not. For this purpose it is evidently important to construct self-dual complex solutions with finite real actions, providing the weight factor e^{-S} for the corresponding paths and thus most directly generalizing the class of real instanton solutions. We present a class of such solutions in this paper. Let us at this point make it quite explicit what type of complexification we are aiming at.

(1) We are not complexifying the coordinates x_μ of the flat Euclidean space. They will have their usual real domains. Our solutions must be complex with the real Euclidean space as base. Complexification of the base space can be of interest in a still broader context and will automatically lead to complex fields. But in this paper we will present explicit examples of SU(2) fields which remain complex in spite of the above-mentioned restriction. So in this context we cannot start with real gauge potentials and complexify them through complex coordinate transformations. We will have to find some less evident method.

(2) We will also exclude solutions obtained by formal *complex* gauge transformations of real solutions. Our solutions must be, in this sense, intrinsically complex. This will be demonstrated by explicitly constructing gauge invariant expressions which are complex for our class of solutions. Our class will, however, contain interesting particular cases which can indeed be rendered real through complex gauge transformations. But we will separate them out, providing explicitly the gauge transformations in question. This will serve to pinpoint the level at which our class of solutions becomes intrinsically complex. See Sec. 4 for examples of complex gauge transformation [SU(2)→SL(2,c), as discussed in Ref. 1].

(3) Neither will we resort to complexification of parameters of known real solutions. In fact such a naive approach usually fails. To take the simplest example, if in the generating function in the 't Hooft gauge of the one instanton solution, namely in

$$\rho = 1 + \lambda^2/(x - \xi)^2 \tag{1.1}$$

the center ξ_μ is complexified, the action diverges. The coordinates x_μ being defined to be real, the imaginary part of ξ_μ cannot be absorbed by a translation of the origin. The situation concerning the $(8n - 3)$ parameters of the general solutions, for arbitrary index n, is naturally more involved, which we will not attempt to analyze here. In sharp contrast to the foregoing line of thought we will present solutions whose action will diverge if a certain purely imaginary parameter is made real. The solutions must be complex for the action to be finite. They will be intrinsically complex also in this sense. The imaginary parameter will be found to enter in a nonperturbative way (in Sec. 2, $K = i\zeta$, where $\zeta > 1$ for the action to be finite). The general solution cannot hence be continuously deformed to a real one with finite action.

(4) We will construct our solutions such that though (except for particular cases to be pointed out) they are intrinsically complex according to the above-mentioned criteria, *the corresponding actions will be real.* The situation will be shown to be as follows. Gauge invariant expressions such as

$$\mathrm{Tr}(F_{\mu\nu} F_{\alpha\beta}) = \mathrm{Tr}(UF_{\mu\nu} U^{-1} UF_{\alpha\beta} U^{-1}), \tag{1.2}$$

where U is any gauge transformation, real or complex, will be in general complex. These are not invariant under coordinate transformations. But since, as explained before, we permit only real coordinate transformations, the set of such tensor components cannot be rendered real by that means. Evidently the inverse real transformation cannot make a real tensor complex. Hence tensors such as (1.2) exhibit the intrinsically complex nature of our solutions. But as the sum over indices is performed to form the action density $\mathrm{Tr}\, F_{\mu\nu} F^{\mu\nu}$ the imaginary parts cancel out. Indeed, our solutions satisfy a deeper constraint. The study of the third and the fourth order invariants in Appendix A would seem to indicate that all polynomials in $F_{\mu\nu}$, *invariant under both gauge and coordinate transformations*, are real for our class of solutions (1.3). This, in spite of the situation described in the remarks following (1.2).

For all physical consequences only the quantities invariant under both gauge and coordinate transformations are relevant.

Hence, in this respect (presumably, at least so far as

[a] Division de Physique Théorique, Institut de Physique Nucléaire, Université Paris XI, 91406 Orsay Cedex, France.

 0022-2488/83/030659-15$02.50

such invariant polynomials are concerned) our solutions are on the same footing as real, self-dual ones. Yet they provide minima of action which would have been missed had one looked only for real solutions. Such properties should make a thorough study of the possible roles of such solutions worthwhile.

In the following section we will introduce an *ansatz* inspired by the work of Manton on complex monopoles[3] and our previous results exploiting static solutions in De Sitter space.[1,4] The approach using coordinates which are familiar in the study of De Sitter space and make its static, spherical symmetry explicit has been commented upon in detail in our previous papers. We will not repeat such comments here. The following sections will again illustrate how this technique enables one to select classes of solutions which combine interesting properties with a relative simplicity permitting explicit calculations.

In the Introduction we have mostly tried to explain the type of complexification we are aiming at. Other important properties of our constructions are discussed in Sec. 5 and 6. In particular we show that a new possibility that arises for complex solutions (as opposed to real ones), namely the possibility of obtaining a finite action even for divergent field strengths, is realized in a particularly simple fashion in our class of solutions.

II. *ANSATZ* AND SOLUTIONS

We consider the flat Euclidean space with the line element

$$ds^2 = dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{2.1}$$

where

$$-\infty < t < \infty, \quad 0 \leqslant r < \infty, \quad 0 \leqslant \theta \leqslant \pi, \quad 0 \leqslant \phi < 2\pi.$$

We introduce the coordinates τ and ρ given by

$$(t + ir) = \tan\left(\frac{\tau + i\rho}{2}\right). \tag{2.2}$$

The full domains of t and r in (2.1) now correspond to the domains

$$0 \leqslant \tau < 2\pi, \quad 0 \leqslant \rho < \infty. \tag{2.3}$$

These coordinates, well known in the study of De Sitter space,[1,4] will be used here in the context of flat space to achieve a simple form (independent of τ) of the gauge potentials. The finite domain of τ leaves open the possibility of "static" solutions (involving ρ, θ, and ϕ) having finite actions. The solutions will be time dependent when expressed in terms of t, r, θ, and ϕ, which will do ultimately. Since only t will be considered as the Euclidean time, our solutions will be found to be neither really static nor periodic in time. These solutions can, of course, be considered (due to the well-known conformal properties of gauge fields) in the context of conformally flat spaces of nonzero curvature. But in this paper we will be interested in flat Euclidean space (2.1) using the transformation (2.2) as a trick to simplify intermediate steps in calculations.

We now proceed to give the *ansatz* for the components A_τ, A_ρ, A_θ, and A_ϕ.

Let $f(\rho,\theta,\phi)$ be a function independent of τ. Let K be a constant parameter which may be *complex*. (In fact our finite action solutions will correspond to $K = i\zeta$, where ζ is real). Let, in terms of the Pauli matrices σ_1, σ_2, σ_3,

$$\begin{aligned}
\sigma_\rho &= \sin\theta\cos\phi\,\sigma_1 + \sin\theta\sin\phi\,\sigma_2 + \cos\theta\,\sigma_3,\\
\sigma_\theta &= \cos\theta\cos\phi\,\sigma_1 + \cos\theta\sin\phi\,\sigma_2 - \sin\theta\,\sigma_3,\\
\sigma_\phi &= -\sin\phi\,\sigma_1 + \cos\phi\,\sigma_2.
\end{aligned} \tag{2.4}$$

We set (with $f_\rho \equiv \partial_\rho f$ and so on)

$$\begin{aligned}
A_\tau &= f_\rho\frac{\sigma_\rho}{2} + \frac{1}{\sinh\rho} f_\theta \frac{\sigma_\theta}{2} + \frac{1}{\sinh\rho\sin\theta} f_\phi \frac{\sigma_\phi}{2},\\
A_\rho &= K\frac{\sigma_\rho}{2} + \frac{1}{\sinh\rho\sin\theta} f_\phi \frac{\sigma_\theta}{2} - \frac{1}{\sinh\rho} f_\theta \frac{\sigma_\phi}{2},\\
\frac{A_\theta}{\sinh\rho} &= -\frac{1}{\sinh\rho\sin\theta} f_\phi \frac{\sigma_\rho}{2} + K\frac{\sigma_\theta}{2} + \left(f_\rho - \frac{\cosh\rho - 1}{\sinh\rho}\right)\frac{\sigma_\phi}{2},\\
\frac{A_\phi}{\sinh\rho\sin\theta} &= \frac{1}{\sinh\rho} f_\theta \frac{\sigma_\rho}{2} - \left(f_\rho - \frac{\cosh\rho - 1}{\sinh\rho}\right)\frac{\sigma_\theta}{2} + K\frac{\sigma_\phi}{2}.
\end{aligned} \tag{2.5}$$

The corresponding $F_{\mu\nu}$ are given in Appendix A.

For simplicity we have normalized a possible scaling factor to one. This can be introduced as follows:

$$\begin{aligned}
&\tau \to \lambda\tau', \quad \rho \to \lambda\rho',\\
&A_\tau \to \frac{1}{\lambda} A'_{\tau'}, \quad A_\rho \to \frac{1}{\lambda} A'_{\rho'}.
\end{aligned} \tag{2.6}$$

Introducing also $K \to (1/\lambda)\,K'$ and making $\lambda \to 0$, (2.7)

one obtains (with $\sigma_\rho = \sigma_{\rho'}$, f' being the limiting form of f, and so on)

$$\begin{aligned}
A'_{\tau'} &= f'_{\rho'}\frac{\sigma_{\rho'}}{2} + \frac{1}{\rho'} f'_\theta \frac{\sigma_\theta}{2} + \frac{1}{\rho'\sin\phi} f'_\phi \frac{\sigma_\phi}{2},\\
A'_{\rho'} &= K'\frac{\sigma_{\rho'}}{2} + \frac{1}{\rho'\sin\phi} f'_\phi \frac{\sigma_\theta}{2} - \frac{1}{\rho'} f'_\theta \frac{\sigma_\phi}{2},\\
\frac{A'_\theta}{\rho'} &= -\frac{1}{\rho'\sin\theta} f'_\phi \frac{\sigma_{\rho'}}{2} + K'\frac{\sigma_\theta}{2} + f'_{\rho'}\frac{\sigma_\phi}{2},\\
\frac{A'_\phi}{\rho'\sin\theta} &= \frac{1}{\rho'} f'_\theta \frac{\sigma_{\rho'}}{2} - f'_{\rho'}\frac{\sigma_\theta}{2} + K'\frac{\sigma_\phi}{2}.
\end{aligned} \tag{2.8}$$

We note that correspondingly the line element (2.1) which is in terms of τ and ρ,

$$ds^2 = \frac{1}{(\cos\tau + \cosh\rho)^2} \times [d\tau^2 + d\rho^2 + \sinh^2\rho(d\theta^2 + \sin^2\theta d\phi^2)], \quad (2.9)$$

reduces in the limit $\lambda \to 0$ to

$$\frac{4ds^2}{\lambda^2} = ds'^2 = d\tau'^2 + d\rho'^2 + \rho'^2(d\theta^2 + \sin^2\theta\, d\phi^2). \quad (2.10)$$

In the flat metric given by ds'^2, (2.8) can be verified to be the *ansatz* of Manton,[3] leading to complex monopoles, which we recover (adapted to our notations and conventions) as a limiting case.[5] It will be noted that the terms with the factor $(\cosh\rho - 1)/\sinh\rho = \tanh(\rho/2)$ in (2.5) disappear in (2.8). *These supplementary terms turn out to be essential to assure self-duality.* Hence a naive generalization of (2.8) by replacing ρ' by $\sinh\rho$ in the explicit factors is not adequate. Thanks to the conformal properties of gauge fields (which permit us in formal construction of solutions to ignore an overall conformal factor in the line element) an *ansatz* in terms of τ and ρ can lead one to solutions in terms of standard flat space coordinates in two distinct ways, namely by a coordinate transformation such as (2.2) and through a limiting process such as (2.6) and (2.7). This permits us to construct sequences of finite action solutions which have as limits interesting static finite energy solutions (but naturally of infinite action due to the trivial time integration over $-\infty < t < \infty$ or equivalently over τ'). Examples can be found in Refs. 4 and in the present paper. For the time being we just note that for finite $K' = \lambda K$, $K \to \infty$ as $\lambda \to 0$. This will be found related to infinite action in the limit.

Let us now study the self-duality constraints for (2.5). Defining

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + i[A_\mu, A_\nu] \quad (\mu,\nu = \tau,\rho,\theta,\phi), \quad (2.11)$$

these constraints are given by[6]

$$F_{\tau\rho} = \frac{1}{\sinh^2\rho \sin\theta} F_{\theta\phi},$$
$$F_{\tau\theta} = \frac{1}{\sin\theta} F_{\phi\rho}, \quad (2.12)$$
$$\frac{F_{\tau\phi}}{\sin\theta} = F_{\rho\theta}.$$

For the *ansatz* (2.5), these can be shown to reduce to the single equation [with $f_{\rho\rho} = (\partial^2/\partial\rho^2)f$, and so on]

$$(f_{\rho\rho} + 2\coth\rho f_\rho - f_\rho^2) + \frac{1}{\sinh^2\rho}(f_{\theta\theta} + \cot\theta f_\theta - f_\theta^2) + \frac{1}{\sinh^2\rho \sin^2\theta}(f_{\phi\phi} - f_\phi^2) = (K^2 + 1). \quad (2.13)$$

Setting

$$f = -\ln H, \quad (2.14)$$

one obtains

$$\Delta H \equiv \frac{1}{\sinh^2\rho}\partial_\rho(\sinh^2\rho\,\partial_\rho H) + \frac{1}{\sinh^2\rho\sin\theta}\partial_\theta(\sin\theta\,\partial_\theta H) + \frac{1}{\sinh^2\rho\sin^2\theta}\partial_\phi^2 H = -(K^2+1)H \quad (2.15)$$

[henceforth the operator Δ will be defined by (2.15)]. Real values of K do not lead to a finite action.

As in Refs. 3 and 5, solutions with interesting properties are obtained for *imaginary* K, namely

$$K = i\zeta, \quad (2.16)$$

where ζ is real. One can assume ζ to be positive without real loss of generality and thus avoid introducing $|\zeta|$ in many subsequent results. In fact, interesting properties will be found to be obtained for *integer values* of $\zeta > 1$ ($\zeta = 2,3,\ldots$). *Henceforth, unless otherwise stated, we will impose this restriction.* (We will come back to the question of nonintegral ζ later on in the concluding remarks of Sec. 6). We have reduced our problem to that of solving a *linear* equation, as one is able to do for 't Hooft gauge instantons.

The equation

$$\Delta H = (\zeta^2 - 1)H, \quad (2.17)$$

has solutions expressible in terms of Legendre functions as

$$H = \frac{1}{(\sinh\rho)^{1/2}} \sum_{j,k} c_{jk} \times P_{\zeta-1/2}^{-(k+1/2)}(\cosh\rho)\, P_k(\cos\psi_j), \quad (2.18)$$

where

$$\cos\psi_j = \cos\theta_j\cos\theta + \sin\theta_j\sin\theta\cos(\phi - \phi_j) \quad (2.19)$$

and c_{jk}, θ_j, and ϕ_j are arbitrary parameters. Equation (2.17) was also encountered in Ref. 7. There, however, *real* linear deformations of static spherically symmetric solutions were being considered. Here we obtain the same equation for *imaginary* K.

Having exhibited the general solution (2.18) let us now concentrate on one particular class of finite action solutions. In Ref. 3, the solutions of (2.8) considered were for the limiting equation

$$\Delta' H' \equiv \frac{1}{\rho'^2}\partial_{\rho'}(\rho'^2\partial_{\rho'} H') + \frac{1}{\rho'^2\sin\theta}\partial_\theta(\sin\theta\,\partial_\theta H') + \frac{1}{\rho'^2\sin^2\theta}\partial_\phi^2 H' = \zeta'^2 H', \quad (2.20)$$

$$H' = \sum_j a'_j \frac{\sinh\zeta'\rho'_j}{\rho'_j}, \quad (2.21)$$

where

$$\rho_j'^2 = \rho'^2 + c_j'^2 - 2\rho' c'_j\cos\psi_j. \quad (2.22)$$

Here the terms in the linear superposition (2.21) are obtained by a translation of the origin in one basic solution,

$$H' = a' \frac{\sinh\zeta'\rho'}{\rho'}. \quad (2.23)$$

For this spherically symmetric solution one can easily verify that the evident modification

$$H = a\frac{\sinh\zeta\rho}{\sinh\rho} \tag{2.24}$$

gives indeed the corresponding basic spherically symmetric solution of (2.17).

It remains now to construct an adequate generalization of the translated form (2.22). This is done in Appendix B. For the readers' convenience we have collected there many results concerning "pseudotranslations" suitable for our purpose. To show that they furnish the necessary solutions we start with a "one-center" form

$$H = a\frac{\sinh\zeta\eta}{\sinh\eta}, \tag{2.25}$$

where η is to be determined. Then (2.17) leads to

$$\begin{aligned}\Delta H &= \Delta\eta\frac{dH}{d\eta} + (\vec{\nabla}\eta)^2\frac{d^2H}{d\eta^2}\\ &= (\zeta^2 - 1)H,\end{aligned} \tag{2.26}$$

where

$$(\vec{\nabla}\eta)^2 \equiv \eta_\rho^2 + \frac{1}{\sinh^2\pi}\left(\eta_\theta^2 + \frac{1}{\sin^2\theta}\eta_\phi^2\right) \tag{2.27}$$

and Δ is given by (2.15). This is satisfied if

$$(\vec{\nabla}\eta)^2 = 1 \tag{2.28}$$

and

$$\Delta\eta = 2\coth\eta. \tag{2.29}$$

It can now be verified that η, given by

$$\cosh\eta = \cosh c\cosh\rho - \sinh c\sinh\rho\cos\psi, \tag{2.30}$$

satisfies (2.28) and (2.29) where

$$\cos\psi = \cos\theta'\cos\theta + \sin\theta'\sin\theta\cos(\phi - \phi'), \tag{2.31}$$

and c, θ', ϕ' are constants. The signficance of (2.30) is discussed in detail in Appendix B. For $c = 0$ we get back the spherically symmetric form (2.24). Now, finally, we can take a linear superposition

$$H = \sum_j a_j\frac{\sinh\zeta\eta_j}{\sinh\eta_j}, \tag{2.32}$$

where

$$\begin{aligned}\cosh\eta_j &= \cosh c_j\cosh\rho - \sinh c_j\sinh\rho\cos\psi_j,\\ \cos\psi_j &= \cos\theta_j\cos\theta + \sin\theta_j\sin\theta\cos(\phi - \phi_j),\end{aligned} \tag{2.33}$$

and a_j, c_j, θ_j, ϕ_j are free parameters.

At this point we note a remarkable consequence of the restriction of ζ to *integer values*.

It would seem that our class of solution (even for a fixed integer ζ) can accommodate an infinite number of free parameters (a_j, c_j, etc.). But, in fact, the maximum number of parameters is finite and is determined by ζ. Using an expansion formula for Gegenbauer polynomials[8] it can be shown that for integer ζ,

$$\begin{aligned}\frac{\sinh\zeta\eta}{\sinh\eta} &= C_{\zeta-1}^1(\cosh\eta)\\ &= C_{\zeta-1}^1(\cosh c\cosh\rho - \sinh c\sinh\rho\cos\psi)\\ &= \sum_{k=0}^{\zeta-1}(-1)^k\frac{\pi}{2}\frac{(2k+1)\Gamma(\zeta+k+1)}{(\sinh c\sinh\rho)^{1/2}}\\ &\quad\times P_{\zeta-1/2}^{-(k+1/2)}(\cosh\rho)P_{\zeta-1/2}^{-(k+1/2)}(\cosh c)\,P_k(\cos\psi).\end{aligned} \tag{2.34}$$

Thus we have a finite series with ζ terms. (This is no longer true for noninteger ζ, when one has to deal with an infinite series.) Using this expansion for each term in (2.32) and grouping the coefficient, the number of terms is seen to be effectively finite. First let us consider the case when each $\theta_j = 0$ and hence $\cos\psi_j = \cos\theta$ for all j while the c's are all different. For this case, using (2.34), (2.32) reduces to the form

$$\begin{aligned}H = &\frac{1}{(\sinh\rho)^{1/2}}\sum_{k=0}^{\zeta-1}b_k\\ &\times P_{\zeta-1/2}^{-(k+1/2)}(\cosh\rho)\,P_k(\cosh\theta).\end{aligned} \tag{2.35}$$

The number of parameters b_k are thus at most ζ. Adding more terms in (2.32) simply amounts to altering the b_k's. But the b_k's are arbitrary to start with. Hence the situation remains essentially unaltered.

For θ_j's $\neq 0$ in general, using the expansion[8]

$$\begin{aligned}P_k(\cos\psi_j) = &P_k(\cos\theta)\,P_k(\cos\theta_j)\\ &+ 2\sum_{m=1}^{k}\frac{(k-m)!}{(k+m)!}\\ &\times P_k^m(\cos\theta)\,P_k^m(\cos\theta_j)\cos m(\phi - \phi_j),\end{aligned} \tag{2.36}$$

one sees similarly that *the maximum number of parameters is* $\sum_{k=0}^{\zeta-1}(2k+1) = \zeta^2$. We will come back to this point later on (Sec. 6). It should be remembered that since only logarithmic derivatives of H appear in A_μ, one overall factor can be normalized in H. This reduces the effective number of parameters by one. In the limit (2.21) of (2.32) the maximum number of parameters is no longer limited. This corresponds to the fact, already noted before, that a finite $\zeta' = \lambda\zeta$ implies $\zeta\to\infty$ as $\lambda\to 0$. A direct expression of this is provided by the fact that in contrast to (2.34) the corresponding expansion of $(\sinh\zeta'_{\rho'}/\rho')$ leads to an infinite series. Thus the infinite action, complex, static monopoles of Manton[3] can depend on an unlimited number of parameters.

We will show in the following section that ζ will determine the aciton to be $S = 8\pi^2(\zeta - 1)$.

So far we have expressed everything in terms of the set (ρ,θ,ϕ) since they permit us to construct finite action solutions independent of the periodic τ. Let us now express our *ansatz* in terms of the coordinates (t,r,θ,ϕ). Using (2.2) one obtains, since

$$\frac{\partial\tau}{\partial t} = \frac{\partial\rho}{\partial r} = (1 + \cosh\rho\cos\tau)$$

and

$$\frac{\partial\tau}{\partial r} = -\frac{\partial\rho}{\partial t} = \sinh\rho\sin\tau,$$

$$A_t = A_\tau(1+\cosh\rho\cos\tau) - A_\rho\sinh\rho\sin\tau, \tag{2.37}$$

$$A_r = A_\rho(1+\cosh\rho\cos\tau) + A_\tau\sinh\rho\sin\tau, \tag{2.38}$$

while A_θ and A_ϕ do not change. Expressing the coefficients in terms of r and t, and with

$$D \equiv [\{t^2+(r+1)^2\}\{t^2+(r-1)^2\}]^{1/2}, \tag{2.39}$$

one has

$$A_t = \frac{2}{D^2}\{(1+t^2-r^2)A_\tau - (2rt)A_\rho\},$$

$$A_r = \frac{2}{D^2}\{(2rt)A_\tau + (1+t^2-r^2)A_\rho\},$$

$$\frac{1}{r}A_\theta = \frac{2}{D}\left(\frac{1}{\sinh\rho}A_\theta\right),$$

$$\frac{A_\phi}{r\sin\theta} = \frac{2}{D}\left(\frac{1}{\sinh\rho\sin\theta}A_\phi\right). \tag{2.40}$$

The terms involving ρ in (2.35) and its ϕ-dependent generalizations can be expressed, using (B37), as

$$\left(\frac{2r}{D}\right)^{-1/2} P^{-(k+1/2)}_{\zeta-1/2}\left(\frac{1+r^2+t^2}{D}\right). \tag{2.41}$$

This provides one way of expressing the right-hand sides of (2.41) entirely in terms of t, r, θ, and ϕ. Calculations in studying the properties of these solutions are of course performed very much more conveniently in terms of the coordinates ρ, θ, and ϕ. Certain apparent singularities in H as $\rho\to\infty$ (when $t\to 0$ and $r\to 1$) do not prevent the action from being finite, at least under certain simple restrictions, since only logarithmic derivatives of H are involved. This we proceed to demonstrate in Sec. 3. More discussions concerning the related properties of the gauge potentials (2.40) will be found in Sec. 5.

III. THE ACTION

One has

$$S = \int_{-\infty}^{+\infty} dt\int_0^\infty dr\int_0^\pi d\theta\int_0^{2\pi} d\phi \times(r^2\sin\theta)(\tfrac{1}{2}\operatorname{Tr}F_{\mu\nu}F^{\mu\nu})\,(\mu,\nu=t,r,\theta,\phi) \tag{3.1}$$

$$= \int_0^{2\pi} d\tau\int_0^\infty d\rho\int_0^\pi d\theta\int_0^{2\pi} d\phi \times\sqrt{g}(\tfrac{1}{2}\operatorname{Tr}F_{\alpha\beta}F^{\alpha\beta})\,(\alpha,\beta=\tau,\rho,\theta,\phi), \tag{3.2}$$

where $\sqrt{g}$ is given by (2.9). The τ integration factors out trivially for our class of *ansatz*. Moreover, using the self-duality constraints and τ independence one has

$$\begin{aligned}\sqrt{g}\,\tfrac{1}{2}\operatorname{Tr}F_{\alpha\beta}F^{\alpha\beta} &= 2\operatorname{Tr}(F_{\tau\rho}F_{\theta\phi}+F_{\tau\theta}F_{\phi\rho}F_{\tau\phi}F_{\rho\theta})\\ &= -2\operatorname{Tr}\left\{\partial_\rho(A_\tau F_{\tau\rho}\sinh^2\rho\sin\theta) + \partial_\theta(A_\tau F_{\tau\theta}\sin\theta) + \partial_\phi\left(A_\tau F_{\tau\phi}\frac{1}{\sin\theta}\right)\right\}\\ &= \tfrac{1}{2}\sinh^2\rho\sin\theta\,\Delta(\vec{\nabla}f)^2\\ &= \tfrac{1}{2}\sinh^2\rho\sin\theta\,\Delta\left(\frac{\vec{\nabla}H}{H}\right)^2.\end{aligned} \tag{3.3}$$

The operators Δ and $\vec{\nabla}$ are defined by (2.15) and (2.27), respectively. As is to be expected for such "static" self-dual solutions, S is expressed in terms of

$$(\vec{\nabla}f)^2 = 2\operatorname{Tr}A_\tau^2. \tag{3.4}$$

Hence, finally,

$$\begin{aligned}S &= \pi\int_0^\infty d\rho\int_0^\pi d\theta\int_0^{2\pi} d\phi \times(\sinh^2\rho\sin\theta)\,\Delta\left(\frac{\vec{\nabla}H}{H}\right)^2\\ &= \pi\int_0^\pi\sin\theta\,d\theta\int_0^{2\pi} \times[\sinh^2\rho\,\partial_\rho\vec{\nabla}(\ln H)^2]_{\lim\rho\to\infty},\end{aligned} \tag{3.5}$$

provided there are no singularities within the surface bounded by the asymptotic value of ρ. For our class of solutions (2.32) H itself has no singularities for finite ρ. Let us now study the asymptotic expansion of η_j and $\ln H$. It can be shown that as $\rho\to\infty$, η given by (2.30) and satisfying (2.28) behaves as follows:

$$\lim_{\rho\to\infty}\eta = \rho+\xi_1+e^{-2\rho}\xi_2+O(e^{-4\rho}), \tag{3.6}$$

where

$$\xi_1 = \ln(\cosh c-\sinh c\cos\psi) \tag{3.7}$$

and

$$\xi_2 = (\partial_\theta\xi_1)^2+\frac{1}{\sin^2\theta}(\partial_\phi\xi_1)^2 \tag{3.8}$$

$$= \frac{\sinh^2 c\sin^2\psi}{(\cosh c-\sinh c\cos\psi)^2}. \tag{3.9}$$

Using (3.6) and developing (2.32) systematically one obtains an analogous form for $\ln H$, namely (since we have $\zeta>1$)

$$\lim_{\rho\to\infty}(\ln H) = (\zeta-1)\rho+h_1(\theta,\phi)+e^{-2\rho}h_2(\theta,\phi)+O(e^{-4\rho}), \tag{3.10}$$

where

$$h_1 = \ln\left[\sum a_j(\cosh c_j-\sinh c_j\cos\psi_j)^{\zeta-1}\right] \tag{3.11}$$

and

$$h_2 = \frac{\sum a_j \{(\zeta - 1)\sinh^2 c_j \sin^2 \psi_j + 1\} (\cosh c_j - \sinh c_j \cos \psi_j)^{(\zeta - 3)}}{\sum a_j (\cosh c_j - \sinh c_j \cos \psi_j)^{\zeta - 1}} \tag{3.12}$$

We will not really need the explicit forms of h_1 and h_2 for calculating S. What is necessary is that they be finite. For a finite number [see the comment following (2.34)] a_j's$\geqslant 0$ this is evident [for example, the argument of the logarithm in (3.11) is positive definite, and so on]. Such a stringent condition is not essential but we will impose it for simplicity. We assume that the choice of parameters is such that such conditions are fulfilled. Now we note that

$$\Delta \ln H = -(\vec{\nabla} \ln H)^2 + \frac{\Delta H}{H}$$
$$= (\zeta^2 - 1) - (\vec{\nabla} \ln H)^2 \tag{3.13}$$

and

$$\Delta \ln H = \left\{2 \coth \rho \ \partial_\rho + \partial_\rho^2 + \frac{1}{\sinh^2 \rho} \Delta_2\right\} (\ln H),$$

where

$$\Delta_2 \equiv \frac{1}{\sin \theta} \partial_\theta (\sin \theta \, \partial_\theta) + \frac{1}{\sin^2 \theta} \partial_\phi^2. \tag{3.14}$$

for $\rho \to \infty$, from (3.10), keeping terms $O(e^{-2\rho})$,

$$\Delta \ln H \equiv \{2(1 + 2e^{-2\rho}) \partial_\rho + \partial_\rho^2 + 4e^{-2\rho} \Delta_2\} (\ln H)$$
$$= 2(1 + 2e^{-2\rho}) [(\zeta - 1)2e^{-2\rho} h_2]$$
$$+ 4e^{-2\rho} h_2 + 4e^{-2\rho} \Delta_2 h_1$$
$$= 2(\zeta - 1) + 4e^{-2\rho} \{(\zeta - 1) + \Delta_2 h_1\}. \tag{3.15}$$

From (3.13) and (3.16)

$$(\vec{\nabla} \ln H)^2 = (\zeta - 1)^2 - 4e^{-2\rho} (\zeta - 1) - 4e^{-2\rho} \Delta_2 h_1. \tag{3.16}$$

The elimination of h_2 is due to the constraint (2.13) [or (3.13)], which implies

$$\Delta_2 h_1 + (\vec{\nabla}_2 h_1)^2 = (\zeta - 1)(h_2 - 1), \tag{3.17}$$

where

$$(\nabla_2 h_1)^2 \equiv (h_{1,\theta})^2 + \frac{1}{\sin^2 \theta} (h_{1,\phi})^2.$$

Substituting this in (3.5),

$$S = 2\pi \int_0^\pi \sin \theta \, d\theta \int_0^{2\pi} d\phi \, [(\zeta - 1) + \Delta_2 h_1]. \tag{3.18}$$

Here h_1 is nonsingular and depends on periodic functions of θ and ϕ. It can be expanded in spherical harmonics. Hence the term $\Delta_2 h_1$ does not contribute to the integral. Thus, finally,

$$S = 8\pi^2 (\zeta - 1) \quad (\zeta > 1). \tag{3.19}$$

(If negative sign of ζ is allowed, $|\zeta|$ replaces ζ in H and hence also in S.) Thus we see that in the integrand of (3.18) a complicated contribution from the angular terms survives, but it does not contribute to the total integral. This has been checked explicitly for the relatively simple cases with one and two "centers" in (2.32). The required cancellations are indeed found to occur but even at this stage the integration becomes laborious.

Though the foregoing considerations are sufficient to derive the action, it is interesting to see how the coordinate transformation of Appendix B shows immediately that the action cannot depend on parameters other than ζ in (2.32). Let us consider the consequences of the transformation of coordinates (B3) and (B5) which, of course, leaves the total action S invariant. For simplicity let us first consider the case of axial symmetry, namely, when in (2.32) $\cos \psi_j = \cos \theta$ for all j. Performing a transformation corresponding to the parameter $-c_1$ and using the group property (B27),

$$\cosh \eta_1 \to \cosh \rho \tag{3.20}$$

and, in general,

$$\cosh \eta_j \to \cosh(c_j - c_1) \cosh \rho - \sinh (c_j - c_1) \sinh \rho \cos \theta. \tag{3.21}$$

The conformal invariance of S and the property (B24) assures that it is sufficient to transform H as a scalar in (3.5). Hence H will now depend on the parameters c in the combination $c_j - c_1$. Hence the action can depend only on these combinations. But since they are quite arbitrary to start with, $c_j \to (c_j - c_1)$ amounts to a redefinition. Hence the conclusion is that S cannot depend on c_1. By symmetry it cannot depend on any of the c's. Setting all the c's equal to zero the a's are grouped together into a single overall factor which can be normalized away. Hence S cannot depend on the a's. Alternatively, one can use a parallel argument replacing the differences of the c's by the ratios of the a's. When there is no restriction to axial symmetry, a rotation $\phi_j \to (\phi_j - \phi_1)$ and an analogous argument would complete the picture. For higher order invariants the overall conformal factor of the metric is not irrelevant. The change in this factor will depend on c_1 for (3.21). Hence the above arguments do not hold. In the first approach the angular contributions to integrands of such invariants will not, in general, be grouped into a term, as in (3.18), giving vanishing integral.

IV. GAUGE TRANSFORMATIONS TO REAL FORMS OF ONE CENTER SOLUTIONS

Our gauge potentials are given by (2.5), (2.16), and (2.32). For the spherically symmetric case,

$$H = \frac{\sinh \zeta\rho}{\sinh \rho}, \tag{4.1}$$

a complex gauge transformation (as in the corresponding case of Refs. 3 and 5) by

$$U = \exp\left(-\zeta\rho \frac{\sigma_\rho}{2}\right) \tag{4.2}$$

with (U^{-1} now being distinct from U^{+})

$$B_\mu = U A_\mu U^{-1} + (i\partial_\mu U) U^{-1}, \tag{4.3}$$

reduces our solution to a familiar real form (i.e., with components B_μ^j real or the matrices B_μ-Hermitic) studied at length in a series of papers.[4,5,9] This solution is

$$\begin{aligned}
B_\tau &= -(\zeta \coth \zeta\rho - \coth\rho)\frac{\sigma_\rho}{2} \\
&= \chi_\rho \frac{\sigma_\rho}{2}, \\
B_\rho &= 0, \\
B_\theta &= -\left(\frac{\zeta \sinh\rho}{\sinh\zeta\rho} - 1\right)\frac{\sigma_\phi}{2} \\
&= -(e^x - 1)\frac{\sigma_\phi}{2}, \\
\frac{B_\phi}{\sinh\theta} &= \left(\frac{\zeta\sinh\rho}{\sinh\zeta\rho} - 1\right)\frac{\sigma_\theta}{2} \\
&= (e^x - 1)\frac{\sigma_\theta}{2},
\end{aligned} \tag{4.4}$$

with

$$e^x = \frac{\zeta\sinh\rho}{\sinh\zeta\rho}.$$

In Appendix B we have studied in detail a pseudotranslation that leads from spherical to axial symmetry (and eventually to ϕ dependence through an additional rotation). Let us now study the generalization of (4.2) for the one-center axially symmetric solution. Let us start directly with the coordinates (τ,η,ω,ϕ) and

$$H = \frac{\sinh\zeta\eta}{\sinh\eta}, \tag{4.5}$$

i.e., with (using evident definitions in terms of η,ω)

$$\begin{aligned}
A_\tau &= f_\eta \frac{\sigma_\eta}{2}, \quad A_\eta = i\zeta\frac{\sigma_\eta}{2} \\
\frac{A_\omega}{\sinh\omega} &= i\zeta\frac{\sigma_\omega}{2} \\
&\quad + \left(f_\eta - \frac{\cosh\eta - 1}{\sinh\eta}\right)\frac{\sigma_\phi}{2}, \\
\frac{A_\phi}{\sinh\omega\sin\theta} &= i\zeta\frac{\sigma_\phi}{2} \\
&\quad - \left(f_\eta - \frac{\cosh\eta - 1}{\sinh\eta}\right)\frac{\sigma_\omega}{2},
\end{aligned} \tag{4.6}$$

where

$$\begin{aligned}
f_\eta &= -(\zeta\coth\zeta\eta - \coth\eta), \\
\sigma_\eta &= \sin\omega\cos\phi\frac{\sigma_1}{2} \\
&\quad + \sin\omega\sin\phi\frac{\sigma_2}{2} + \cos\omega\frac{\sigma_3}{2},
\end{aligned}$$

and so on.

Now using the coordinate transformation (see Appendix B)

$$(\eta,\omega) \to (\rho,\theta),$$

when

$$\begin{aligned}
\cosh\rho &= \cosh c\cosh\eta + \sinh c\sinh\rho\cos\omega, \\
\sinh\rho\sin\theta &= \sinh\eta\sin\omega,
\end{aligned} \tag{4.7}$$

and so on, one obtains

$$\begin{aligned}
A'_\tau &= f_\eta\left[\cos(\omega-\theta)\frac{\sigma_\rho}{2} + \sin(\omega-\theta)\frac{\sigma_\theta}{2}\right], \\
A'_\rho &= i\zeta\left[\cos(\omega-\theta-\gamma)\frac{\sigma_\rho}{2} + \sin(\omega-\theta-\gamma)\frac{\sigma_\theta}{2}\right] \\
&\quad - \frac{1}{\sinh\rho}f_\theta\frac{\sigma_\phi}{2} - (\cosh\eta - 1)\frac{\partial\omega}{\partial\rho}\frac{\sigma_\phi}{2}, \\
\frac{A'_\phi}{\sinh\rho} &= i\zeta\left[\cos(\omega-\theta-\gamma)\frac{\sigma_\theta}{2}\right. \\
&\quad \left. - \sin(\omega-\theta-\gamma)\frac{\sigma_\rho}{2}\right] \\
&\quad + f_\rho\frac{\sigma^\phi}{2} - (\cosh\eta - 1) \\
&\quad \times\frac{1}{\sinh\rho}\frac{\partial\omega}{\partial\theta}\frac{\sigma_\phi}{2}, \\
\frac{A'_\theta}{\sinh\rho\sin\theta} &= i\zeta\frac{\sigma_\phi}{2} \\
&\quad - \left(f_\eta - \frac{\cosh\eta - 1}{\sinh\eta}\right)\left[\cos(\omega-\theta)\frac{\sigma_\theta}{2}\right. \\
&\quad \left. - \sin(\omega-\theta)\frac{\sigma_\rho}{2}\right].
\end{aligned} \tag{4.8}$$

All the results used, including the definition of γ, can be found in Appendix B.

This is not in the gauge used in our Sec. 2. It can be shown that a further (real) gauge transformation by

$$U = e^{i(\omega-\theta-\gamma)\sigma_\phi/2} \tag{4.9}$$

converts (4.8) to the gauge implied by our *ansatz*. Namely, U through

$$A_\mu = U A'_\mu U^{-1} + (i\partial_\mu U) U^{-1}$$

leads to

$$\begin{aligned}
A_\tau &= f_\eta\left(\cos\gamma\frac{\sigma_\rho}{2} + \sin\gamma\frac{\sigma_\theta}{2}\right) \\
&= f_\rho\frac{\sigma_\rho}{2} + \frac{f_\theta}{\sinh\rho}\frac{\sigma_\theta}{2},
\end{aligned}$$

and similarly to

$$\begin{aligned}
A_\rho &= i\zeta\frac{\sigma_\rho}{2} - \frac{f_\theta}{\sinh\rho}\frac{\sigma_\phi}{2}, \\
\frac{A_\theta}{\sinh\rho} &= i\zeta\frac{\sigma_\theta}{2} + \left(f_\rho - \frac{\cosh\rho - 1}{\sinh\rho}\right)\frac{\sigma_\phi}{2}, \\
\frac{A_\phi}{\sinh\rho\sin\theta} &= i\zeta\frac{\sigma_\phi}{2} \\
&\quad + \frac{f_\theta}{\sinh\rho}\frac{\sigma_\rho}{2} - \left(f_\rho - \frac{\cosh\rho - 1}{\sinh\rho}\right)\frac{\sigma_\theta}{2}.
\end{aligned} \tag{4.10}$$

In deriving (4.10) one has to use the relations

$$\frac{\partial\gamma}{\partial\rho} = \cosh\eta\,\frac{\partial\omega}{\partial\rho}, \quad \frac{\partial\gamma}{\partial\theta} = \cosh\eta\,\frac{\partial\omega}{\partial\theta} - \cosh\rho, \tag{4.11}$$

$$\cos\theta = \cosh\eta\sin\omega\sin\gamma + \cos\omega\cos\gamma, \tag{4.12}$$

$$\cosh\rho\sin\theta = \cosh\eta\sin\omega\cos\gamma - \cos\omega\sin\gamma. \tag{4.13}$$

From the above results one finds that a gauge transformation by

$$V = e^{-\zeta\eta\,\sigma_\eta/2}\,e^{-i(\omega-\theta-\gamma)\,\sigma_\phi/2} = e^{-i(\omega-\theta)\,\sigma_\phi/2}\,e^{-\zeta\eta\,\sigma_\rho/2}\,e^{i\gamma\,\sigma_\phi/2} \tag{4.14}$$

reduces (4.10) to a real form. Moreover, these real gauge potentials are those obtained by the coordinate transformation (4.7) starting from

$$B_\tau = \chi_\eta\,\frac{\sigma_\eta}{2}, \quad B_\eta = 0,$$

$$B_\omega = -(e^x - 1)\,\frac{\sigma_\phi}{2}, \quad \frac{B_\phi}{\sin\omega} = (e^x - 1)\,\frac{\sigma_\omega}{2}, \tag{4.15}$$

with $e^x = \zeta\sinh\eta/\sinh\zeta\eta$, and so on. After the transformation one obtains

$$B_\tau = \chi_\eta\,\frac{\sigma_\eta}{2}, \quad B_\rho = B_\omega\,\frac{\partial\omega}{\partial\rho},$$

$$B_\theta = B_\omega\,\frac{\partial\omega}{\partial\theta}, \quad B_\phi = \sin\omega(e^x - 1)\,\frac{\sigma_\omega}{2}, \tag{4.16}$$

where

$$\sigma_\eta = \cosh(\omega-\theta)\,\sigma_\rho + \sin(\omega-\theta)\,\sigma_\theta,$$

$$\sigma_\omega = \cos(\omega-\theta)\,\sigma_\theta - \sin(\omega-\theta)\,\sigma_\rho, \tag{4.17}$$

$$\frac{\partial\omega}{\partial\rho} = -\frac{\sinh c\sin\theta}{\sinh^2\eta},$$

$$\frac{\partial\omega}{\partial\theta} = -\frac{\sinh\rho}{\sinh^2\eta}(\sinh c\cosh\rho\cos\theta - \cosh c\sinh\rho). \tag{4.18}$$

The gauge potentials (4.15) (or 4.4) has been exploited in a series of papers to obtain many interesting results. The axial deformation (4.16) will naturally have analogous interesting properties while including one more parameter (c of B4). For completeness we add that under the limiting process (B14), (B15) (4.16) reduces to the "translated" Prasad–Sommerfield form, namely, to

$$B_t = (1 - K\eta\coth K\eta)\times\frac{1}{\eta^2}\left[(r - c\cos\theta)\,\frac{\sigma_r}{2} + c\sin\theta\,\frac{\sigma_\theta}{2}\right],$$

$$B_r = \left(\frac{K\eta}{\sinh\eta} - 1\right)\frac{1}{\eta^2}(c\sin\theta)\,\frac{\sigma_\phi}{2},$$

$$\frac{B_\theta}{r} = -\left(\frac{K\eta}{\sinh\eta} - 1\right)\frac{(c\cos\theta - r)}{\eta^2}\,\frac{\sigma_\phi}{2},$$

$$\frac{B_\phi}{r\sin\theta} = \left(\frac{K\eta}{\sinh K\eta} - 1\right)\times\frac{1}{\eta^2}\left[(r - c\cos\theta)\,\frac{\sigma_\theta}{2} - c\sin\theta\,\frac{\sigma_\rho}{2}\right], \tag{4.19}$$

where now

$$\eta^2 = r^2 + c^2 - 2cr\cos\theta.$$

Finally, we note that for $\zeta = 2$, (2.32) always reduces to the one-center form. Using results such as (2.34) and (2.35) with $\zeta = 2$, or more directly, one has for $\zeta = 2$, from (2.32) and (2.33),

$$H = \sum_j 2a_j\cosh\eta_j$$

$$= \sum_j 2a_j(\cosh c_j\cosh\rho - \sinh c_j\sinh\rho\cos\psi_j). \tag{4.20}$$

Grouping the coefficients of $\cosh\rho$, and so on, and normalizing suitably one gets back the standard one-center form. Hence the case $\zeta = 2$ can also be reduced to a real form through complex gauge transformations. We demonstrate in Appendix A that for more than one center in (2.32) and $\zeta = 3,4,\ldots$, this is no longer possible. Such solutions are, in this sense, intrinsically, complex.

V. ON THE POSSIBILITY OF FINITE ACTION FOR DIVERGENT, COMPLEX GAUGE POTENTIALS

For complex potentials a new possibility arises. The signs of different terms in the action density $2\,\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu})$ can be different. Hence mutual cancellations can render the density finite and even integrable, though certain terms may diverge individually. Here we are not talking about singularities in A_μ which can be gauged away, but which are essential and can survive in other gauge invariant quantities other than the action density. [The energy momentum tensor of course vanishes. This is a general property of all (anti) self-dual solutions on the Euclidean section.] Such a possibility is of course excluded for real solutions. There, for essentially divergent solutions, the weight factor e^{-S} in the path integral formalism is vanishing. Applications have nevertheless been envisaged for meron type solutions with logarithmically divergent action.[10] For paths corresponding to certain classes of divergent, complex solutions, S can be finite and even real (without oscillations in e^{-S} due to an imaginary part of S). There seems to be no *a priori* reason permitting us to ignore the possible contributions of such paths. We will now make it explicit how our class of solutions (for more than one center) furnishes an example of the above mentioned phenomenon in the simplest possible fashion, namely, through the role played by one single purely imaginary parameter.

This is not everywhere evident in terms of the coordinates τ and ρ. But the metric (2.9) is itself singular as $\rho\to\infty$. [For the De Sitter metric, where the conformal factor is $(\cosh\rho)^{-2}$, this limit corresponds to the cosmological horizon.] We have to study the properties of A_μ and $F_{\mu\nu}$ in terms of the usual flat space spherical coordinates (t,r,θ,ϕ). (Since there is no special trouble as $r\to 0$, it is not necessary to go over to flat Cartesian coordinates.) Let us concentrate on the region where $\rho\to\infty$ and hence [see (B39)] $t\to 0$ and $r\to 1$.

As $\rho\to\infty$, from (2.5), (A28), (A29), and (A30),

$$A_\tau = -(\zeta - 1)\frac{\sigma_\rho}{2} + O(e^{-\rho}),$$

$$A_\rho = i\zeta\frac{\sigma_\rho}{2} + O(e^{-\rho}), \tag{5.1}$$

$$\frac{A_\theta}{\sinh\rho} = i\zeta\tfrac{1}{2}(\sigma_\theta + i\sigma_\phi) + O(e^{-\rho}),$$

$$\frac{A_\phi}{\sinh\rho\sin\theta} = \zeta\tfrac{1}{2}(\sigma_\theta + i\sigma_\phi) + O(e^{-\rho}).$$

Hence these all remain finite. Hence A_θ and A_ϕ diverge. Now from (2.37) to (2.40), as $\rho\to\infty$,

$$A_t\to(A_\tau\cos\tau - A_\rho\sin\tau)\frac{e^\rho}{2},$$

$$A_r\to(A_\rho\cos\tau + A_\tau\sin\tau)\frac{e^\rho}{2}, \tag{5.2}$$

$$A_\theta\to\frac{e^\rho}{2}(i\zeta)\tfrac{1}{2}(\sigma_\theta + i\sigma_\phi),$$

$$A_\phi\to\sin\theta\cdot\frac{e^\rho}{2}\cdot\frac{\zeta}{2}(\sigma_\theta + i\sigma_\phi).$$

Now all the A_μ's are seen to diverge as e^ρ, which means that in terms of t and r they diverge linearly [as D^{-1} as $t\to 0$ and $r\to 1$, as is evidenct from (B36) and (B37)].

Divergences in A_μ can be a gauge effect. Indeed, for one center these divergences can be gauged away. The transformation to the real forms given in Sec. 4, accomplish this. This follows from (4.4) and (4.16). It should not be forgotten in this context that unless otherwise stated we are restricting ζ to positive integral values [see the remarks following (2.16)]. The role of integral values of ζ in assuring regularity properties of (4.4) has been discussed elsewhere[11] and can be directly extended to the axially symmetric case of Sec. 4. We will not repeat these considerations here.

We will now show that solutions with more than one center, which have been shown in Appendix A to be intrinsically complex [see (A20), for example] are also essentially divergent. The technique is similar to that used in Appendix A, namely, it is sufficient to construct one single gauge invariant quantity which is divergent as $\rho\to\infty$. Such a quantity is furnished, among others, by $\mathrm{Tr}(F_{t\theta})^2$. For simplicity we put $f_\phi = 0$, when from (A2), using (A28)–(A36) and also (3.17) with $f_\phi = 0$, one obtains

$$2\,\mathrm{Tr}\left(\frac{F_{\tau\theta}}{\sinh\rho}\right)^2 = 4\zeta e^{-2\rho}\{h_{1,\theta}^2 + (\zeta - 1)(h_{1,\theta\theta} - \cot\theta\, h_{1,\theta})\} + O(e^{-4\rho}), \tag{5.3}$$

where h_1 is given by (3.11) with $\psi_j = \theta$. We note that for one center, namely, for

$$h_1 = (\zeta - 1)\ln(\cosh c - \sinh c\cos\theta), \tag{5.4}$$

$$h_{1,\theta}^2 + (\zeta - 1)(h_{1,\theta\theta} - \cot\theta\, h_{1,\theta}) = 0.$$

But it is no longer zero for more than one center and $\zeta > 2$ [compare the situation presented by (A18)–(A22)]. Hence $\mathrm{Tr}(F_{\tau\theta})^2$ tends to $O(1)$ as $\rho\to\infty$. A similar result holds for $\mathrm{Tr}(F_{\tau\phi})^2$. Now one has

$$F_{tr} + \frac{1}{r^2\sin\theta}F_{\theta\phi} = \frac{\sinh^2\rho}{r^2}F_{\tau\rho}, \tag{5.5}$$

$$F_{t\theta} = \frac{F_{r\phi}}{\sin\theta} = F_{\tau\theta}(1 + \cosh\rho\cos\tau) - F_{\rho\theta}\sinh\rho\sin\tau,$$

$$F_{t\phi} = \sin\theta\, F_{r\theta} = F_{\tau\phi}(1 + \cosh\rho\cos\tau) - F_{\rho\phi}\sinh\rho\sin\tau.$$

Hence evidently $\mathrm{Tr}(F_{t\theta})^2$ diverges as $\rho\to\infty$ or as $t\to 0$ and $r\to 1$. A similar result holds for $\mathrm{Tr}(F_{t\phi})^2$ [though $\mathrm{Tr}(F_{tr})^2$ can be shown to remain finite as $\rho\to\infty$]. Thus we have established the announced result. That the action is finite even under these circumstances has already been established in Sec. 3. But the role of the imaginary value of K can be made more explicit in the following way. Using the results of Appendix A one obtains

$$\frac{1}{\sinh^2\rho\sin\theta}2\mathrm{Tr}(F_{\tau\rho}F_{\theta\phi} + F_{\tau\theta}F_{\phi\rho} + F_{\tau\phi}F_{\rho\theta}) = 2K^2(f_1^2 + f_2^2 + f_3^2) + 2(X_1^2 + X_2^2 + X_3^2) + (Y_1^2 + Y_2^2 + Y_3^2). \tag{5.6}$$

Using the asymptotic forms (A28)–(A36)

$$f_1^2 + f_2^2 + f_3^2 = (\zeta - 1)^2 + 4e^{-2\rho}\{(\vec{\nabla}_2 h_1)^2 - (\zeta - 1)h_2\} + O(e^{-4\rho}),$$

$$X_1^2 + X_2^2 + X_3^2 = 4e^{-2\rho}\zeta^2(\vec{\nabla}_2 h_1)^2 + O(e^{-4\rho}),$$

and

$$Y_1^2 + Y_2^2 + Y_3^2 = 2\zeta^2(\zeta - 1)^2 - 8\zeta^2(\zeta - 1)e^{-2\rho}h_2 + O(e^{-4\rho}), \tag{5.7}$$

where in the last equation (3.17) has been used. Hence, as $\rho\to\infty$,

$$2\,\mathrm{Tr}(F_{\tau\rho}F_{\theta\phi} + F_{\tau\theta}F_{\phi\rho} + F_{\tau\phi}F_{\rho\theta}) = (\sinh^2\rho\sin\theta)[2(\zeta^2 + K^2)\{(\zeta - 1)^2 + 4e^{-2\rho} \times(|\vec{\nabla}_2 h_1)^2 - (\zeta - 1)h_2)\} + O(e^{-4\rho})]. \tag{5.8}$$

Since $K^2 = -\zeta^2$, both divergence and nonintegrability are eliminated and we obtain a finite, real action $8\pi^2(\zeta - 1)$, as shown in Sec. 3.

For comparison let us consider the third order invariant (A12), namely,

$$\mathrm{Tr}(F_{\alpha\beta}[F^\beta{}_\gamma, F^{\gamma\alpha}]) = \frac{24}{\sqrt{g}\,g_{\tau\tau}}\mathrm{Tr}\,F_{\tau\rho}[F_{\tau\theta}, F_{\tau\phi}]), \tag{5.9}$$

where $g_{\mu\nu}$ are given by (2.9). Using the results of Appendix A,

$$\frac{\mathrm{Tr}\{F_{\tau\rho}[F_{\tau\theta}, F_{\tau\phi}]\}}{(\sinh^2\rho\sin\theta)} = O(e^{-4\rho}). \tag{5.10}$$

The coefficients of the two leading terms on the right-hand side vanish again as for the action density. But since

$$\sqrt{g}\,g_{\tau\tau} = \frac{\sinh^2\rho\sin\theta}{(\cosh\rho + \cos\tau)^6}, \tag{5.11}$$

the invariant (5.9) diverges. The factor $\sqrt{g}$ in the invariant

volume element in terms of (τ,ρ,θ,ϕ) still gives

$$\sqrt{g}\,\mathrm{Tr}\,(F_{\alpha\beta}[F^{\beta}{}_{\gamma},F^{\gamma\alpha}])\simeq O(1),\quad \rho\to\infty \tag{5.12}$$

and hence is not integrable. The effect of the conformal factor cancels out for the action density leading to finiteness and integrability.

So far we have been considering the situation as $\rho\to\infty$ and hence, on our chosen scale, $t\to 0$ and $r\to 1$. To compare with the well-known properties of real instanton solutions let us finally briefly note the situation as $t^2+r^2\equiv x^2\to\infty$. From (B36)–(B39) we see that as $x^2\to\infty$, $\rho\to 0$ and $\tau\to\pm\pi$ according to the sign of t. Developing H in powers of ρ and using (2.37) and (2.38), one obtains the corresponding properties of A_μ. It is convenient to express the results in terms of the Cartesian components, where $x^2=x_1^2+x_2^2+x_3^2+x_4^2$. It is easy to verify that A_1, A_2, A_3, and A_4 all are of $O(1/x^2)$ as $x\to\infty$. This may be compared with the fact that for real instantons they are of $O(1/x)$ in the Jackiw–Nohl–Rebbi gauge and of $O(1/x^3)$ in the 't Hooft gauge. Our A_μ's remain of course complex valued due to the presence of K. For one-center solutions only they have been shown to be reducible to real instantons. The Cartesian components of $F_{\mu\nu}$ can be shown, for the case of axial symmetry, to be of $O(1/x^4)$ as for the above mentioned real instantons. But for ϕ dependence ($f_\phi\neq 0$) some of them are of $O(1/x^3)$. This does not spoil integrability but some special cancellations [giving $O(1/x^4)$] no longer take place. Thus here again ϕ dependence brings about a discontinuity as concerning the number of parameters [see (2.35) and (2.36)].

VI. REMARKS

The imaginary part (the antihermitic one in the matrix formulation) enters into our *ansatz* in an extremely simple way, namely, as projection of $K(\sigma/2)$, where K is a purely imaginary *constant*. This suffices to lead to many remarkable features. K ($=i\zeta$, where ζ is real) turns out to play a nonperturbative role. The properties of the solutions do not everywhere vary smoothly with ζ, though one might have the contrary impression on looking at the form of the *ansatz* (2.5). For our class of solutions the action diverges for $\zeta<1$. [More precisely, for $|\zeta|<1$. But as our solutions are obtained in complex conjugate pairs it is sufficient to consider, as in (2.16), non-negative values.] This divergence is most immediately evident, in the gauge (4.4), for the simplest, namely, the spherically symmetric case, as $\rho\to\infty$. But this is true in general. (The effect of the logarithmic derivatives of H entering into the *ansatz* is such that H must diverge as $\rho\to\infty$ for the action to be finite. The case $\zeta=1$ or $H=$ constant is evidently trivial.) The action can indeed be verified to be finite [and still given by (3.19)] for all values of $\zeta>1$. *Thus, if only the criterion of finite action is imposed one obtains a continuous spectrum*. But in this spectrum the integer values ($\zeta=2,3,\ldots$) correspond to special properties. Discontinuous changes occur at these points. The effective number of parameters in H becomes finite for integer ζ, otherwise they are unlimited. [See the discussion leading to Eqs. (2.34)–(2.36).] We have shown elsewhere that the spherically symmetric solution (4.4) is regular only for integer ζ.[11] This holds for the one-center solution with angular dependence, studied in detail in Sec. 4, since this is related to the spherically symmetric case through a coordinate transformtion. Solutions with more than one center [of H in (2.32)] are both essentially complex and singular, though of finite action. They are not regular even for integer values of ζ. (See Sec. 5.) Hence the criterion of regularity no longer singles out the integer values of ζ. But such integer values are still set apart by their drastic consequence concerning the parameters in H. The linear deformations of such parameters would, for a given ζ fixed by the action, lead to a finite number of modes for integer ζ. We intend to study elsewhere such modes and fermions with our complex backgrounds.

Another source of discontinuous features is the relaxation of symmetry. For spherical symmetry [$H(\rho)$] there are no free parameters for a given action. For axial symmetry [$H(\rho,\theta)$], normalized H has effectively $(\zeta-1)$ parameters for integer ζ. As axial symmetry is broken this number jumps to ζ^2-1. We note that for axial symmetry the number $\zeta-1=n$ say, is well under the maximal number of parameters in real instanton solutions of index n, which is $8n-3$. But ϕ dependence changes the situation and this limit is crossed for $\zeta=7$. But in any case even the axially symmetric solutions, for more than one center, are not only essentially complex but also divergent (though of finite action). Presumably they do not fall within the framework of standard index theorems. The asymptotic behavior (as $x^2=t^2+r^2\to\infty$) also exhibits a discontinuity as axial symmetry is relaxed.

We have studied many important aspects of our *ansatz*. But we are far from being able to make categoric statements concerning the full significance of the existence of such solutions. Certain possibilities of further interesting explorations has already been mentioned above. Much more can be envisaged. It would be desirable to generalize our class of complex conjugate pairs of solutions and to try to construct the corresponding measures, though this would be difficult to achieve. One should, moreover, try to better understand the significance of integer values of ζ. Our solutions can also lead to explicit construction of complex periodic solutions. These would be generalizations of the corresponding results for the spherically symmetric case discussed elsewhere.[4] Certain results obtained here can again be useful concerning certain aspects of real instanton solutions. Let us explain this point and why we consider the possibility interesting. One crucial step in our construction was a suitable generalization of the flat (3-space) Helmholtz equation and of a class of its solutions. (See Sec. 2 and Appendix B). The flat Helmholtz Eq. (2.20) arises in the context of singly charged, complex monopoles of Manton,[3] but also in certain approaches to the construction of real multicharged monopoles.[12] (It is present again in the study of real, linear deformations of the real spherically symmetric monopoles. Our generalization has already been exploited in this context.[7] We have studied in preceding papers[4] the properties of a class of instantons ("summable chains") which tends very simply in an infinite action limit to the spherically symmetric, singly charged Prasad–Sommerfield monopole. (More precisely, to the case where the Euclidean A_t replaces the Higgs scalar.) This class is obtained from (4.4). In the standard 't Hooft gauge this

corresponds to the generating function, for $\zeta = n$ an integer

$$\left\{1 + \sum_{k=1}^{n-1} \frac{\csc^2(k\pi/n)}{[(t - \cot(k\pi/n))^2 + r^2]}\right\}. \tag{6.1}$$

Their simple properties lead, for example, to totally explicit expressions for Green's functions and quantum fluctuation determinants. Thus they provide the first example where the correction to the "dilute gas approximation" can be obtained really explicitly for arbitrary index.[4] Encouraged by this one may try to select out other classes of finite action self-dual solutions which have other interesting static (t-independent) solutions as limits. In this paper we have constructed the class which thus tends to the complex generalization of the $P - S$ monopole. But for reasons mentioned above our solutions (2.32) of (2.15) should prove helpful in the construction of a hierarchy of real self-dual solutions which tends in a relatively simple fashion to multicharged monopoles in an infinite action limit. One cannot have a simplicity comparable to (6.1) for these cases. But one fundamental criterion of simplicity will continue to hold. These solutions will all be "static" (τ-independent) when expressed in coordinates (τ,ρ,θ,ϕ). Such a property should continue to have interesting consequences. This is again another topic we intend to study elsewhere.

APPENDIX A: $F_{\mu\nu}$ AND GAUGE INVARIANT EXPRESSIONS, REAL AND COMPLEX

Using (2.5) and (2.11) the $F_{\mu\nu}$'s can be written compactly as follows.

$$F_{\tau\rho} = \frac{1}{\sinh^2\rho \sin\theta} F_{\theta\phi} = Y_1 \frac{\sigma_\rho}{2} + (X_3 - f_3 K)\frac{\sigma_\theta}{2} + (X_2 + f_2 K)\frac{\sigma_\phi}{2}, \tag{A1}$$

$$\frac{F_{\tau\theta}}{\sinh\rho} = \frac{F_{\phi\rho}}{\sinh\rho \sin\theta} = (X_3 + f_3 K)\frac{\sigma_\rho}{2} + Y_2 \frac{\sigma_\theta}{2} + (X_1 - f_1 K)\frac{\sigma_\phi}{2}, \tag{A2}$$

$$\frac{F_{\tau\phi}}{\sinh\rho \sin\theta} = \frac{F_{\rho\theta}}{\sinh\rho} = (X_2 - f_2 K)\frac{\sigma_\rho}{2} + (X_1 + f_1 K)\frac{\sigma_\theta}{2} + Y_3 \frac{\sigma_\phi}{2} \tag{A3}$$

where, by definition,

$$f_1 \equiv f_\rho, \quad f_2 \equiv \frac{f_\theta}{\sinh\rho}, \quad f_3 \equiv \frac{f_\phi}{\sinh\rho \sin\theta}, \tag{A4}$$

$$X_1 \equiv \frac{-1}{\sinh^2\rho \sin\theta}\{f_{\theta\phi} + (f_\theta - \cot\theta) f_\phi\}, \tag{A5}$$

$$X_2 \equiv \frac{-1}{\sinh\rho \sin\theta}\{f_{\phi\rho} + f_\phi(f_\rho - \coth\rho)\}, \tag{A6}$$

$$X_3 \equiv \frac{-1}{\sinh\rho}\{f_{\rho\theta} + (f_\rho - \coth\rho) f_\theta\}, \tag{A7}$$

$$Y_1 \equiv -f_{\rho\rho} + \frac{1}{\sinh^2\rho}\left(f_\theta^2 + \frac{f_\phi^2}{\sin^2\theta}\right), \tag{A8}$$

$$Y_2 \equiv -\frac{f_{\theta\theta}}{\sinh^2\rho} + \frac{f_\phi^2}{\sinh^2\rho \sin^2\theta} + f_\rho(f_\rho - \coth\rho), \tag{A9}$$

$$Y_3 \equiv \frac{-f_{\phi\phi}}{\sinh^2\rho \sin^2\theta} + \frac{f_\theta}{\sinh^2\rho}(f_\theta - \cot\theta) + f_\rho(f_\rho - \coth\rho). \tag{A10}$$

Introduction of the f_j, x_j and the Y_j's facilitate the calculation of certain gauge invariant expressions we want to discuss. The cyclic symmetry makes the cancellations of certain terms evident without substituting the detailed expressions (A4)–(A10).

Using (2.13) one obtains

$$Y_1 + Y_2 + Y_3 = (\vec{\nabla} f)^2 - (K^2 + 1) \tag{A11}$$

when the operator $\vec{\nabla}$ is defined by (2.27).

We have already seen that the action density (Sec. 3) is real. Indeed our *ansatz* was motivated to assure this reality. (See also the discussion in Sec. 5.) Let us now examine the higher order invariants. Using (A1)–(A3) one can easily calculate the invariants

$$\mathrm{Tr}\,(F_{\alpha\beta}\,[F^\beta{}_\gamma, F^{\gamma\alpha}]), \tag{A12}$$

and

$$\mathrm{Tr}\,J^2, \tag{A13}$$

where

$$J^{ij} = F^i_{\alpha\beta} F^{j\alpha\beta}, \tag{A14}$$

i,j being isospin indices. Using (2.9) and (2.12) the invariants can be calculated in terms of $F_{\tau\rho}$, $F_{\tau\theta}$, and $F_{\tau\phi}$. One finds that the third and fourth order invariant are both real (for $K = i\zeta$ with ζ real), since the coefficient of K^3 and K both turn out to be zero.

Since we are considering self-dual fields, the reality of the action density, (A12) and (A13), indicates that all polynomials in $F_{\mu\nu}$ invariant under both gauge and coordinate transformations are real for our solutions. This should follow from a suitable adaptation of the results of Ref. 13 to the Euclidean section. We, however, will not attempt a full discussion of this point. The reality of the action density is what is most important for us.

Let us now look at certain gauge invariant tensorial expressions of the type

$$\mathrm{Tr}\,(F_{\tau i} F_{\tau j}) \quad (i,j = \rho,\theta,\phi). \tag{A15}$$

From (A1)–(A10) one obtains for a purely imaginary $K\,(= i\zeta)$

$$2\,\mathrm{Tr}\,(F_{\tau\rho} F_{\tau\rho}) = Y_1^2 + X_2^2 + X_3^2 - (f_1^2 + f_3^2)\zeta^2 + 2i\zeta(f_2 X_2 - f_3 X_3), \tag{A16}$$

$$2\,\mathrm{Tr}\left(F_{\tau\rho}\frac{F_{\tau\theta}}{\sinh\rho}\right) = X_1 X_2 + (Y_1 + Y_2) X_3 + f_1 f_2 \zeta^2 + i\zeta\{(Y_1 - Y_2) f_3 + X_1 f_2 - X_2 f_1\}, \tag{A17}$$

and so on.

It can be verified that *when there are more than one center*, i.e., when in (2.32) there are more than one distinct terms, the imaginary parts in (A15) do not all vanish. This is true even when the centers are aligned, leading to axial sym-

metry without ϕ dependence. The significance of this is discussed in Sec. 1.4. An exception is the case $\zeta = 2$.

The imaginary parts vanish for $\zeta = 2$ because then (2.32) always reduces to a one-center form. This is discussed in Sec. 4, where (for one center in H) gauge and coordinate transformations relating the complex gauge potentials to real ones are given explicitly. Using the simplest nontrivial case, where

$$H = \frac{\sinh \zeta\rho}{\sinh \rho} + a \frac{\sinh \zeta\eta}{\sinh \eta} \tag{A18}$$

with

$$\cosh \eta = \cosh c \cosh \rho - \sinh c \sinh \rho \cos \theta, \tag{A19}$$

one can show that

$$\mathrm{Tr}\left(F_{\tau\rho} \frac{F_{\tau\phi}}{\sinh \rho \sin \theta} \right) \tag{A20}$$

$$= \frac{i\zeta}{a} \frac{f_\theta^2}{\sinh c \sinh \rho \sin \theta} \partial_\rho \times \frac{\sinh^3 \eta (\zeta \cosh \zeta\rho \sinh \rho - \sinh \zeta\rho \cosh \rho)}{\sinh^3 \rho \, (\zeta \cosh \zeta\eta \sinh \eta - \sinh \zeta\eta \cosh \eta)}$$

$$= 0 \quad \text{for } \zeta = 2, \tag{A21}$$

$$\neq 0 \quad \text{for } \zeta = 3,4, \ldots . \tag{A22}$$

Manton[3] establishes the intrinsically complex nature of his monopole solutions by studying the 't Hooft tensor defining the magnetic field. This is found to be complex except for spherically symmetric solutions and equivalent ones. Since we are not dealing with Higgs fields the situation here is different in certain respects, in spite of many formal analogies.

We are not motivated for defining a tensor exactly in the same way and the transformation properties of A_τ are different from those of a Higgs field Φ. Let us, however, note the following results for comparison. Let us consider the traces

$$-4i \, \mathrm{Tr} \{ A_\tau [F_{\tau j}, F_{\tau k}] \} \quad (j,k = \rho,\theta,\phi), \tag{A23}$$

and let us note the corresponding imaginary parts by I_{jk}. Then one can show that

$$I_{\rho\theta} = \frac{i\zeta}{\sinh \rho} (f_\rho \partial_\theta - f_\theta \partial_\rho) (\vec{\nabla} f)^2, \tag{A24}$$

$$I_{\theta\varphi} = \frac{i\zeta}{\sinh^2 \rho \sin \theta} (f_\theta \partial_\varphi - f_\varphi \partial_\theta) (\vec{\nabla} f)^2, \tag{A25}$$

$$I_{\varphi\rho} = \frac{i\zeta}{\sinh \rho \sin \theta} (f_\varphi \partial_\rho - f_\rho \partial_\varphi) (\vec{\nabla} f)^2, \tag{A26}$$

where

$$(\vec{\nabla} f)^2 = f_\rho^2 + \frac{1}{\sinh^2 \rho} \left(f_\theta^2 + \frac{1}{\sin^2 \theta} f_\varphi^2 \right).$$

For the general case (i.e., more than one center in (2.32) and $\zeta > 2$) these are again nonzero. However, while $i \, \mathrm{Tr} \{ \Phi [F_{\tau j}, F_{\tau k}] \}$, where Φ is a scalar, is totally gauge invariant, (A23) is invariant only under *static* (i.e., τ-independent) gauge transformations. The vanishing of the I's for $f_\theta = f_\varphi = 0$, i.e., for spherical symmetry, is evident. The cases $f_\rho = f_\theta = 0$ and $f_\varphi = f_\rho = 0$ correspond to infinite action when nontrivial.

Let us finally note the asymptotic forms of f_i, X_i, and Y_i [from (A4) to (A10)] as $\rho \to \infty$. They will be utilized in Sec. 5. We have, as $\rho \to \infty$ [see (3.10)],

$$-f = \ln H = (\zeta - 1)\rho + h_1 + e^{-2\rho} h_2 + O(e^{-4\rho}), \tag{A27}$$

where h_1 and h_2 depend on θ and φ. Hence we obtain the following developments:

$$f_1 = -(\zeta - 1) + 2 e^{-2\rho} h_2 + O(e^{-4\rho}), \tag{A28}$$

$$f_2 = -2 e^{-\rho} h_{1,\theta} + O(e^{-3\rho}), \tag{A29}$$

$$f_3 = -2 e^{-\rho} \frac{h_{1,\theta}}{\sin \theta} + O(e^{-3\rho}), \tag{A30}$$

$$X_1 = -\frac{4 e^{-2\rho}}{\sin \theta} \{ -h_{1,\theta\varphi} + (h_{1,\theta} + \cot \theta) h_{1,\varphi} \} + O(e^{-4\rho}), \tag{A31}$$

$$X_2 = -2\zeta e^{-\rho} \frac{h_{1,\theta}}{\sin \theta} + O(e^{-3\rho}), \tag{A32}$$

$$X_3 = -2\zeta e^{-\rho} h_{1,\theta} + O(e^{-3\rho}), \tag{A33}$$

$$Y_1 = 4 e^{-2\rho} \left(h_2 + h_{1,\theta}^2 + \frac{h_{1,\varphi}^2}{\sin \theta} + O(e^{-4\rho}) \right), \tag{A34}$$

$$Y_2 = \zeta(\zeta - 1) - 2 e^{-2\rho} \{ (\zeta - 1)(h_2 - 1) + \zeta h_2 \} + 4 e^{-2\rho} \left(h_{1,\theta\theta} + \frac{h_{1,\varphi}^2}{\sin^2 \theta} \right) + O(e^{-4\rho}), \tag{A35}$$

$$Y_3 = \zeta(\zeta - 1) - 2 e^{-2\rho} \{ (\zeta - 1)(h_2 - 1) + \zeta h_2 \} + 4 e^{-2\rho} \left\{ \frac{h_{1,\varphi\varphi}}{\sin^2 \theta} + (h_{1,\theta} + \cot \theta) h_{1,\theta} \right\} + O(e^{-4\rho}). \tag{A36}$$

APPENDIX B: A COORDINATE TRANSFORMATION

In Sec. 2 we introduced the coordinates (τ,ρ,θ,ϕ). Here we collect together some results on a useful coordinate transformation which sheds light on the structure of our solutions. In a certain limit, through rescaling, it will be shown to reduce to a translation in 3-space. It will thus provide a nontrivial generalization of the complex solutions of Ref. 3. This will furnish the key to the solutions constructed in Sec. 2. The solutions in Sec. 2 are linear superpositions of one basic solution expressed in terms of η, where

$$\cosh \eta = \cosh c \cosh \rho - \sinh c \sinh \rho \cos \psi, \tag{B1}$$

with

$$\cos \psi = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\phi - \phi'), \tag{B2}$$

c, θ', ϕ' being arbitrary parameters. We now display the significance of η in the context of a systematic coordinate transformation. First a simple rotation can bring the direction (θ', ϕ') on the third axis. Hence it is sufficient for our purpose to consider a transformation

$$(\tau,\rho,\theta,\phi) \to (\tau,\eta,\omega,\phi).$$

We define

$$\cosh \eta = \cosh c \cosh \rho - \sinh c \sinh \rho \cos \zeta, \tag{B3}$$

and

$$\cosh\rho = \cosh c \cosh\eta + \sinh c \sinh\eta \cos\omega. \tag{B4}$$

The second relation follows by imposing a symmetric inverse with $c \to -c$. Since we are aiming at generalizing translations this is essential. From (B4) and (B5) one obtains

$$\tan\omega = \frac{\sinh\rho \sin\theta}{\cosh c \cos\theta \sinh\rho - \sinh c \cosh\rho}. \tag{B5}$$

Useful relations are

$$\sinh\eta \sin\omega = \sinh\rho \sin\theta \tag{B6}$$

and

$$\sinh^2\eta = (\sinh c \cosh\rho - \cosh c \sinh\rho \cos\theta)^2 + \sinh^2\rho \sin^2\theta \tag{B7}$$

$$= (\cosh c \sinh\rho - \sinh c \cosh\rho \cos\theta)^2 + \sinh^2 c \sin^2\theta. \tag{B8}$$

Equations (B7) and (B8) are related by interchange of ρ and c and the different terms have the following significance: One has

$$\cos\omega = (\cosh c \sinh\rho \cos\theta - \sinh c \cosh\rho)/\sinh\eta, \tag{B9}$$

$$\sin\omega = \sinh\rho \sin\theta/\sinh\eta, \tag{B10}$$

and

$$\cos\gamma \equiv \frac{\partial\eta}{\partial\rho} = (\cosh c \sinh\rho - \sinh c \cosh\rho \cos\theta)/\sinh\eta, \tag{B11}$$

$$\sin\gamma \equiv \frac{1}{\sinh\rho}\frac{\partial\eta}{\partial\theta} = \sinh c \sin\theta/\sinh\eta. \tag{B12}$$

Indeed η satisfies the relation

$$(\vec{\nabla}\eta)^2 = \left(\frac{\partial\eta}{\partial\rho}\right)^2 + \frac{1}{\sinh^2\rho}\left(\frac{\partial\eta}{\partial\theta}\right)^2 = 1. \tag{B13}$$

Upon rescaling

$$(\rho,\eta,c) \to (\lambda\rho,\lambda\eta,\lambda c) \tag{B14}$$

and making

$$\lambda \to 0, \tag{B15}$$

(B4), (B5), and (B6) reduce, respectively, to

$$\eta^2 = \rho^2 + c^2 - 2c\rho\cos\theta, \tag{B16}$$

$$\rho^2 = \eta^2 + c^2 + 2c\rho\cos\omega, \tag{B17}$$

and

$$\eta \sin\omega = \rho \text{ sub } \theta. \tag{B18}$$

Also in this limit $(\omega - \theta) \to \gamma$. Thus, in this limit (rescaling τ analogously for completeness) one arrives at a translation in flat metric. This translation is along the third, namely the $\theta = 0$, axis. For generalizing translations in arbitrary directions the rotation mentioned after (B2) should be included.

We note that

$$d\eta = \frac{1}{\sinh\eta}[(\cosh c \sinh\rho - \sinh c \cosh\rho \cos\theta)\, d\rho + \sinh c \sin\theta \sinh\rho\, d\theta] \tag{B19}$$

and

$$\sinh\eta\, d\omega = \frac{1}{\sinh\eta}[-\sinh c \sin\theta\, d\rho + (\cosh c \sinh\rho - \sinh c \cosh\rho\cos\theta) \times \sinh\rho\, d\theta], \tag{B20}$$

implying

$$\frac{\partial\eta}{\partial\rho} = \frac{\sinh\eta}{\sinh\rho}\frac{\partial\omega}{\partial\theta}, \tag{B21}$$

$$\frac{1}{\sinh\rho}\frac{\partial\eta}{\partial\theta} = -\sinh\eta\frac{\partial\omega}{\partial\rho} \tag{B22}$$

with the Jacobian

$$\left\{\frac{\partial\eta}{\partial\rho}\frac{\partial\omega}{\partial\theta} - \frac{\partial\eta}{\partial\theta}\frac{\partial\omega}{\partial\rho}\right\} = \frac{\sinh\rho}{\sinh\eta} \tag{B23}$$

and

$$d\eta^2 + \sinh^2\eta(d\omega^2 + \sin^2\omega\, d\phi^2) = d\rho^2 + \sinh^2\rho(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{B24}$$

In

$$ds = dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2) = \frac{1}{(\cosh\rho + \cos\tau)^2}[d\tau^2 + d\rho^2 + \sinh^2\rho(d\theta^2 + \sin^2\theta\, d\varphi^2)], \tag{B25}$$

the conformal factor in (B25) is, however, changed, as indicated by (B5). [For De Sitter space the conformal factor would be $(\cosh\rho)^{-2}$.]

Let us also note the following group property: Composing two successive transformations one has for $(\rho,\theta) \to (\eta_1,\omega_1) \to (\eta_2,\omega_2)$

$$\cosh\eta_1 = \cosh c_1 \cosh\rho - \sinh c_1 \sinh\rho\cos\theta \tag{B26}$$

$$\cosh\eta_2 = \cosh c_2 \cosh\eta_1 - \sinh c_2 \sinh\eta_1 \cos\omega_1 = \cosh(c_1 + c_2)\cosh\rho - \sinh(c_1 + c_2)\sinh\rho\cos\theta \tag{B27}$$

and

$$\sinh\rho \sin\theta = \sinh\eta_1 \sin\omega_1 = \sinh\eta_2 \sin\omega_2. \tag{B28}$$

For completeness we display the implication of the transformation in terms of the embedding (normalized for simplicity)

$$z_1 = \frac{1}{\cosh\rho}\sin\tau, \quad z_2 = \frac{1}{\cosh\rho}\cos\tau,$$
$$z_3 = \tanh\rho \sin\theta \sin\phi, \quad z_4 = \tanh\rho \sin\theta \cos\phi,$$
$$z_5 = \tanh\rho\cos\theta, \tag{B29}$$

with

$$z_1^2 + z_2^2 + z_3^2 + z_4^2 + z_5^2 = 1. \tag{B30}$$

Defining analogously

$$\xi_1 = \frac{1}{\cosh\eta}\sin\tau, \quad \xi_2 = \frac{1}{\cosh\eta}\cos\tau,$$
$$\xi_3 = \tanh\eta \sin\omega \sin\phi,$$
$$\xi_4 = \tanh\eta \sin\omega \cos\phi, \tag{B31}$$
$$\xi_5 = \tanh\eta\cos\omega,$$

with

$$\xi_1^2 + \xi_2^2 + \xi_3^2 + \xi_4^2 + \xi_5^2 = 1, \tag{B32}$$

one obtains

$$\xi_i = \frac{z_i}{(\cosh c - z_5 \sinh c)}, \quad i = 1,2,3,4 \tag{B33}$$

$$\xi_5 = \frac{(z_5 \cosh c - \sinh c)}{(\cosh c - z_5 \sinh c)}. \tag{B34}$$

We have used the term "pseudotranslation" to emphasize the limit (B14)–(B16), giving translation in flat space. This is the essential feature for our purpose. But in the embedding space it is a nonlinear rotation of the sphere (B30). The infinitesimal form is

$$\xi_i = z_i + \epsilon z_i z_5$$

$$\xi_5 = z_5 - \epsilon\Big(\sum z_i^2\Big), \quad i = 1,2,3,4. \tag{B35}$$

The points $z_5 = \pm 1, z_i = 0$ are fixed.

Finally, we want to study the chain of transformations

$$(1)\ t + ir = \tan\left(\frac{\tau + i\rho}{2}\right)$$

with θ,φ unchanged.

$$(2)\ \cosh\eta = \cosh c \cosh\rho - \sinh c \sinh\rho \cos\theta,$$

$$\tan\omega = \frac{\sinh\rho \sin\theta}{\cosh c \sinh\rho \cos\theta - \sinh c \cosh\rho}$$

with τ,ϕ unchanged.

$$(3)\ t' + ir' = \tan\left(\frac{\tau + i\eta}{2}\right)$$

with ω,ϕ unchanged.

$$\begin{array}{ccc} (\tau,\rho,\theta,\phi) & \rightarrow & (\tau,\eta,\omega,\phi) \\ \uparrow & & \downarrow \\ (t,r,\theta,\phi) & \rightarrow & (t',r',\omega,\phi) \end{array}$$

For (1) some useful relations are, defining

$$D = [\{t^2 + (r+1)^2\}\{t^2 + (r-1)^2\}]^{1/2} \tag{B36}$$

$$\sinh\rho = \frac{2r}{D}, \quad \cosh\rho = \frac{1 + r^2 + t^2}{D}, \tag{B37}$$

$$\sin\tau = \frac{2t}{D}, \quad \cos\tau = \frac{1 - t^2 - r^2}{D}, \tag{B38}$$

$$t = \frac{\sin\tau}{\cosh\rho + \cos\tau}, \quad r = \frac{\sinh\rho}{\cosh\rho + \cos\tau}. \tag{B39}$$

There are analogous relations involving (t',r') and (τ,η). Using all these one can show that

$$\tan\omega = \frac{2r\sin\theta}{\cosh c\,(2r\cos\theta) - \sinh c\,(1 + t^2 + r^2)}, \tag{B40}$$

$$\cosh\eta = \frac{1}{D}[\cosh c\,(1 + t^2 + r^2) - \sinh c(2r\cos\theta)], \tag{B41}$$

$$t' = \frac{\sin\tau}{\cosh\eta + \cos\tau} = \frac{2t}{\cosh c\,(1 + t^2 + r^2) - \sinh c(2r\cos\theta) + (1 - t^2 - r^2)}, \tag{B42}$$

$$r' = \frac{\sinh\eta}{\cosh\eta + \cos\tau} = \frac{\{[\sinh c\,(1 + t^2 + r^2) - \cosh c(2r\cos\theta)]^2 + 4r^2\sin^2\theta\}^{1/2}}{\cosh c\,(1 + t^2 + r^2) - \sinh c(2r\cos\theta) + (1 - t^2 - r^2)}. \tag{B43}$$

Corresponding to the limit $\rho \rightarrow \infty$, when $\eta \rightarrow \infty$ we note that for

$$t = 0, r = 1 \quad \text{one has} \quad t = 0, r' = 1. \tag{B44}$$

Concerning the line element we note that

$$dt'^2 + dr'^2 + r'^2(d\omega^2 + \sin^2\omega\, d\phi^2)$$

$$= \frac{1}{(\cosh\eta + \cos\tau)^2}[d\tau^2 + d\eta^2 + \sinh^2\eta(d\omega^2 + \sin^2\omega\, d\phi^2)] \tag{B45}$$

$$= \frac{1}{(\cosh c\cosh\rho - \sinh c\sinh\rho\cos\theta + \cos\tau)^2}[d\tau^2 + d\rho^2 + \sinh^2\rho(d\theta^2 + \sin^2\theta\, d\phi^2)] \tag{B46}$$

$$= \frac{(\cosh\rho + \cos\tau)^2[dt^2 + dr^2 + r^2(d\theta + \sin^2\theta\, d\phi^2)]}{(\cosh c\cosh\rho - \sinh c\sinh\rho\cos\theta + \cos\tau)^2} \tag{B47}$$

$$= \frac{4[dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)]}{[\cosh c\,(1 + t^2 + r^2) - \sinh c\,(2r\cos\theta) + (1 - t^2 - r^2)]^2}. \tag{B48}$$

Various applications of the transformations of this Appendix will be found in the foregoing sections.

[1] H. Boutaleb-Joutei, A. Chakrabarti, and A. Comtet, Phys. Lett. B **101**, 249 (1981).

[2] See the relevant comments and references of Ref. 1.

[3] N. S. Manton, Nucl. Phys. B **135**, 319 (1978).

[4] H. Boutaleb-Joutei, A. Chakrabarti, and A. Comtet, (i) Phys. Rev. D **23**, 1781 (1981); (ii) Phys. Rev. D **24**, 3146 (1981).

[5] A related *ansatz* was used elsewhere for multicenter spaces. H. Boutaleb-joutei, A. Chakrabarti, and A. Comtet, (i) Phys. Rev. D **21**, 2280 (1980); (ii) **21**, 2285 (1980).

[6] Since the metric is diagonal the self-duality constraints reduce to $F_{ri}/(g_{rr}\, g_{ii})^{1/2} = [F_{jk}/(g_{jj}\, g_{kk})^{1/2}]\ \frac{1}{2}\epsilon_{ijk}$ $(i,j,k = \rho,\theta,\phi)$.

[7] A. Chakrabarti and K. S. Viswanathan, Phys. Rev. D **22**, 2569 (1980).

[8] See, for example, I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals Series and Products* (Academic, New York, 1965).

[9] H. Boutaleb-Joutei, A. Chakrabarti, and A. Comtet, Phys. Rev. D**20**, 1898 (1979).

[10]C. G. Callan, R. F. Dashen, and D. J. Gross, Phys. Rev. D **17**, 2717 (1978).
[11]See Sec. II of Ref. 5 (ii). There mostly Kruskal type coordinates was used. But, as mentioned there explicitly following its Eq. (2.15), the same conclusion follows in the context of the metric (2.1).
[12]M. K. Prasad, Commun. Math. Phys. **80**, 137 (1981).
[13]J. Anandan and R. Roskies, J. Math. Phys. **19**, 2614 (1978).

Some solutions of the Yang–Mills equations generalizing the Wu–Yang monopole

P. A. Hogan
Mathematical Physics Department, University College, Belfield, Dublin 4, Ireland
School of Theoretical Physics, Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Ireland

(Received 31 August 1982; accepted for publication 15 October 1982)

We present a new form of the Wu–Yang magnetic monopole solution of the SU(2) Yang–Mills equations on S_2 which is generalized to yield a solution of the SU($n+1$) Yang–Mills equations on $\mathbb{C}P_n$ for each integer $n \geqslant 1$.

PACS numbers: 11.15. – q, 11.30.Jw, 11.30.Ly, 14.80.Hv

1. INTRODUCTION

In 1968 Wu and Yang published a solution of the SU(2) Yang–Mills equations [solution (12a) in Ref. 1] which describes a Dirac magnetic monopole[2] embedded in SU(2) Yang–Mills theory. The embedded solution is free from the wire singularity which necessarily appears in the potential of Dirac's monopole. The gauge transformation relating the Dirac monopole to the Wu–Yang monopole is such that the pure gauge term in the transformation has a wire singularity. This is discussed by Wu and Yang[3] and by Actor.[4] The embedding of Dirac monopoles into SU(2) monopoles has recently been examined by Quirós[5] and Tafel.[6]

The Wu–Yang potential, considered as a 1-form field on Minkowskian space-time, in rectangular Cartesian coordinates and time, was originally given in the form[1]

$$A_\alpha = -\epsilon_{iaj}(x^j/r^2)dx^i, \tag{1.1}$$

where $\alpha = 1,2,3$ is the isotopic spin index, $i,j = 1,2,3$ with $x^i = (x,y,z), r^2 = x^2 + y^2 + z^2$, and the summation convention applies. A new form of this solution, for which the relationship with the Dirac monopole is perhaps more transparent, is given in Sec. 2.

We may visualize the Wu–Yang solution as a 1-form field on $S_2 (= \mathbb{C}P_1)$ with values in the Lie algebra of SU(2) (see Sec. 2). This is generalized in Sec. 3 to 1-form fields on $\mathbb{C}P_n$ with values in the Lie algebra of SU($n+1$), for each integer $n \geqslant 1$, which are shown to satisfy the Yang–Mills equations on $\mathbb{C}P_n$ with Fubini–Study metric.[7] This is the main result of the present paper which ends with a discussion of our results in Sec. 4.

2. THE WU–YANG MONOPOLE

We take the line-element of Minkowskian space-time in the form

$$ds^2 = 4r^2(1+\zeta\bar{\zeta})^{-2}d\zeta\, d\bar{\zeta} - 2dr\, d\sigma - d\sigma^2, \tag{2.1}$$

where ζ is a complex coordinate, with conjugate $\bar{\zeta}$, related to the polar angles θ,ϕ by $\zeta = e^{i\phi}\cot(\theta/2)$, while r is the radial polar coordinate and $\sigma = t - r$ is retarded time. We introduce the basis 1-forms

$$\theta^1 = \sqrt{2}r(1+\zeta\bar{\zeta})^{-1}d\zeta, \quad \theta^2 = \sqrt{2}r(1+\zeta\bar{\zeta})^{-1}d\bar{\zeta},$$

$$\theta^3 = d\sigma, \quad \theta^4 = dr + \tfrac{1}{2}d\sigma, \tag{2.2}$$

and note the dual

$$*(\theta^1 \wedge \theta^2) = -\theta^3 \wedge \theta^4. \tag{2.3}$$

Finally we define the operators

$$\Delta = (1+\zeta\bar{\zeta})^2\frac{\partial^2}{\partial\zeta\partial\bar{\zeta}}, \quad \partial = (1+\zeta\bar{\zeta})\frac{\partial}{\partial\zeta},$$

$$\bar{\partial} = (1+\zeta\bar{\zeta})\frac{\partial}{\partial\bar{\zeta}}. \tag{2.4}$$

As gauge potential we consider a 1-form field on Minkowskian space-time with values in su(2), the Lie algebra of SU(2), of the form

$$A = i\left(\frac{\partial P}{\partial\zeta}d\zeta - \frac{\partial P}{\partial\bar{\zeta}}d\bar{\zeta}\right) \tag{2.5a}$$

$$= \frac{i}{r\sqrt{2}}(\partial P\theta^1 - \bar{\partial}P\theta^2), \tag{2.5b}$$

where $P = P(\zeta,\bar{\zeta}) \in$ su(2). From (2.5b), A is singular on $r = 0$ (the t-axis) which is the world-line of the source. However, A may in general have wire singularities. It follows from (2.5a) that we may consider A as a 1-form field on S_2 with line-element given by (2.1) with r, σ both constant.

The gauge field is

$$F = dA + \tfrac{1}{2}[A,A] = -r^{-2}(i\Delta P - \tfrac{1}{2}[\partial P,\bar{\partial}P])\theta^1 \wedge \theta^2, \tag{2.6}$$

and so, using (2.3), the Yang–Mills equations

$$D^*F = d^*F + [A,^*F] = 0 \tag{2.7}$$

read

$$2i\partial(\Delta P) - 2[\partial P,\Delta P] - \partial[\partial P,\bar{\partial}P] - i[\partial P,[\partial P,\bar{\partial}P]] = 0. \tag{2.8}$$

We now asume that P satisfies

$$\Delta P + 2P = 0, \tag{2.9}$$

and as solution we take

$$P = (\zeta\bar{\zeta}+1)^{-1}\{(\zeta\bar{\zeta}-1)T_1 + i(\zeta-\bar{\zeta})T_2 + (\zeta+\bar{\zeta})T_3\}, \tag{2.10}$$

where T_1, T_2, T_3 are basis vectors of su(2) satisfying the commutation relations $[T_\alpha,T_\beta] = \epsilon_{\alpha\beta\gamma}T_\gamma$. It can then be shown that P satisfies the equations

$$[\partial P,\bar{\partial}P] = -2iP \tag{2.11}$$

and

$$[\partial P,P] = i\partial P. \tag{2.12}$$

 0022-2488/83/030674-03$02.50

Substitution of (2.9), (2.11), and (2.12) into (2.8) shows that P given by (2.10) satisfies (2.8). The gauge potential is given by (2.5) and (2.10) while substitution of (2.10) into (2.6) yields the gauge field

$$F = ir^{-2}P\theta^1 \wedge \theta^2. \tag{2.13}$$

Thus F is a Dirac monopole field multiplied by P. We have thus obtained a new form of the Wu–Yang monopole. The expression (1.1) may be recovered from (2.5) and (2.10) by introducing rectangular Cartesian coordinates.

A gauge transformation takes the form

$$A \to Q^\dagger AQ + Q^\dagger dQ, \tag{2.14a}$$

$$F \to Q^\dagger FQ, \tag{2.14b}$$

with $Q \in SU(2)$. By (2.10), $P^2 = -\frac{1}{4}1$, and so in view of (2.13) we can choose $Q \in SU(2)$ such that

$$Q^\dagger PQ = T_1 = \frac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \tag{2.15}$$

This is found to be given by

$$Q = \begin{pmatrix} -\rho\zeta & \rho \\ -\rho & -\rho\bar{\zeta} \end{pmatrix} \tag{2.16}$$

with $\rho^{-2} = 1 + \zeta\bar{\zeta}$. A straightforward calculation now reveals that

$$Q^\dagger AQ + Q^\dagger dQ = i\rho^2(\zeta d\bar{\zeta} - \bar{\zeta}d\zeta)T_1, \tag{2.17}$$

which is the potential of a Dirac monopole of strength $g = \frac{1}{2}$ (see, for example Ref. 12).

We note that when (2.10) is substituted into (2.5b), A is then singular only when $r = 0$, while (2.17), when expressed on the basis (2.2), is singular not only at $r = 0$ but also when $\zeta = \infty$ (i.e., when $\theta = 0$). This latter is the wire singularity mentioned in the Introduction and is due to the fact that $Q^\dagger\, dQ$ in (2.17) is singular when $\theta = 0$.

3. SU($n+1$)—SOLUTIONS ON $\mathbb{C}P_n$

As a generalization of (2.5a) and (2.10) we consider a 1-form field on $\mathbb{C}P_n, n \geqslant 1$, of the form

$$A = i\left(\frac{\partial P}{\partial \zeta_a}d\zeta_a - \frac{\partial P}{\partial \bar{\zeta}_a}d\bar{\zeta}_a\right), \tag{3.1}$$

where, in this section, Latin indices take values 1,2,3,...,n while Greek indices take values 0,1,2,...,n and the summation convention applies unless otherwise stated. To generalize (2.10) we take

$$P_{\alpha\beta} = i[(n+1)^{-1}\delta_{\alpha\beta} + \rho^2(-1)^{\alpha+\beta+1}\bar{\zeta}_\alpha\zeta_\beta] \in su(n+1), \tag{3.2}$$

where there is no summation over repeated Greek indices, $\zeta_0 = 1$, and

$$\rho^{-2} = 1 + \zeta_a\bar{\zeta}_a. \tag{3.3}$$

We take the Fubini–Study line-element on $\mathbb{C}P_n$ given by

$$ds^2 = \rho^4(\rho^{-2}d\bar{\zeta}_a d\zeta_a - \zeta_a\bar{\zeta}_b d\bar{\zeta}_a d\zeta_b). \tag{3.4}$$

The main result of the present paper can now be stated in the following theorem.

Theorem: The 1-form field A given by (3.1) and (3.2) on $\mathbb{C}P_n (n \geqslant 1)$, with Fubini–Study metric given by (3.4), satisfies the Yang–Mills equations (2.7) for each integer $n \geqslant 1$. The case $n = 1$ corresponds to the Wu–Yang monopole.

Proof: It is a simple matter to check that if $n = 1$, (3.1) and (3.2) reduce to (2.5a) and (2.10) with $\zeta = \zeta_1$ while (3.4) becomes the line-element on S_2. For $n > 1$ the following analogs of (2.11) and (2.12), respectively, are found to be satisfied:

$$(\delta_{bc} + \bar{\zeta}_b\zeta_c)\left[\frac{\partial P}{\partial \zeta_a}, \frac{\partial P}{\partial \bar{\zeta}_b}\right] = i(\delta_{bc} + \bar{\zeta}_b\zeta_c)\frac{\partial^2 P}{\partial \zeta_a \partial \bar{\zeta}_b} \tag{3.5}$$

and

$$\frac{\partial P}{\partial \zeta_a} + i\left[\frac{\partial P}{\partial \zeta_a}, P\right] = 0. \tag{3.6}$$

As basis 1-forms on $\mathbb{C}P_n$ we choose

$$\sigma^{2a-1} = (1/\sqrt{2})\rho^2(\rho^{-2}\delta_{ab} - \bar{\zeta}_a\zeta_b)d\bar{\zeta}_b, \quad \sigma^{2a} = (1/\sqrt{2})\rho^2 d\zeta_a, \tag{3.7}$$

and thus (3.4) reads

$$ds^2 = 2\sigma^{2a-1}\sigma^{2a}. \tag{3.8}$$

Now, using (3.5) and (3.6), the gauge field can be written

$$F = 2i\rho^{-2}(\delta_{bc} + \bar{\zeta}_b\zeta_c)\frac{\partial^2 P}{\partial \zeta_a \partial \bar{\zeta}_b}\sigma^{2c-1} \wedge \sigma^{2a} \tag{3.9}$$

with the matrix $P = (P_{\alpha\beta})$ given by (3.2).

In order to substitute (3.9) into the Yang–Mills equations (2.7) we have to evaluate expressions of the form $d^*(\sigma^{2c-1} \wedge \sigma^{2a})$ and $d\zeta_e \wedge {}^*(\sigma^{2c-1} \wedge \sigma^{2a})$. This is done using (3.7) and the quantities $\eta^{2a} = {}^*\sigma^{2a}, \eta^{2a-1} = {}^*\sigma^{2a-1}$, $\eta^{2c-1,2a} = {}^*(\sigma^{2c-1} \wedge \sigma^{2a})$ which satisfy the equations (see Refs. 8 and 9 for definitions and derivations)

$$d\eta^{2c-1,2a} = \sqrt{2}(n\bar{\zeta}_c\eta^{2a} - \delta_{ac}\bar{\zeta}_d\eta^{2d}) \tag{3.10}$$

and

$$d\zeta_e \wedge \eta^{2c-1,2a} = -\sqrt{2}\rho^{-2}\delta_{ce}\eta^{2a}. \tag{3.11}$$

Using these we find

$$\begin{aligned} D^*F = {} & 2\sqrt{2}i\rho^{-2}(\delta_{bc} + \bar{\zeta}_b\zeta_c)\frac{\partial^2 P}{\partial \zeta_a \partial \bar{\zeta}_b}(n\bar{\zeta}_c\eta^{2a} \\ & - \delta_{ac}\bar{\zeta}_d\eta^{2d}) \\ & - 2\sqrt{2}i\rho^{-2}\left\{\frac{\partial}{\partial \zeta_c}\left(\rho^{-2}(\delta_{bc} + \bar{\zeta}_b\zeta_c)\frac{\partial^2 P}{\partial \zeta_a \partial \bar{\zeta}_b}\right)\right. \\ & \left. + i\rho^{-2}\left[\frac{\partial P}{\partial \zeta_c}, (\delta_{bc} + \bar{\zeta}_b\zeta_c)\frac{\partial^2 P}{\partial \zeta_a \partial \bar{\zeta}_b}\right]\right\}\eta^{2a} \\ & + 2i\left\{\frac{\partial}{\partial \bar{\zeta}_e}\left(\rho^{-2}(\delta_{bc} + \bar{\zeta}_b\zeta_c)\frac{\partial^2 P}{\partial \zeta_a \partial \bar{\zeta}_b}\right)\right. \\ & \left. - i\rho^{-2}\left[\frac{\partial P}{\partial \bar{\zeta}_e}, (\delta_{bc} + \bar{\zeta}_b\zeta_c)\frac{\partial^2 P}{\partial \zeta_a \partial \bar{\zeta}_b}\right]\right\} \\ & \times d\bar{\zeta}_e \wedge \eta^{2c-1,2a}. \end{aligned} \tag{3.12}$$

Referring now to the formulas quoted, for convenience, in the Appendix we find, using (A3), that the coefficient of $d\bar{\zeta}_e \wedge \eta^{2c-1,2a}$ vanishes while use of (A1) and (A2) implies that the remaining terms cancel and so we have

$$D^*F = 0 \tag{3.13}$$

and the theorem is established.

4. DISCUSSION

We wish to point out the following two properties of the solutions of the Yang–Mills equations given in Sec. 3:

Equation (3.6) together with its Hermitian conjugate can be written in the form

$$DP = dP + [A,P] = 0, \tag{4.1}$$

which is the infinitesimal form of the invariance of the potential, given by (3.1) and (3.2), under gauge transformations generated by P (see Refs. 10 and 11).

Trautman[12] has proved that the natural connections on the U(1) Hopf bundles $S_{2n+1} \to \mathbb{C}P_n (n \geqslant 1)$ satisfy Maxwell's equations, with Fubini–Study metric on $\mathbb{C}P_n$. The case $n = 1$ is Dirac's monopole and this prompts one to ask if our solutions correspond to Trautman's Maxwell fields embedded in SU($n + 1$) Yang–Mills theory for $n > 1$? In general the answer is "no." For example, in the case $n = 2$ Trautman's Maxwell field (an "electromagnetic instanton") on $\mathbb{C}P_2$ is self-dual[12] whereas a simple calculation reveals that our su(3) field on $\mathbb{C}P_2$ is not.

ACKNOWLEDGMENTS

I wish to thank Dr. Jacek Tafel of the University of Warsaw for many informative and stimulating discussions on Wu–Yang monopoles and Dr. J. D. McCrea for his helpful comments during the course of this work.

APPENDIX

We list here three formulas required for the proof of the theorem in Sec. 3:

$$\begin{aligned}
&\frac{\partial}{\partial \zeta_c}\left(\rho^{-2}(\delta_{bc} + \bar{\zeta}_b\zeta_c)\frac{\partial^2 P_{\alpha\gamma}}{\partial\zeta_a\partial\bar{\zeta}_b}\right) \\
&\quad + i\rho^{-2}\left[\frac{\partial P}{\partial \zeta_c}, (\delta_{bc} + \bar{\zeta}_b\zeta_c)\frac{\partial^2 P}{\partial\zeta_a\partial\bar{\zeta}_b}\right]_{\alpha\gamma} \\
&= i\rho^4(n-1)(-1)^{\alpha+\gamma}\bar{\zeta}_a\bar{\zeta}_\alpha\zeta_\gamma + i\rho^2 n(-1)^{\gamma+1}\delta_{\alpha 0}\bar{\zeta}_a\zeta_\gamma \\
&\quad + i\rho^2\delta_{\alpha\gamma}\bar{\zeta}_a + i\rho^2 n(-1)^{\alpha+\gamma+1}\delta_{\gamma a}\bar{\zeta}_\alpha \\
&\quad + in(-1)^\gamma\delta_{\alpha 0}\delta_{\gamma a}
\end{aligned} \tag{A1}$$

$$= X_{\alpha\gamma a},$$

$$(\delta_{bc} + \bar{\zeta}_b\zeta_c)\frac{\partial^2 P_{\alpha\gamma}}{\partial\zeta_a\partial\bar{\zeta}_b}(n\bar{\zeta}_c\eta^{2a} - \delta_{ac}\bar{\zeta}_d\eta^{2d}) = X_{\alpha\gamma a}\eta^{2a}, \tag{A2}$$

and

$$\begin{aligned}
&\frac{\partial}{\partial\bar{\zeta}_e}\left(\rho^{-2}(\delta_{bc} + \bar{\zeta}_b\zeta_c)\frac{\partial^2 P}{\partial\zeta_a\partial\bar{\zeta}_b}\right) \\
&\quad = i\rho^{-2}\left[\frac{\partial P}{\partial\bar{\zeta}_e}, (\delta_{bc} + \bar{\zeta}_b\zeta_c)\frac{\partial^2 P}{\partial\zeta_a\partial\bar{\zeta}_b}\right],
\end{aligned} \tag{A3}$$

where $P = (P_{\alpha\beta})$ is given by (3.2).

[1] T. T. Wu and C. N. Yang, "Some Solutions of the Classical Isotopic Gauge Field Equations," in *Properties of Matter under Unusual Conditions*, edited by H. Mark and S. Fernbach (Interscience, New York, 1968), p. 349.
[2] P. A. M. Dirac, Proc. R. Soc. London Sec. A **133**, 60 (1931).
[3] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).
[4] A. Actor, Rev. Mod. Phys. **51**, 461 (1979).
[5] M. Quirós, "Embedding Dirac Monopoles into SU(2) n-Monopoles" in *Monopoles in Quantum Field Theory*, edited by N. S. Craigie, P. Goddard, and W. Nahm (World Scientific, Singapore, 1982), p. 281.
[6] J. Tafel, "Wu–Yang Monopoles Seminar," University College, Dublin, May 1982 (unpublished).
[7] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1969), Vol. II, p. 160.
[8] A. Trautman, Bull. de L'Académie Polonaise de Sciences **20**, 185 (1972).
[9] P. A. Hogan, "Some Yang–Mills Fields Constructed from Maxwell Fields," J. Math. Phys. (to appear).
[10] P. G. Bergmann and E. J. Flaherty, J. Math. Phys. **19**, 212 (1978).
[11] A. Trautman, J. Phys. A: Math. Gen. **13**, L1 (1980).
[12] A. Trautman, Int. J. Theor. Phys. **16**, 561 (1977).

Normalization of Bethe-ansatz states in the chiral-invariant Gross–Neveu model

C. Destri
Department of Physics, New York University, New York University, New York, New York 10003[a]

J. H. Lowenstein
Laboratoire de Physique Théorique et Hautes Energies, Université Pierre et Marie Curie, Tour 16-ler étage, 4, place Jussieu, 75230 Paris Cedex 05, France and Department of Physics, New York University, New York, New York 10003[a]

(Received 27 May 1982; accepted for publication 7 October 1982)

The norms of the Bethe-ansatz eigenstates of the chiral-invariant Gross–Neveu Hamiltonian are calculated.

PACS numbers: 11.30.Rd, 11.30.Jw, 11.15.Bt

I. INTRODUCTION

In Refs. 1 and 2 the chiral-invariant Gross–Neveu Hamiltonian

$$H = -i(\psi^\dagger_{a+}\partial_x\psi_{a+} - \psi^\dagger_{a-}\partial_x\psi_{a-}) + 4g\psi^\dagger_{a+}\psi^\dagger_{b-}\psi_{a-}\psi_{b+} \tag{1}$$

was diagonalized using the Bethe–ansatz technique. For this purpose, as well as for that of extracting the S-matrix for physical particles,[3] there was no necessity to normalize the energy eigenstates. However, if we are to make further progress in the direction of a complete solution of the field-theoretic model (in the form of correlation functions of renormalized fields), we shall certainly require a detailed knowledge of the norms of our basis states. Our aim in this note is to obtain useful and instructive norm formulas for the model, making use of the general results (valid for a wide class of models closely related to the one-dimensional Heisenberg ferromagnet) established recently by Korepin.[4] In addition, we shall exploit a newly established classification of all physically relevant solutions of the Bethe-ansatz equations for SU(2) internal symmetry.[5]

II. BETHE–ANSATZ EIGENSTATES AND THEIR GENERAL NORM FORMULA

Restricting ourselves to SU(2) internal symmetry ("color"), we may write the energy eigenstates constructed in Refs. 1 and 2 in the form[6]

$$|F\rangle = \int d^N x F(x_1\cdots x_N)_{a_1\cdots a_N} \times \prod_{i=1}^{N_+}\psi^\dagger_{a_i+}(x_i)\prod_{j=N_++1}^{N_++N_-}\psi^\dagger_{a_j-}(x_j)|0\rangle, \tag{2}$$

where $|0\rangle$ is the Fock reference state annihilated by the canonical fermion fields $\psi_{a\alpha}(x)$, $a = \pm\frac{1}{2}$, $\alpha = \pm 1$. More specifically, the function F, interpreted as the wave function for $N(=N_+ + N_-)$ pseudoparticles on a reentrant line segment of length L, is of the form

$$L^{-N/2}\exp\left(\frac{2\pi i}{L}\sum_{j=1}^{N_+}n_j^+ x_j + \frac{2\pi i}{L}\sum_{l=1}^{N}n_l^- x_{N_++l}\right) \times\exp\left(ik_+\sum_{j=1}^{N_+}x_j + ik_-\sum_{l=1}^{N_-}x_{N_++l}\right) \times\sum_{Q\in S_N}[U(Q)\Phi]_{a_1\cdots a_N}\theta(x_Q), \tag{3}$$

where

(a) the sum Σ_Q is over all permutations Q of 1,2,...,N;

(b) the step function $\theta(x_Q)$ is equal to unity for $x_{Q_1} < x_{Q2} < \cdots < x_{QN}$ and vanishes otherwise;

(c) for each sign of the chirality α, $n_j^\alpha, j = 1,...,N_\alpha$, are distinct integers;

(d) $U(Q)$ is a matrix which is unitary with respect to the inner product

$$(\xi,\xi') = \sum_{a_1\cdots a_N}\xi^*_{a_1\cdots a_N}\xi'_{a_1\cdots a_N};$$

(e) the quantity $\Sigma_Q[U(Q)\Phi]\theta(x_Q)$ is symmetric with respect to interchange of x_i, a_i with x_j, a_j, provided $\alpha_i = \alpha_j$;

(f) the spinorial tensor $\Phi_{a_1\cdots a_N}$ and the collective momenta $k_\pm$ are the functions of M complex parameters $\chi_1,...,\chi_M$ (for a state of total color spin $S = \frac{1}{2}N - M$) satisfying the "Bethe-ansatz equations"

$$\varphi_j(\chi_1\cdots\chi_M) = \text{integer}, \quad j = 1,...,M, \tag{4}$$

where

$$2\pi\varphi_j(\chi_1\cdots\chi_M) = i\sum_{l=1}^{N}\ln\left(\frac{\chi_j - \lambda_l + i\pi/2}{\chi_j - \lambda_l - i\pi/2}\right) - i\sum_{\substack{j'=1\\ j'\neq j}}^{M}\ln\left(\frac{\chi_j - \chi_{j'} + i\pi}{\chi_j - \chi_{j'} - i\pi}\right)$$

$$\lambda_l = \frac{\alpha_l\pi}{c}; \quad c = \frac{4g}{1-g^2}.$$

Making use of the canonical anticommutation relations, the norm of a state F with the properties described above is calculated to be

[a] Research supported in part by National Science Foundation Grant No. PHY-8116102.

 0022-2488/83/030677-06$02.50

$$\langle F|F\rangle = \int d^N x \sum_{a_1\cdots a_N} \sum_{P\in S_N\times S_N} (-)^P F^*(x_1\cdots x_N)_{a_1\cdots a_N} \times F(x_{P1}\cdots x_{PN})_{a_{P1}\cdots a_{PN}}$$
$$= \sum_{Q\in S_N} \theta(Q)(U(Q)\Phi, U(Q)\Phi) = (\Phi,\Phi).$$

For a given solution set $\{\chi_j\}$ of (5), the corresponding Φ comprises a color multiplet, $\{\Phi(S,S_z); S_z = -S\cdots +S\}$. Given $\Phi(S,S)$, the remaining members of the multiplet may be obtained by repeated application of the spin lowering operator $S^- = \sum_{n=1}^N S_n^-$:

$$\Phi(S,S_z) = [(S+S_z)!]^{1/2}[(2S)!(S-S_z)!]^{-1/2}(S^-)^{S-S_z}\Phi(S,S). \tag{6}$$

With the normalization given in (6), all members of the same multiplet have the same norm and we may henceforth restrict our attention to those Φ with maximal S_z. According to Ref. 7 such a Φ has the explicit form

$$\Phi = \sum_{1\le y_1<y_2<\cdots<y_M\le N} \Phi(y_1,\cdots,y_M)\prod_{j=1}^M S_{y_j}^-\Phi_0,$$

where

$$\Phi_{0a_1\cdots a_N} = \prod_{l=1}^N \delta_{+1/2,a_l},$$
$$\Phi(y_1\cdots y_M) = \sum_{P\in S_M} A(\chi_{P1}\cdots\chi_{PM})\prod_{j=1}^M f(\chi_{Pj},y_j), \tag{7}$$
$$f(\chi,y) = \frac{-i\pi}{\chi-\lambda_y-i\pi/2}\prod_{l=1}^{y-1}\frac{\chi-\lambda_l+i\pi/2}{\chi-\lambda_l-i\pi/2},$$
$$A(\chi_1,\cdots,\chi_M) = \prod_{j<k}\left(1+\frac{i\pi}{\chi_j-\chi_k}\right).$$

In Ref. 6, the discrete wave function $\Phi(y_1,\cdots,y_M)$ was determined up to an arbitrary y-independent overall factor. In (7) this factor [partially absorbed in the definition of $f(\chi,y)$] has been chosen to better conform to the conventions of the Quantum Inverse Scattering formalism used by Korepin and many other authors.[4,8] In particular, we may write

$$\Phi = B(\chi_1+i\pi/2)B(\chi_2+i\pi/2)\cdots B(\chi_M+i\pi/2)\Phi_0, \tag{8}$$

where $B(\chi)$ is an element of the so-called monodromy matrix

$$T(\chi) = \begin{pmatrix} A(\chi) & B(\chi) \\ C(\chi) & D(\chi)\end{pmatrix} = L_N(\chi-\lambda_N)\cdots L_1(\chi-\lambda_1),$$
$$L_n(\chi) = a(\chi)\mathbb{1} + b(\chi)P_n,$$
$$a(\chi) = \chi(\chi-i\pi)^{-1}, a(\chi)+b(\chi) = 1, \tag{9}$$
$$P_n = \tfrac{1}{2}(1+\sigma^\alpha\otimes\sigma_n^\alpha) = \frac{1}{2}\begin{pmatrix} 1+\sigma_n^3 & \sigma_n^1-i\sigma_n^2 \\ \sigma_n^1+i\sigma_n^2 & 1-\sigma_n^3\end{pmatrix}.$$

Here σ_n^α, $\alpha = 1,2,3$ are the usual Pauli matrices, associated with site n on a one-dimensional lattice of length N, with

$$[\sigma_n^\alpha,\sigma_m^\beta] = 2i\delta_{nm}\epsilon_{\alpha\beta\gamma}\sigma_n^\gamma.$$

The commutation relations among $A(\chi), B(\chi), C(\chi), D(\chi)$ may be expressed concisely as

$$R(\chi-\chi')[T(\chi)\otimes T(\chi')] = [T(\chi')\otimes T(\chi)]R(\chi-\chi'), \tag{10}$$

where $R(\chi)$ is the 4×4 matrix

$$R(\chi) = a(\chi)P + b(\chi)\mathbb{1}, \quad P = \tfrac{1}{2}(1+\sigma^\alpha\otimes\alpha^\alpha). \tag{11}$$

Actually, in Ref. 4, Korepin used a different normalization for $L_n(\chi)$. According to our normalization, his general formula for the norms of the XXX-type models (based on the conjecture of Gaudin, McCoy, and Wu[8] reads

$$(\Phi,\Phi) = \pi^M\left[\prod_{j\ne k}^M a(\chi_j-\chi_k)\right]^{-1}\det(2\pi\varphi_{jk})\cdot(\Phi_0,\Phi_0), \tag{12}$$

where (Φ_0,Φ_0), the norm of the reference state, is in this case just 1, and, from (4),

$$\varphi_{jj} \equiv \frac{\partial\varphi_j}{\partial\chi_j} = \sum_{\alpha=\pm} N_\alpha D_{1/2}\left(\alpha\frac{\pi}{c}-\chi_j\right) - \sum_{\substack{k=1\\k\ne j}}^M D_1(\chi_j-\chi_k),$$
$$\varphi_{jk} \equiv \frac{\partial\varphi_j}{\partial\chi_k} = \frac{\partial\varphi_k}{\partial\chi_j} = D_1(\chi_j-\chi_k), \quad j\ne k, \tag{13}$$
$$D_a(\chi) = \frac{a}{\chi^2+a^2\pi^2}.$$

III. CALCULATION OF THE JACOBIAN DETERMINANT

At first sight, the matrix elements φ_{jk} may appear too complicated to allow straightforward calculation of the determinant $D = \det(\varphi_{jk})$ in (12). Fortunately, the fact that we are interested in D only for asymptotically large L and $K \equiv \pi N/2L$ comes to our rescue (as it did in the solution[9] of the Bethe-ansatz equations themselves), allowing us to introduce integrals over *densities* of real χ_j in place of sums over individual χ_j.

For a given solution of (4), $\chi_1,\cdots,\chi_M$, consider the equation

$$\exp\{2\pi i\varphi(\chi;\chi_1,\cdots,\chi_M)\} = -1, \tag{14}$$

where

$$2\pi\varphi(\chi;\chi_1,\cdots,\chi_M) = i\sum_{l=1}^N \ln\left(\frac{\chi-\lambda_l+i\pi/2}{\chi-\lambda_l-i\pi/2}\right) - i\sum_{j=1}^M \ln\left(\frac{\chi-\chi_j+i\pi}{\chi-\chi_j-i\pi}\right).$$

The set of χ satisfying (14) will of course include all χ_j, $j = 1,\ldots,M$, but in addition will usually—the vacuum state being the only exception—contain other members as well, and these are known as *holes*. Unless otherwise specified, we shall concern ourselves only with holes on the real axis. Physically, a state describing n particles with rapidities $\theta_1,\ldots,\theta_n$ corresponds to a solution of (4) having precisely n holes located at $\chi = \theta_i$, $i = 1,\ldots,n$. For real χ, $\varphi(\chi;\chi_1,\cdots,\chi_M)$ is a monotone increasing function (for the states of interest) which runs through successive half-odd integers as χ runs through successive solutions (χ_j and holes) of (14). A smooth density of real χ_j and holes is thus given by

$$\rho(\chi) = \frac{d}{d\chi}\varphi(\chi;\chi_1,\cdots,\chi_M). \tag{15}$$

For holes at $\theta_1,\ldots,\theta_n$, we may also define a density of real χ_j, $j=1,\ldots,M_r$, as

$$\sigma(\chi)=\rho(\chi)-\sum_{i=1}^{n}\delta(\chi-\theta_i). \tag{16}$$

The introduction of the densities (15) and (16) simplifies the calculation of the determinant in two important ways:

(i) for real χ_j,χ_k we may now write

$$\varphi_{jk}=R_{jk}=\rho(\chi_j)\delta_{jk}+K_{jk},\quad K_{jk}=D_1(\chi_j-\chi_k). \tag{17}$$

(ii) for nonreal χ_k, the sum over real χ_j may be performed explicitly (in fact, it has already been performed in Ref. 5) by replacing $\Sigma_{j=1}^{M_r}\cdots$ by $\int d\chi\,\sigma(\chi)\cdots$.

Further simplification arises from the exponential growth, for asymptotically large L, of certain matrix elements of (φ_{jk}). To see this, we recall[5] that the nonreal χ_j can be classified according to the following scheme:

Two-string:

$$\{\chi_s,\chi_{s^*}\},\quad \chi_s=\psi_s+i(\pi/2+\tfrac{1}{2}\epsilon_s),\quad \chi_{s^*}=\chi_s^*, \tag{18}$$
$$\psi_s=\psi_s^*,\quad \epsilon_s=\epsilon_s^*,\quad \epsilon_s=O(e^{-\kappa L}),\quad \kappa>0.$$

Quartet:

$$\{\chi_q,\chi_{\tilde q},\chi_{\tilde q^*},\chi_{q^*}\},\quad \chi_q=\xi_q+i\eta_q,\quad \chi_{q^*}=\chi_q^*,$$
$$\chi_{\tilde q}=\xi_q+i(\pi-\eta_q+\epsilon_q),\quad \chi_{\tilde q^*}=\chi_{\tilde q}^*, \tag{19}$$
$$\xi_q=\xi_q^*,\eta_q=\eta_q^*,\quad \pi/2<\eta_q<\pi,|\epsilon_q|=O(e^{-\kappa' L}),$$
$$\kappa'>0.$$

Wide pair:

$$\{\chi_w,\chi_{w^*}\},\quad \chi_w=\xi_w+i\omega_w,\quad \chi_{w^*}=\chi_w^*,$$
$$\xi_w=\xi_w^*,\quad \omega_w=\omega_w^*,\quad \omega_w\geqslant\pi. \tag{20}$$

From (13), (18), and (19) we see that the following matrix elements (and only those) have exponential growth for large L (here s and q are used as generic indices for a two-string and a quartet, respectively):

$$\varphi_{ss}\sim(1/2\pi\epsilon_s)\sim\varphi_{s^*s^*},$$
$$\varphi_{ss^*}=\varphi_{s^*s}\sim-(1/2\pi\epsilon_s), \tag{21}$$
$$\varphi_{qq}\sim(1/2\pi\epsilon_q^*)\sim\varphi_{\tilde q^*\tilde q^*},\quad \varphi_{q^*q^*}\sim(1/2\pi\epsilon_q)\sim\varphi_{\tilde q\tilde q},$$
$$\varphi_{q\tilde q^*}=\varphi_{\tilde q^*q}\sim-(1/2\pi\epsilon_q^*),$$
$$\varphi_{q^*\tilde q}=\varphi_{\tilde q q^*}\sim-(1/2\pi\epsilon_q).$$

The $1/\epsilon$ singularities of $\det(\varphi_{jk})$ for $L\to\infty$ can be isolated by noting

$$\det(\varphi_{jk})=\det(\varphi'_{jk}), \tag{22}$$

where (φ'_{jk}) is obtained from (φ_{jk}) by the following determinant-preserving manipulations: for each two-string $\{\chi_s,\chi_{s^*}\}$ add row s^* to row s and column s^* to column s; for each quartet $\{\chi_q,\chi_{\tilde q},\chi_{\tilde q^*},\chi_{q^*}\}$, add row (column) $\tilde q^*$ to row (column) q, and row (column) $\tilde q$ to row (column) q^*. The only remaining $O(1/\epsilon)$ elements of (φ'_{jk}) are $\varphi'_{s^*s^*}$, $\varphi'_{\tilde q\tilde q},\varphi'_{\tilde q^*\tilde q^*}$ (equal, respectively, to their unprimed counterparts). With an exponentially small error, these quantities may be factored out of the determinant to yield

$$D=\prod_s\left(\frac{1}{2\pi\epsilon_s}\right)\prod_q\left(\frac{1}{4\pi^2|\epsilon_q|^2}\right)\det(\varphi''_{ij})\times(1+O(e^{-\kappa'' L})),$$
$$\kappa''>0, \tag{23}$$

where (φ''_{ij}) is (φ'_{ij}) with rows and columns labeled by $s^*,\tilde q,\tilde q^*$ (but not s,q,q^*) deleted. We will completely neglect the $O(e^{-\kappa L})$ corrections in Eq. (23), because, as we shall see, to each pole in (23) there corresponds an equivalent zero coming from $[\Pi_{j\neq k}^{M}a(\chi_j-\chi_k)]^{-1}$ in (12). (φ''_{ij}) has the structure

$$\begin{array}{c} \\ \text{real} \\ sqq^*ww^* \end{array}\begin{array}{c} \begin{array}{cc} \text{real}' & s'q'q'^*w'w'^* \end{array} \\ \begin{pmatrix} R & U \\ U^T & A \end{pmatrix} \end{array}$$

with the matrix R of the real χ_j given in (17) and the matrices U and A assigned matrix elements as follows (for convenience we introduce the notation $\hat\chi=\chi-i\pi/2$ for $\operatorname{Im}\chi\geqslant\pi/2$):

χ	χ'	$U_{\chi\chi'}$ or $A_{\chi\chi'}$
real	χ_s,χ_q,χ_{q^*}	$D_{3/2}(\chi-\hat\chi')+D_{1/2}(\chi-\hat\chi')$
real	χ_w,χ_{w^*}	$D_1(\chi-\chi')$
χ_s,χ_q,χ_{q^*}	$\chi_{s'},\chi_{q'},\chi_{q'^*}(\neq\chi)$	$2D_1(\hat\chi-\hat\chi')+D_2(\hat\chi-\hat\chi')$
χ_s,χ_q,χ_{q^*}	χ_w,χ_{w^*}	$D_1(\hat\chi-\chi'+i\pi/2)=D_1(\hat\chi-\chi'-i\pi/2)$
χ_w,χ_{w^*}	$\chi_{w'},\chi_{w'^*}(\neq\chi)$	$D_1(\chi-\chi')$
χ_w,χ_{w^*}	χ	$\left[\frac{\partial}{\partial z}\varphi(z;\{\chi_j\})+D_1(z-\chi)\right]_{z=\chi}$
χ_s,χ_q,χ_{q^*}	χ	$\left[\frac{\partial}{\partial z}(\varphi(z+i\pi/2;\{\chi_j\})+\varphi(z-i\pi/2;\{\chi_j\}))\right.$ $\left.+D_1(z-\hat\chi+i\pi)+2D_1(z-\hat\chi)+D_1(z-\hat\chi-i\pi)\right]_{z=\chi}$

(24)

The expressions for $\varphi(z;\{\chi_j\})$ in the Eq. (24) may be evaluated using the formulas of Ref. 5. The calculation is straightforward and we omit the details. The resulting expression for the diagonal elements of A is remarkably simple when expressed in terms of the parameters $\hat\chi_j$ and the holes $\theta_1,\ldots,\theta_n$:

$$A_{\chi\chi}=\sum_{h=1}^{n}D_{1/2}(\theta_h-\hat\chi)-\sum_{\hat\chi_j\neq\hat\chi}D_1(\hat\chi_j-\hat\chi)-(3/2\pi^2)$$
$$\times\theta(\pi/2-|\operatorname{Im}\hat\chi|), \tag{25}$$

where

$$\theta(x)=\begin{cases}1, & x>0\\ 0, & x\leqslant 0\end{cases}.$$

To proceed with the calculation note that

$$\det\begin{pmatrix}R & U\\ U^T & A\end{pmatrix}=\det\left[\begin{pmatrix}R & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ U^T & 1\end{pmatrix}\begin{pmatrix}1 & R^{-1}U\\ 0 & A-U^TR^{-1}U\end{pmatrix}\right]$$
$$=\det R\cdot\det(A-B),\quad B=U^TR^{-1}U. \tag{26}$$

Since

$R^{-1}=\rho^{-1/2}(1+\rho^{-1/2}K\rho^{-1/2})^{-1}\rho^{-1/2}$, $\rho_{ij}=\rho(\chi_j)\delta_{ij}$, we have [setting $U_{\chi\chi_j}=u_j(\chi)$]

$$B_{ij}=\int d\chi\,\frac{\sigma(\chi)}{\rho(\chi)}u_i(\chi)u_j(\chi)-\iint d\chi_1 d\chi_2\,\frac{\sigma(\chi_1)}{\rho(\chi_1)}\frac{\sigma(\chi_2)}{\rho(\chi_2)}\times u_i(\chi_1)K(\chi_1-\chi_2)u_j(\chi_2)+\cdots$$

$$=\iint d\chi d\chi'\,u_i(\chi)(1+K)^{-1}(\chi-\chi')u_j(\chi')+\iiint d\chi d\chi' d\chi''\,u_i(\chi)(1+K)^{-1}(\chi-\chi')(\sigma_h(\chi')/\rho(\chi'))(1+K)^{-1}(\chi'-\chi'')u_j(\chi'')+O\left(\frac{\sigma_h^2}{\rho^2}\right), \quad (27)$$

where $\sigma_h=\rho-\sigma$ is the density of holes. But for an n-particle state,

$$\frac{\sigma_h(\chi)}{\rho(\chi)}=\sum_{h=1}^{n}\frac{1}{\rho(\theta_h)}\delta(\chi-\theta_h),$$

with $\rho(\theta_h)^{-1}=O(L^{-1})$, and hence we need keep only the first term in the expansion (27) in powers of σ_h/ρ. Inserting the expressions for elements of U given in (24), we obtain, for example, for $\chi,\chi'=\chi_s,\chi_q,\chi_{q^*}$,

$$B_{\chi\chi'}=D_1(\hat\chi-\hat\chi')+D_2(\hat\chi-\hat\chi') \quad (28)$$

and hence, from (24) and (28),

$$(A-B)_{\chi\chi}=\sum_{h=1}^{n}D_{1/2}(\theta_h-\hat\chi)-\sum_{\hat\chi_j\neq\hat\chi}D_1(\hat\chi-\hat\chi_j),$$

$$(A-B)_{\chi\chi'}=D_1(\hat\chi-\hat\chi'),\quad \chi\neq\chi'. \quad (29)$$

Calculating the remaining matrix elements, (i.e., those involving at least one wide pair), one finds that *all* matrix elements of $A-B$ are of the form (29).

The result (29) has a most satisfying interpretation. In Ref. 5 we found that the introduction of the density σ into the Bethe-ansatz equations (4) led, upon integration, to a set of "higher level Bethe-ansatz" equations,

$$\hat\varphi_j(\hat\chi_1,\cdots,\hat\chi_{\hat M})=\text{integer},\quad j=1,\ldots,\hat M,$$

where

$$2\pi\hat\varphi_j(\hat\chi_1,\ldots,\hat\chi_{\hat M})=i\sum_{h=1}^{n}\ln\left(\frac{\hat\chi_j-\theta_h+i\pi/2}{\hat\chi_j-\theta_h-i\pi/2}\right)-i\sum_{\substack{l=1\\ l\neq j}}^{\hat M}\ln\left(\frac{\hat\chi_j-\hat\chi_l+i\pi}{\hat\chi_j-\hat\chi_l-i\pi}\right). \quad (30)$$

We see from (29) that $\det(A-B)$ is nothing but the Jacobian determinant of these equations.

To evaluate the first factor in (26), we write

$$\det R=\prod_{i=1}^{M_r}\rho(\chi_i)\exp[\operatorname{Tr}\ln(1+\rho^{-1/2}K\rho^{-1/2})]$$

$$\operatorname{Tr}\ln(1+\rho^{-1/2}K\rho^{-1/2})=\sum_{i=1}^{M_r}\rho(\chi_i)^{-1}\Big(D_1(0)-\frac{1}{2}\int d\chi D_1(\chi_i-\chi)\frac{\sigma(\chi)}{\rho(\chi)}D_1(\chi-\chi_i)+\cdots\Big) \quad (31)$$

$$=\left(\sum_{i=1}^{M_r}\rho(\chi_i)^{-1}\right)\sum_{n=1}^{\infty}(-)^{n+1}(\pi n)^{-2}=\left(\sum_i\rho(\chi_i)^{-1}\right)\cdot\frac{1}{12}.$$

To arrive at the last line, we have used

$$\frac{\sigma(\chi)}{\rho(\chi)}=1+O(L^{-1}),$$

$$\int d\chi''D_a(\chi-\chi'')D_b(\chi''-\chi')=D_{a+b}(\chi-\chi'),$$

$$D_a(0)=(1/\pi^2 a).$$

The treatment of $\Sigma_{i=1}^{M_r}\rho(\chi_i)^{-1}$ is a bit tricky, since insertion of $\int d\chi\sigma(\chi)\cdots$ in place of Σ_i is not justified here (it leads to a linearly divergent integral). However, we can write

$$\sum_{i=1}^{M_r}\rho(\chi_i)^{-1}=\int_{-\mu}^{\mu}d\chi\sigma(\chi)\rho(\chi)^{-1}+\rho(\chi_1)^{-1}+\rho(\chi_{M_r})^{-1} \quad (32)$$

with μ determined by

$$\int_{\mu}^{\infty}d\chi\sigma(\chi)=1 \quad (33)$$

and $\rho(\chi_1)^{-1},\rho(\chi_{M_r})^{-1}$ are of order unity. Inserting the asymptotic behavior of $\sigma(\chi)$ for $\chi\geqslant\mu$ (see Refs. 1 and 6),

$$\sigma(\chi)\simeq(N/2\pi)e^{\pi/c-\chi}+h(\chi),\quad |h(\chi)|<h_0\chi^{-2},\quad h_0=O(1),$$

we obtain

$$\mu=(\pi/c)+\ln N+O(1)$$

and hence

$$\sum_{i=1}^{M_r}\rho(\chi_i)^{-1}=2\left(\frac{\pi}{c}+\ln N\right)+O(1). \quad (34)$$

Thus, using the renormalization rule

$$mL=2Ne^{-\pi/c},$$

$$\det R=\left(\prod_{i=1}^{M_r}\rho(\chi_i)\right)\left(\frac{2N^2}{mL}\right)^{1/6}\cdot A_0,$$

where A_0 is a cutoff independent numerical factor.

IV. EVALUATION OF THE REMAINING FACTORS

Referring back to the full norm formula (12), we now turn to the problem of calculating

$$\ln\prod_{j\neq k}a(\chi_j-\chi_k)=\sum_{\nu=1}^{3}\sum_{j\neq k}^{(\nu)}\frac{1}{2}\ln\left(\frac{(\chi_j-\chi_k)^2}{\pi^2+(\chi_j-\chi_k)^2}\right), \quad (35)$$

where we break up the double summation into three sets of terms: $\nu=1$ for both χ_j and χ_k real; $\nu=2$ for χ_j real and χ_k complex, or vice versa; and $\nu=3$ for both χ_j and χ_k complex.

For $\nu=1$ (real-real case), one might try to replace the summation over j and k in (35) by

$$\iint d\chi d\chi'\,\sigma(\chi)\sigma(\chi')\ln a(\chi-\chi')-\int d\chi\,\sigma(\chi)\ln a(0). \quad (36)$$

Unfortunately, both terms in (36) are meaningless; the sec-

ond because of the strictly infinite $\ln a(0)$, and the first because it contains a piece $\Sigma_h \sigma(\theta_h)^2 \ln a(0)$ (sum over holes θ_h), which is again infinite. We can, however, use the densities in a more judicious fashion, first writing

$$\sum_{j\neq k}^{(1)} \ln(\chi_j-\chi_k)^2 = \sum_{j=1}^{M_r}\Big\{\int d\chi\,\rho(\chi)\ln(\chi_j-\chi)^2 - \sum_{h=1}^{n}\ln(\chi_j-\theta_h)^2 - \int_{\chi_j-\delta_j}^{\chi_j+\delta_j} d\rho(\chi)\ln(\chi_j-\chi)^2\Big\},$$
$$\delta_j = \tfrac{1}{2}\rho(\chi_j)^{-1}. \tag{37}$$

Except near $\chi = \pm\infty, \delta_j = O(L^{-1})$ and we can evaluate the last term in brackets, obtaining

$$\int_{\chi_j-\delta_j}^{\chi_j+\delta_j} d\chi\rho(\chi)\ln(\chi_j-\chi)^2 = \rho(\chi_j)\int_{-\delta_j}^{\delta_j} dx\,\ln x^2 + O(L^{-1}) = 2\rho(\chi_j)\ln 2e + O(L^{-1}). \tag{38}$$

Adopting this expression for all χ_j introduces an overall $O(1)$ error which, like that in (34), depends only on the asymptotic form of $\sigma_0(\chi)$ and not on the parameters $\theta_i, \hat{\chi}_j$ or the cutoffs N.

If the sum over j in (37) is now replaced by an integral, the first and third terms in brackets give no problem, but the second term must be treated by the same procedure we used to derive (37) and (38). The result is

$$\sum_{j\neq k}^{(1)} \ln a(\chi-\chi_k) = \iint d\chi\,d\chi'\rho(\chi)\rho(\chi')\frac{1}{2}\ln\frac{(\chi-\chi')^2}{(\chi-\chi')^2+\pi^2} + \ln\frac{\Pi_{i=1}^{M_r}\rho(\chi_i)}{\Pi_{h=1}^{n}\rho(\theta_h)} + \sum_{h\neq h'}^{n}\ln a(\theta_h-\theta_{h'}) - \int d\chi\,\rho(\chi)\sum_{h=1}^{n}\ln\frac{(\chi-\theta_h)^2}{(\chi-\theta_h)^2+\pi^2} + (M_r-n)\ln(2\pi e) + O(1)$$
$$= \int_0^\infty \frac{dp}{p}(1-e^{-\pi p})\Big\{\sum_{h=1}^{n}(\tilde{\rho}(p)e^{ip\theta_h}+\tilde{\rho}(-p)e^{-ip\theta_h}) - \tilde{\rho}(p)\tilde{\rho}(-p)\Big\} + \ln\Big\{\prod_{i=1}^{M_r}\rho(\chi_i)\prod_{h=1}^{n}\rho(\theta_h)^{-1}\Big\} + (M_r-n)\ln(2\pi e) + O(1), \tag{39}$$

where $\tilde{\rho}(p)$, the Fourier transform of $\rho(\chi)$, is given explicitly by[9]

$$\tilde{\rho}(p) = \tilde{\sigma}(p) + \sum_{h=1}^{n} e^{-ip\theta_h} = \tilde{\rho}_1(p) + \tilde{\rho}_2(p),$$
$$\tilde{\rho}_1(p) = [2\cosh(\pi p/2)]^{-1}\Big\{N_+ e^{-ip\pi/c} + N_- e^{ip\pi/c} + \sum_{h=1}^{n} e^{-|p|\pi/2 - ip\theta_h}\Big\}$$
$$\tilde{\rho}_2(p) = -\sum_s e^{-|p|\pi/2 - ip\psi_s} - \sum_q [e^{-|p|\eta_q} + e^{-|p|(\pi-\eta_q)}]e^{-ip\xi_q} - \sum_w [e^{-|p|\omega_w} - e^{-|p|(\omega_w-\pi)}]e^{-ip\zeta_w}. \tag{40}$$

In contrast to the subtleties of the $\nu = 1$ case, the expression for the $\nu = 2$ (real-complex) summation in (35) in terms of densities is obtained in a completely straightforward manner. One simply replaces the sum over real χ_j by $\int d\chi\,\sigma(\chi)\cdots$, then groups the terms according to the two-strings, quartets, and wide pairs. This gives

$$\sum_{j,k}^{(2)} \ln a(\chi_j-\chi_k) = \int_0^\infty \frac{dp}{p}\{\tilde{\sigma}(p)\tilde{\rho}_2(-p) + \tilde{\sigma}(-p)\tilde{\rho}_2(p)\}(1-e^{-\pi p}), \tag{41}$$

which may be combined with (39) to yield

$$\sum_{\nu=1}^{2}\sum_{j\neq k}^{(\nu)} \ln a(\chi_j-\chi_k) = \sum_{h\neq h'}^{n}\ln a(\theta_h-\theta_{h'}) + \int_0^\infty \frac{dp}{p}(1-e^{-\pi p})\Big\{\sum_{h=1}^{n}[\tilde{\rho}_1(p)e^{ip\theta_h} + \tilde{\rho}_1(-p)e^{-ip\theta_h}] - \tilde{\rho}_1(p)\tilde{\rho}_1(-p) + \tilde{\rho}_2(p)\tilde{\rho}_2(-p)\Big\} + \ln\Big\{B_0(2\pi e)^{M_r-n}\prod_{i=1}^{M_r}\rho(\chi_i)\prod_{h=1}^{n}\rho(\theta_h)^{-1}\Big\}, \tag{42}$$

where B_0 is a pure number of order unity.

To complete this phase of the calculation, we insert the expressions (40) for $\tilde{\rho}_1, \tilde{\rho}_2$ into (42) and perform as many of the integrals as possible. Combining the result with the $\nu = 3$ (complex-complex) contribution, we obtain a result whose dependence on the parameters $\hat{\chi}_j$ is particularly simple:

$$\prod_{j\neq k}^{M} a(\chi_j-\chi_k)^{-1}=B_0(2\pi e)^{-M_r+n}\prod_{i=1}^{M_r}\rho(\chi_i)^{-1}\prod_{h=1}^{n}\left(\frac{mL}{2\pi}\cosh\theta_h\right)\cdot G(\boldsymbol{\theta})\cdot\prod_{j\neq k}^{\hat{M}} a(\hat{\chi}_j-\hat{\chi}_k)^{-1}\prod_s\frac{\epsilon_s}{\pi}\prod_q\frac{|\epsilon_q|^2}{\pi^2}\cdot(1+O(L^{-1})), \tag{43}$$

where

$$\ln G(\boldsymbol{\theta})=\ln G(\theta_1,\ldots,\theta_n)=\frac{1}{2}\sum_{\alpha,\beta=\pm}N_\alpha N_\beta I\left(\frac{\pi}{2},\lambda_\alpha-\lambda_\beta\right)-\sum_{\alpha=\pm}\sum_{h=1}^{n}N_\alpha I(0,\lambda_\alpha-\theta_h)$$
$$+\sum_{h<h'}^{n}\left[2I\left(\frac{\pi}{2},\theta_h-\theta_{h'}\right)+I((3\pi/2),\theta_h-\theta_{h'})\right]; \tag{44}$$

$$I(\mu,x)=\int_0^\infty\frac{dp}{p}\frac{\sinh(p\pi/2)}{\cosh^2(p\pi/2)}e^{-\mu p}\cos px.$$

V. FINAL EXPRESSION FOR THE NORM

Putting together the above calculations, we obtain, finally, a formula for the norm of our Bethe-ansatz basis state, in the limit $L, K\to\infty$:

$$\langle F|F\rangle=(\Phi,\Phi)=\pi^{\hat{M}}\left[\prod_{j\neq k}a(\hat{\chi}_j-\hat{\chi}_k)\right]^{-1}\det(2\pi\hat{\varphi}_{jk})\langle\boldsymbol{\theta}|\boldsymbol{\theta}\rangle, \tag{45}$$

where

$$\langle\boldsymbol{\theta}|\boldsymbol{\theta}\rangle=C_0\left(\frac{2N^2}{mL}\right)^{1/6}\left(\frac{\pi}{e}\right)^{M_r-n}G(\boldsymbol{\theta})\prod_{h=1}^{n}(\pi Lm\cosh\theta_h)$$
$$\times\prod_{h\neq h'}^{n}a(\theta_h-\theta_{h'})^{-1},$$

with $G(\boldsymbol{\theta})$ given by (44) and C_0 a purely numerical factor common to all states. The result (45) agrees with the hypothesis, put forward in Ref. 9, that each n-massive particle sector is characterized by an asymptotic higher Bethe-ansatz structure. $|\boldsymbol{\theta}\rangle$ is the reference state, the one with highest color ($S=n/2$, i.e., no parameter $\hat{\chi}$), in each of these sectors. Its norm, given by (46), correctly depends on $\boldsymbol{\theta}\equiv\{\theta_1,\cdots,\theta_n\}$ and the cutoffs L and K, and could be renormalized to one for the whole sector; all other states of the sector have the norm given in (45), identical in structure to (12).

[1] N. Andrei and J. H. Lowenstein, Phys. Rev. Lett. **43**, 1698 (1979); Phys. Lett. B **90**, 106 (1980).
[2] A. A. Belavin, Phys. Lett. B **87**, 117 (1979).
[3] N. Andrei and J. H. Lowenstein, Phys. Lett. B **91**, 401 (1980).
[4] V. E. Korepin, Commun. Math. Phys. **86**, 391 (1982).
[5] C. Destri and J. H. Lowenstein, Nucl. Phys. B **205** (FS5), 369 (1982).
[6] J. H. Lowenstein, Surv. High Energy Phys. **2**, 207 (1981).
[7] C. N. Yang, Phys. Rev. Lett. **19**, 1312 (1967).
[8] P. P. Kulish and N. Yu. Reshetikhin, Zh. Eksp. Teor. Fiz. **80**, 214 (1981) [Sov. Phys. JETP **53**, 108 (1981)] and references cited therein.
[9] M. Gaudin, B. M. McCoy, and T. T. Wu, Phys. Rev. D **23**, 417 (1981).

Coulomb-modified nuclear scattering. II

B. Talukdar and S. Saha
Department of Physics, Visva-Bharati University, Santiniketan 731235, West Bengal, India
T. Sasakawa
Department of Physics, Tohoku University, Sendai 980, Japan

(Received 15 October 1981; accepted for publication 23 September 1982)

Exact analytical expressions for the off-shell Jost function $f_l(k, q)$ for the Coulomb and Coulomb-distorted nuclear potentials are constructed in terms of the sum of Gaussian hypergeometric functions by using an integral representation for $f_l(k, q)$ given by Fuda. It is pointed out that the hypergeometric functions which occur here can be generated by the use of a well-known three-term recurrence relation. With van Haeringen our expressions for $f_l(k, q)$ exhibit discontinuity in the on-shell limit $q \to k$. This discontinuous behavior has been attributed to the long range nature of the Coulomb force.

PACS numbers: 21.30. + y, 21.10.Sf, 02.30. − f

1. INTRODUCTION

Based on a coordinate space approach[1] to the T-matrix, Fuda and Whiting[2] have introduced a generalization of the Jost function.[3] This is often referred to as the off-shell Jost function. The half off-shell T-matrix element can be expressed directly in terms of the off-shell Jost function. Also by exploiting the relations which exist between the fully off-shell T-matrix elements and half-off shell ones, we can write the off-shell T-matrix in terms of the off-shell Jost function. Thus it is of some importance to have in the literature a relatively noncomplicated mathematical prescription to derive an expression for the off-shell Jost function relating to scattering by Coulomb distorted nuclear potentials which are encountered in physical processes like the proton-proton (p-p) bremsstrahlung and p-$2p$ reaction. The present paper is an effort in this direction.

For a central potential $V(r)$ having a short range, the off-shell Jost function $f_l(k, q)$ is obtained from the solution of an inhomogeneous differential equation[2]

$$\left[k^2 + \frac{d^2}{dr^2} - \frac{l(l+1)}{r^2} - V(r)\right] f_l(k, q, r) = (k^2 - q^2) e^{(1/2)il\pi} w_l^{(+)}(qr) \tag{1}$$

by using

$$f_l(k, q) = \frac{q^l e^{-il\pi/2}(2l+1)}{(2l+1)!!} \lim_{r \to 0} r^l f_l(k, q, r). \tag{2}$$

Here we have two momenta $\vec{k}$ and $\vec{q}$, the on-shell momentum $\vec{k}$ is related to the energy $E = k^2$ while $\vec{q}$ is an off-shell momentum. We work in units in which $\hbar^2/2m$ is unity. Note that $w_l^{(+)}(kr)$ is a Reccati–Hankel function in the phase convention of Messiah.[4]

Fuda[5] has introduced two integral representations for $f_l(k, q)$, one of which connects $f_l(k, q)$ with a function that arises in Kowalski's generalization of the Sasakawa's theory[6] and the other that expresses $f_l(k, q)$ in terms of the regular solution of the Schrödinger equation.

The later integral representation is given by

$$f_l(k, q) = 1 + \frac{q^l}{(2l+1)!!} \int_0^\infty dr\, qr\, h_l^{(+)}(qr) V(r) \phi_l(k, r), \tag{3}$$

where $\phi_l(k, r)$ is the regular solution of the radial Schrödinger equation, i.e., of (1) when $q = k$. The functions $w_l^{(+)}(qr)$ and $h_l^{(+)}(qr)$ are related as

$$w_l^{(+)}(qr) = qr h_l^{(+)}(qr) = i \sqrt{\frac{\pi qr}{2}}\, H^{(1)}_{l+1/2}(qr). \tag{4}$$

The Hankel function $H^{(1)}_{l+1/2}(qr)$ of order half-odd integer can be expanded as

$$H^{(1)}_{l+1/2}(qr) = i^{-l} \left(\frac{8qr}{\pi}\right)^{1/2} \sum_{j=0}^{l} \frac{(-1)^j (l+j)!}{j!(l-j)!} \frac{e^{iqr}}{(2iqr)^{j+1}}. \tag{5}$$

One of the important virtues of (3) is that it can be used both for short and long range potentials.[7] In general, (3) will facilitate construction of expressions for $f_l(k, q)$ whenever the associated radial Schrödinger equation is analytically solvable. Here we deal with the Coulomb and Coulomb plus some useful separable nucleon-nucleon potentials. For the pure Coulomb case we have

$$V(r) = V_c(r) = \frac{2\eta k}{r}, \tag{6}$$

with η the Sommerfeld parameter. However, for the Coulomb plus, a rank n separable potential $V(r)$ will be replaced by

$$V_c(r) + \Sigma_{i=1}^n \lambda_i v_l^{(i)}(r) v_l^{(i)}(r')$$

and in this case the integral representation for $f_l(k, q)$ will read

$$f_l(k, q) = 1 + \frac{q^l}{(2l+1)!!} \int_0^\infty qr\, h_l^{(+)}(qr) \left[V_c(r) \phi_l^{cs}(k, r) + \sum_{i=1}^n \lambda_i v_l^{(i)}(r) \int_0^\infty v_l^{(i)}(s) \phi_l^{cs}(k, s) ds \right] dr, \tag{7}$$

where $\phi_l^{cs}(k, r)$ is the regular solution of the Schrödinger equation for the Coulomb plus separable potentials under consideration. We shall use $\phi_l^c(k, r)$ for the Coulomb regular solution. The superscripts c and cs will have similar meanings throughout this article.

In the next section we employ (3) to construct an expression for the l-wave Coulomb off-shell Jost function in terms

 0022-2488/83/030683-04$02.50

of Gaussian hypergeometric functions ${}_2F_1$. This result is already known in the literature. However, generalization of the ${}_2F_1$ functions in the expressions for $f_l^c(k, q)$ without using series summation technique is a nontrivial problem.[7] We achieve this by the use of a three-term recursion relation.[8]

The equation[7]

$$\lim_{q\to k} \omega f_l^c(k, q) = f_l^c(k) \tag{8}$$

has been found to hold good for the Coulomb case. Here $f_l^c(k)$ is the on-shell Coulomb Jost function and

$$\omega = \left(\frac{q-k}{q+k}\right)^{i\eta} \frac{e^{\pi\eta/2}}{\Gamma(1+i\eta)}. \tag{9}$$

We have checked our results for the criterion in (8).

Based on (7), in Sec. 3 we present a similar treatment for (i) Coulomb plus Yamaguchi and (ii) Coulomb plus Mongan case IV potentials[9,10] by using expressions for the regular wave functions derived in the previous paper[11] (hereafter referred to as paper I). For simplicity of presentation we consider the s-wave case only. The results for $f_0^{cs}(k, q)$ appear somewhat complicated. But the ${}_2F_1$ function involved can still be generated by the use of the three-term recursion relation noted earlier. Finally, in Sec. 4 we present some concluding remarks.

2. COULOMB CASE

For the Coulomb potential in (6) the regular solution of the Schrödinger equation is given by[12]

$$\phi_l^c(k, r) = r^{l+1} e^{ikr} {}_1F_1(l+1+i\eta, 2l+2; -2ikr), \tag{10}$$

where the confluent hypergeometric function

$${}_1F_1(a, c; z) = \frac{\Gamma(c)}{\Gamma(a)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)}{\Gamma(c+n)} \frac{z^n}{n!}. \tag{11}$$

Combining (3)–(6) and (10) we get

$$f_l^c(k, q) = 1 + 2\eta k \frac{q^l}{(2l+1)!!} \sum_{j=0}^{l} \frac{i^{j-l}(l+j)!}{j!(l-j)!} \frac{1}{(2q)^j} \int_0^\infty r^{j-l} e^{i(k+q)r} {}_1F_1(l+1+i\eta, 2l+2; -2ikr)\, dr. \tag{12}$$

In order to transform the integral in (12) into more useful forms, it is necessary at this point to introduce explicitly into the integral a convergence factor e^{-sr}. For all values of s such that Re $s \to 0$, the resulting integral is then uniformly convergent. The limit $s \to 0$ will be taken in the final answer and it is this limiting operation that introduces the discontinuous behavior[7] of $f_l^c(k, q)$ at $q = k$. When the integration in (12) is thus performed we arrive at

$$f_l^c(k, q) = 1 + \frac{q^l}{(2l+1)!!} \frac{2\eta k}{(k+q)^{l+1}} \sum_{j=0}^{l} \frac{(l+j)!}{j!} \frac{(k+q)^j}{2q} \times {}_2F_1\left(l+1+i\eta, l-j+1; 2l+2; \frac{2k}{k+q}\right), \tag{13}$$

where the Gaussian hypergeometric function

$${}_2F_1(a, p; b; z) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(p)} \sum_{n=0}^{\infty} \frac{\Gamma(n+a)\Gamma(n+p)}{\Gamma(n+b)} \frac{z^n}{n!}. \tag{14}$$

Evaluation of the closed form expression for $f_l^c(k, q)$ for a particular l requires the generation of a set of hypergeometric functions which differ only in p-values with fixed values of a and b. Note that the parameter a is complex while b is a real integer. As one moves from one partial wave to another, a new set of ${}_2F_1$ functions is needed to construct the expression for $f_l^c(k, q)$. Each set is determined by its value of a and b. For a fixed l, the functions ${}_2F_1$ which occur in (13) can be generated by solving a set of simultaneous equations resulting from the recursion relation[8] for the hypergeometric function,

$$\begin{aligned} &p(1-x){}_2F_1(a, p+1; b; x) \\ &\quad + [(a-p)(1-x) + (b-p-a)]{}_2F_1(a, p; b; x) \\ &\quad + (p-b){}_2F_1(a, p-1; b; x) = 0 \end{aligned} \tag{15}$$

with the known special values

$${}_2F_1(a, o; b; x) = 1 \tag{16a}$$

and

$${}_2F_1(a, b; b; x) = (1-x)^{-a}. \tag{16b}$$

In the following we specialize (13) to a few lower partial waves and demonstrate that our limiting procedure in writing this equation has already introduced the discontinuous behavior of $f_l^c(k, q)$ at $q = k$.

Considering the s-wave case we get a very simple expression for $f_0^c(k, q)$, namely,

$$f_0^c(k, q) = 1 + 2\eta k/(k+q)i\, {}_2F_1(1+i\eta, 1; 2; 2k/(k+q)). \tag{17}$$

Using (15), (16a), and (16b), (17) becomes

$$f_0^c(k, q) = \left(\frac{q+k}{q-k}\right)^{i\eta}. \tag{18}$$

Clearly, at $q = k$, $f_0^c(k, q)$ is discontinuous. This behavior was first observed by van Haeringen.[7] The discontinuity would disappear if we have chosen to work with a screened Coulomb potential. Note that the appropriate wave function $\phi_0^c(k, r)$ for this case has been given by one of us.[13]

Using the approach described above we can also calculate expressions for higher partial wave off-shell Coulomb Jost functions. Interestingly, these results will be in exact agreement with those given by van Haeringen.[7]

3. COULOMB-LIKE POTENTIALS

A. The Coulomb plus Yamaguchi potential

The s-wave regular solution $\phi_0^{cs}(k, r)$ of the Schrödinger equation for the Coulomb plus Yamaguchi potential has been given in (40) of paper I, where we did not use the superscript cs on ϕ_0. The subscript $l = 0$ was also omitted. We now specialize (7) to the rank one case, using Yamaguchi form factors and the expression for $\phi_0^{cs}(k, r)$ just noted. This leads

us to

$$f_0^{cs}(k,q) = \left(\frac{q+k}{q-k}\right)^{i\eta} + \frac{\lambda}{(\alpha - iq)(\alpha^2+k^2)}\left(\frac{\alpha - ik}{\alpha + ik}\right)^{i\eta} - \lambda \frac{\eta e^{2\eta\theta}}{(\alpha^2+k^2)(k+q)D^{cs}(k)} \times \sum_{n=1}^{\infty} \frac{\rho^{n-1}}{(n+1)}\left(\frac{2k}{k+q}\right)^n \times {}_2F_1\left(1, n+1+i\eta; n+2; \frac{2k}{k+q}\right) - \lambda^2 \frac{e^{2\eta\theta}}{(\alpha - iq)(\alpha - ik)^2(\alpha^2+k^2)D^{cs}(k)2ik} \times \sum_{n=1}^{\infty} \rho^{n-1}\left(\frac{-2ik}{\alpha - ik}\right)^n {}_2F_1 \times \left(1, n+1+i\eta; n+1; \frac{-2ik}{\alpha - ik}\right), \quad (19)$$

with

$$\rho = [(\alpha + ik)/2ik], \quad (20)$$

$$\theta = \tan^{-1} k/\alpha, \quad (21)$$

and

$$D^{cs}(k) = 1 + \frac{\lambda}{(\alpha^2+k^2)(\alpha - ik)} \sum_{n=1}^{\infty} (-1)^n \left(\frac{\alpha + ik}{\alpha - ik}\right)^n \times {}_2F_1\left(1, 1+n+i\eta; n+1, \frac{-2ik}{\alpha - ik}\right). \quad (22)$$

A couple of useful checks can be made on the fairly complicated expression (19) for the Coulomb modified Yamaguchi off-shell Jost function. For example, in absence of the Yamaguchi potential, i.e., for $\lambda = 0, f_0^{cs}(k,q)$ goes to the pure Coulomb Jost function $f_0^c(k,q)$ given in (18). Secondly, in the limit of no Coulomb field, $\eta \to 0$, the first term becomes unity, the second term takes up a simple value, $\lambda / (\alpha - iq)(\alpha^2 + k^2)$, the third term vanishes altogether, and the fourth term is modified as

$$\frac{\lambda^2}{(\alpha - iq)(\alpha - ik)(\alpha^2+k^2)^2 D(k)} \sum_{n=0}^{\infty} (-1)^n \left(\frac{\alpha + ik}{\alpha - ik}\right)^n, \quad (23)$$

where

$$D(k) = 1 - \frac{\lambda}{(\alpha^2+k^2)(\alpha - ik)} \sum_{n=0}^{\infty} (-1)^n \left(\frac{\alpha + ik}{\alpha - ik}\right)^n, \quad (24)$$

the value of $D^{cs}(k)$ for $\eta = 0$. In writing (23) and (24) we have made use of (16b). We have also used the value of ρ and changed the summation appropriately. Interestingly, the summations in (23) and (24) represent the binomial series giving

$$\sum_{n=0}^{\infty} (-1)^n \left(\frac{\alpha + ik}{\alpha - ik}\right)^n = \frac{\alpha - ik}{2\alpha}. \quad (25)$$

Equations (23)–(25) facilitate evaluation of the limit $f_0^{cs}(k,q)$ for the no Coulomb case and we arrive at the expression for the Yamaguchi off-shell Jost function.[14] It is of interest to remark that all Gaussian hypergeometric functions which parametrize (19) and (22) can be generated by the use of (15), since ${}_2F_1(\alpha, \beta; \gamma; x) = {}_2F_1(\beta, \alpha; \gamma; x)$.

B. The Coulomb plus Mongan case IV potential

The s-wave regular solution $\phi_0^{cs}(k,r)$ of the Schrödinger equation for the Coulomb plus rank-two Mongan case IV potential as given in paper I can be written in the form

$$\phi_0^{cs}(k,r) = re^{ikr} {}_1F_1(1+i\eta, 2; -2ikr) - \frac{1}{2ik} re^{ikr} \sum_{n=1}^{\infty} \frac{\theta_n(1+i\eta, 2; -2ikr)}{(n-1)!} \times \{d_1(k)\rho_1^{n-1} + d_2(k)\rho_2^{n-1}\}. \quad (26)$$

All functions occurring in the right side of (26) are given in Ref. 11. For example, θ_n is related to the generalized hypergeometric function ${}_2F_2$. The quantities d_r's are given in (52)–(54). Using (26) in the special form of (7) appropriate for the Mongan case IV potential we arrive at the expression for the off-shell Jost function

$$f_0^{cs}(k,q) = \left(\frac{q+k}{q-k}\right)^{i\eta} - \frac{2\eta k}{k+q} \sum_{r=1}^{2} \frac{d_r(k)}{\alpha_r + ik} \sum_{n=1}^{\infty} \frac{i^{-n+1}}{(n+1)}\left(\frac{\alpha_r + ik}{k+q}\right)^n \times {}_2F_1\left(1, n+1+i\eta; n+2; \frac{2k}{k+q}\right) + \sum_{r=1}^{2} \frac{\lambda_r}{(\alpha_r - iq)(\alpha_r^2+k^2)}\left[\left(\frac{\alpha_r + ik}{k+q}\right)^{i\eta} - \frac{d_r(k)}{(\alpha_r - ik)} \sum_{n=1}^{\infty} (-1)^n \left(\frac{\alpha_r + ik}{\alpha_r - ik}\right)^n {}_2F_1\left(1, n+1+i\eta; n+1; \frac{-2ik}{\alpha_r - ik}\right)\right] - \sum_{r=1}^{2} \sum_{\substack{S=1 \\ r \neq S}}^{2} \frac{\lambda_r}{(\alpha_r - iq)(\alpha_r - ik)^2} \frac{d_S(k)}{(\alpha_S + ik)} \sum_{n=1}^{\infty} (-1)^n \left(\frac{\alpha_S + ik}{\alpha_r - ik}\right)^n \times {}_2F_1\left(1, n+1+i\eta; n+1; \frac{-2ik}{\alpha_r - ik}\right). \quad (27)$$

Checks similar to those carried out for (19) can also be made here. In appropriate limits (27) yields results for the pure Coulomb and for the pure Mongan case IV potentials obtained by one of us.[15]

4. CONCLUDING REMARKS

Following the line of thought initiated by Fuda and by van Haeringen we have derived an analytical expression for

the l-wave off-shell Jost function for the Coulomb field in terms of Gaussian hypergeometric functions. These functions could be generated by the use of a three-term recurrence relation which avoids truncation errors resulting from series summation of ${}_2F_1$ functions. We have also derived similar results for (i) Coulomb plus Yamaguchi and (ii) Coulomb plus Mongan case IV potentials for the s-wave case only. These expressions exhibit relatively complicated mathematical structures as compared to the pure Coulomb Jost function. However, hypergeometric functions which enter into the expressions for $f_0^{cs}(k, q)$ can still be generated by the same prescription as used in the pure Coulomb case. To achieve a desired accuracy for the series in (19) or (27) one may proceed as follows.

For a chosen value of n the system of simultaneous equations resulting from (15) should be solved on a digital computer by the Gaussian elimination method and the entire array of functions be stored. The procedure may be repeated for some higher value of n. At a certain stage of such calculations the values of ${}_2F_1$ functions for large n will become negligible. This may be used as a criterion for truncating the series in (19) or (27).

[1]J. M. J. van Leeuwen and A. S. Reiner, Physica **27**, 99 (1961).
[2]M. G. Fuda and J. S. Whiting, Phys. Rev. C **8**, 1255 (1973).
[3]R. Jost, Helv. Phys. Acta. **20**, 256 (1947).
[4]A. Messiah, *Quantum Mechanics* (Wiley-Interscience, New York, 1961, 1962), Vols. 1 and 2.
[5]M. G. Fuda, Phys. Rev. C **14**, 37 (1976).
[6]K. L. Kowalski, Nucl. Phys. A **190**, 645 (1972); T. Sasakawa, Prog. Theor. Phys. Suppl. **27**, 1 (1963).
[7]H. van Haeringen, J. Math. Phys. **20**, 1109 (1979); **20**, 2520 (1979).
[8]C. Snow, Nat. Bur. Standards (U.S.) *Applied Math. Series 19* (U.S. Govt. Printing Office, Washington D.C., 1952).
[9]Y. Yamaguchi, Phys. Rev. **95**, 1628 (1954).
[10]T. R. Mongan, Phys. Rev. **178**, 1597 (1969).
[11]B. Talukdar, D. K. Ghosh, and T. Sasakawa (referred to as paper I), J. Math. Phys. **23**, 1700 (1982).
[12]R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[13]B. Talukdar, R. N. Chaudhuri, U. Das, and P. Banerjee, J. Math. Phys. **19**, 1654 (1978); U. Das, S. Chakravarty, and B. Talukdar, J. Math. Phys. **20**, 887 (1979).
[14]C. S. Warke and M. K. Srivastava, Phys. Rev. C **15**, 561 (1977).
[15]B. Talukdar, U. Das, and S. Mukhopadhyaya, J. Math. Phys. **20**, 372 (1979).

External boundary perturbations in neutron transport theory

F. Rahnema[a] and G. C. Pomraning
School of Engineering and Applied Science, University of California, Los Angeles, Los Angeles, California 90024

(Received 21 April 1982; accepted for publication 1 July 1982)

Two formalisms are presented for estimating the change in the fundamental eigenvalue (the criticality) of the neutron transport equation due to a perturbation in the exterior boundary of the system. The first-order perturbation formulas are derived within the context of energy-dependent transport theory and its energy-dependent diffusion approximation. Several numerical examples are given.

PACS numbers: 28.20.He, 03.65. — w

I. INTRODUCTION

The mathematical description of transport processes can be extremely complex. The transport (or the diffusion) equation characterizing such processes is often far too complex to allow a direct analytical (as well as numerical) solution for any but simple, modeled problems (e.g., simplified geometries or scattering kernels). Therefore, it is not surprising that the theory of neutron transport represents a fertile field for development and application of approximate methods in reactor physics. Some of the widely used techniques are the perturbation or variational methods that make use of the solutions to simple problems to estimate features of the solutions to more complex problems. That is, the complexity in such problems can be represented mathematically as a small perturbation on the diffusion or transport equation characterizing a far simpler problem. For these problems we can use familiar methods from perturbation theory to obtain first-order estimates of integral quantities such as eigenvalues or reaction rates in terms of the solution to the unperturbed problem. For example, in nuclear reactor calculations one is frequently required to estimate the effects of a small change in the reactor core composition or geometry on the multiplication factor K_{eff} (the criticality eigenvalue). If these changes or perturbations are small, it is not necessary to repeat the criticality calculation for the new core configuration; rather, the calculated flux for the unperturbed core can be used to make simple estimates of the change in K_{eff}.[1,2] Similar problems arise in a variety of other transport phenomena. There are also several specialized techniques such as multigroup methods or degenerate kernel expansions which can be used to generate approximate solutions to transport problems.

Our primary concern here is to develop a *boundary* perturbation method that can easily be applied to the analysis of eigenvalue problems both in transport theory and the diffusion approximation. In this case, a perturbation of order-1 amplitude in a small volume on the exterior boundary of a system is prescribed, and the problem is to seek a formalism which estimates the first-order change in the fundamental eigenvalue. We note that, unlike the usual perturbation approach which envisions order-ϵ material composition (e.g., cross sections) changes over a volume of order 1, boundary perturbation problems involve order-1 property changes over a volume of order-ϵ.

The subject of boundary perturbation theory is quite old. However, in the field of reactor physics, Larsen and Pomraning[3] have very recently developed such a boundary perturbation formalism for the first order change in the eigenvalue. Their formalism is restricted to one-speed models both in linear transport theory and the diffusion approximation.

In this paper we present two approaches, different from Larsen and Pomraning's, to the boundary perturbation problem. For the special case of one-speed models we will show that both of these methods yield the same results (formula) as those of Larsen and Pomraning. The applicability of our formalisms to energy-dependent problems make them more useful to the reactor physics community. We also believe that one of these methods is conceptually less complicated and algebraically less involved than that suggested by Larsen and Pomraning. In fact, this method is a natural extension of the ordinary operator perturbation method used extensively in reactor physics and other fields. In the following section, we briefly review the existing literature relevant to this study.

II. SELECTED REVIEW OF PREVIOUS PAPERS CONCERNED WITH BOUNDARY PERTURBATIONS

The subject of boundary perturbations was extensively studied in the time span of the late 1930s and early 1940s. In fact, this subject was perhaps considered first by Rayleigh in the late 1800s. In his famous book *The Theory of Sound*,[4] he investigated how the natural vibration of a uniform membrane is affected by a slight departure from the exact circular form. Rayleigh's approach consists of expanding the perturbed eigenfunction in an infinite series of the unperturbed eigenfunctions. The same approach was taken by Polya and Szego[5] as well as in the early work of Morse and Feshbach.[6] The main interest in the subject of boundary perturbation lies in the fields of acoustics and quantum mechanics. In this case, one imagines a small change in the domain of the problems and seeks a formula (or a set of equations) which estimates the eigenvalue (or the eigenfunction) of the perturbed problem. Of the previous work in this field one must mention the papers by Brillouin,[7] Cabrera,[8] Froelich,[9] Feshbach and

[a] Permanent address: General Electric Co., Nuclear Energy Business Operations 175 Curtner Ave., M/C 151, San Jose, California 95125.

 0022-2488/83/030687-15$02.50

Clogston,[10] Fuchs,[11] Feshbach,[12] Bolt, Feshbach, and Clogston,[13] Wassermann,[14,15] Feshbach and Harris,[16] Keller and Keller,[17] Strang and Berger,[18] Blair,[19] Helton and Ralston,[20] and Larsen and Pamraning.[3] Feshbach and Clogston's as well as Wassermann's developments contain and generalize the methods of other authors except that of Froehlich. In the following, we briefly summarize some of these developments.

Froehlich[9]: In his paper, Froehlich considered the Schrödinger equation

$$\nabla^2\Psi_0 + (E_0 - V)\Psi_0 = 0, \tag{1}$$

with a certain boundary condition at the surface $\partial\mathscr{D}_0$ bounding the domain $\mathscr{D}_0$. His intent was to calculate the eigenvalue $E \equiv E_0 + \epsilon$ corresponding to the Schrödinger operator defined in a new domain $\mathscr{D}$ bounded by the surface $\partial\mathscr{D}$ which is slightly different from $\partial\mathscr{D}_0$,

$$\nabla^2\Psi + (E_0 + \epsilon - V)\Psi = 0. \tag{2}$$

Multiplying Eq. (1) by Ψ and Eq. (2) by Ψ_0 and subtracting one result from the other, he found, after integrating over $\mathscr{D}$ and using Green's theorem,

$$\int_{\partial\mathscr{D}} dS\left(\Psi_0\frac{\partial\Psi}{\partial n} - \epsilon\frac{\partial\Psi_0}{\partial n}\right) + \epsilon\int_{\mathscr{D}} d\mathbf{x}\,\Psi_0\Psi = 0, \tag{3}$$

where $\partial/\partial n \equiv \hat{n}\cdot\nabla$, in which $\hat{n}$ is the unit normal to $\partial\mathscr{D}$ in the outward direction. In an attempt to compute ϵ, Froehlich considered the following boundary conditions:

$$\Psi_0 = 0 \quad \text{on } \partial\mathscr{D}_0 \quad \text{and} \quad \Psi = 0 \quad \text{on } \partial\mathscr{D}, \tag{4a}$$

or

$$\frac{\partial\Psi_0}{\partial n} = 0 \quad \text{on } \partial\mathscr{D}_0 \quad \text{and} \quad \frac{\partial\Psi}{\partial n} = 0 \quad \text{on } \partial\mathscr{D}. \tag{4b}$$

For small changes in the boundary, he found

$$\epsilon = \int_{\partial\mathscr{D}} dS\,\Psi_0\frac{\partial\Psi_0}{\partial n} \tag{5a}$$

or

$$\epsilon = -\int_{\partial\mathscr{D}} dS\,\Psi_0\frac{\partial\Psi_0}{\partial n}, \tag{5b}$$

where Ψ_0 is normalized to unity. Therefore, it is possible to determine the first-order approximation to the energy E by an integration over functions that are known from the unperturbed problem. However, Froehlich failed to mention that in Eqs. (5a) and (5b) one needs to compute the analytic continuation of Ψ_0 onto $\partial\mathscr{D}$ which is a difficult task even for simple geometries. Another shortcoming of these formulas is that the integrations are performed over the perturbed surface (i.e., the intent in any perturbation method is to avoid the perturbed problem as much as possible).

In this paper, we develop a method (identified as Method 1), in which we use Froehlich's cross multiplication idea, to compute eigenvalue changes due to a boundary perturbation. By defining an equation, which relates points on the unperturbed surface to those on the perturbed surface, we overcome the shortcomings mentioned above. As a result, the calculation of the eigenvalue change only requires the eigenfunction and its derivatives at the unperturbed surface. The idea of cross multiplication as used by Froehlich is a common procedure utilized in reactor physics and other fields when the problem involves a perturbation of the (interior) operator. In fact, one may consider Method 1 a natural generalization of the standard operator perturbation method to include boundary changes as well as an extension of Froehlich's method.

Feshbach and Clogston[10]: This method is nearly the same as those of Brillouin and Cabrera. In Feshbach and Clogston's (hereafter referred to as FC) paper, the change in the boundary was introduced into the problem as a small perturbation in the differential operator. This was done in the following manner.

Consider the eigenvalue equation

$$H_0\phi_0 + K_0\phi_0 = 0 \quad \text{in } \mathscr{D}_0, \tag{6}$$

which has been solved with ϕ_0 satisfying some boundary conditions denoted by i on $\partial\mathscr{D}_0$. It is required to find the solution of the equation

$$H_0\psi + K\psi = 0 \quad \text{in } \mathscr{D}, \tag{7}$$

with ψ satisfying other boundary conditions denoted by ii on $\partial\mathscr{D}$. The idea was to let T be some operator, and define a function by

$$\phi = T\psi, \tag{8}$$

where ϕ is to satisfy conditions i on $\partial\mathscr{D}_0$. This is the very condition on T. If one assumes that $T = I + \sigma + \cdots$, where I is the identity operator and σ is a quantity small in first order, then one obtains the equation satisfied by ϕ as

$$(H_0 + H_1 + \cdots)\phi + K\phi = 0, \tag{9}$$

where

$$H_1 = \sigma H_0 - H_0\sigma. \tag{10}$$

Thus, the problem has been reduced to one in which ordinary perturbation methods may be applied. This gives

$$K = K_0 + K_1 + \cdots, \tag{11}$$

where

$$K_1 = \int_{\partial\mathscr{D}_0} d\mathbf{x}\,\phi_0 H_1\phi_0. \tag{12}$$

The formula for the perturbed eigenfunctions is not given here, because of our sole interest in the eigenvalues.

Now, one is to determine $H_1\phi_0$, which as yet is unknown. By introducing the transformation operator R $(= I + \Sigma_i X_i(\partial/\partial x_i) + \cdots$, where X_i is the perturbation), FC made an attempt to compute $H_1\phi_0$ [with $H_0 = \nabla^2 + V(\mathbf{x})$], which involves finding σ according to the boundary conditions on ψ. In their paper, they emphasized that to find a σ satisfying all the boundary conditions [specifically of the type $\partial\psi/\partial n = 0$ or $\partial\Psi/\partial n = F(\mathbf{x})\Psi$ on $\partial\mathscr{D}$ with $F(\mathbf{x})$ small] and of sufficient continuity may be, in all but simple cases, very difficult. They conjecture that the eigenvalues can always be obtained in terms of surface integrals and thus made independent of the volume behavior of σ. However, they were only able to verify this in cases where the boundary was fixed ($\mathscr{D} = \mathscr{D}_0$) and only the boundary conditions were perturbed.

In the following, we use FC's approach to the boundary perturbation problem in transport theory and the diffusion

approximation. By recognizing that the vector normal to the perturbed surface may be expanded as

$$\hat{n} = \hat{n}_0 + \epsilon \mathbf{t}_0 + \cdots, \tag{13}$$

where ϵ is a smallness parameter, we are able to indirectly find a σ that satisfies boundary conditions of the types $\psi + \gamma(\partial\psi/\partial n) = 0$ on $\partial\mathscr{D}$ (in diffusion theory) and $\psi(\mathbf{x},\hat{\Omega}) = 0$ on $\mathscr{D}$ for all $\hat{n}\cdot\hat{\Omega} < 0$ (in transport theory). In the above equation $\mathbf{t}_0$ is a vector tangential to $\partial\mathscr{D}_0$ which is given in terms of the derivatives of the function $f(\mathbf{x}_{s_0}) = 0$ defining the unperturbed surface $\partial\mathscr{D}_0$. This extension (generalization) of FC's approach is identified as Method 2 in Secs. III B and IV B.

Wasserman[15]: In contrast to FC,[10] Wasserman did not utilize an inverse displacement operator, Eq. (8). The introduction of Gaussian parameter u_i on the unperturbed surface avoided the use of the inverse operator and made the general representation more natural. In fact, he established a one-to-one correspondence between the points of the unperturbed surface $\partial\mathscr{D}_0$ and those of the perturbed surface $\partial\mathscr{D}$ by means of a displacement vector $\boldsymbol{\eta}$, which joins corresponding points. However, to obtain a continuous mapping he came across a rather complicated differential equation for $\boldsymbol{\eta}$ that naturally involves the equations of the surfaces $\partial\mathscr{D}_0$ and $\partial\mathscr{D}$. Another complication of his method is that $|\partial\boldsymbol{\eta}/\partial u_i|_{\max}$ is generally of zero order, and in some practical cases they found it to be of order $10|\boldsymbol{\eta}|_{\max}$ ($\boldsymbol{\eta}$ being of first order). This difficulty, however, was overcome by introducing a new mode of separating orders. In his previous work, the zero-order magnitude of the derivatives of the displacement vector was formerly looked upon as first-order quantities.

In summary, Wasserman assumed that the eigenfunctions and eigenvalues of an eigenvalue problem given by an elliptic differential equation are known subject to given boundary conditions on a finite boundary. He then showed how the corresponding quantities can be obtained for a similar problem in which the original differential equation, boundary, and boundary conditions are simultaneously perturbed. The introduction of a surface displacement vector allowed him to Taylor expand all quantities and subsequently separate orders. The problem of finding the eigenfunctions for each order was then reduced to solving an inhomogeneous differential equation subject to known boundary conditions. The equations were solved by a variational method, unlike most of the previous attempts[6,10–13] in this field which depended on solutions in series of the unperturbed eigenfunctions or on solutions involving a Green's function. An application of Green's theorem at each stage enabled him to find the perturbed eigenvalues.

Recent authors (1952–1976): Keller and Keller[17] have considered the problem of obtaining a lower bound on the fundamental eigenvalue of the Helmholtz equation for nearly circular regions. On the other hand, Strang and Berger[18] estimated the change in the solution of $-\nabla^2 u = f$ (with the Dirichlet condition $u = 0$ on the smooth boundary) due to a change in the domain $\mathscr{D}$ of the problem. In particular, $\mathscr{D}$ was replaced with an inscribed polygon $\mathscr{D}_h$ as the perturbed domain.

Blair,[19] later, generalized Strang and Berger's method to obtain bounds for the change in the solutions of second-order elliptic partial differential equations when the boundary of the system is perturbed.

Larsen and Pomraning[3]: Very recently Larsen and Pomraning developed an approach to the boundary perturbation problem in which they obtain a first-order formalism for the change in the eigenvalue $\delta\Lambda$ both in transport theory and the diffusion approximation. The basic idea in their method is to introduce a small dimensionless parameter ϵ into the perturbed problem and obtain expressions that lead to $\delta\Lambda$ by differentiating the eigenvalue equation as well as the boundary conditions with respect to ϵ. Since we will compare LP's results with those obtained in this paper, a brief review of their approach is in order.

Consider the one-group diffusion eigenvalue equation

$$L(\mathbf{x})\phi(\mathbf{x},\epsilon) = \Lambda(\epsilon)G(\mathbf{x})\phi(\mathbf{x},\epsilon), \quad \mathbf{x}\in\mathscr{D}_\epsilon, \tag{14}$$

where

$$L(\mathbf{x}) \equiv -\nabla\cdot D(\mathbf{x})\nabla + \sigma_a(\mathbf{x}) \tag{15}$$

and

$$G(\mathbf{x}) \equiv \nu(\mathbf{x})\sigma_f(\mathbf{x}). \tag{16}$$

In this equation $\phi(\mathbf{x},\epsilon)$ is the scalar flux (fundamental eigenfunction) and $\Lambda(\epsilon)$ is the corresponding eigenvalue (the reciprocal of K_{eff}). The boundary surface $\partial\mathscr{D}_\epsilon$, bounding the domain $\mathscr{D}_\epsilon$, is defined by

$$F(\mathbf{x}_\epsilon,\epsilon) = 0, \tag{17}$$

where F is a function such that ∇F and $\partial F/\partial\epsilon$ are $O(1)$ for $\mathbf{x}$ near $\partial\mathscr{D}_\epsilon$ and $\epsilon \ll 1$. The boundary condition on Eq. (14) is given by

$$\phi(\mathbf{x}_\epsilon,\epsilon) + \gamma(\mathbf{x}_\epsilon)\hat{n}(\mathbf{x}_\epsilon,\epsilon)\cdot\nabla\phi(\mathbf{x}_\epsilon,\epsilon) = 0, \quad \mathbf{x}_\epsilon\in\partial\mathscr{D}_\epsilon, \tag{18}$$

where $\hat{n}(\mathbf{x}_\epsilon,\epsilon)$ is the unit outward normal to $\partial\mathscr{D}_\epsilon$ at the point $\mathbf{x}_\epsilon$ and $\gamma(\mathbf{x}_\epsilon)$ is the linear extrapolation distance. We note that the unperturbed problem (whose solution is assumed known) is immediately obtained by setting $\epsilon = 0$ in the above equations.

They assume in their analysis that $d\Lambda/d\epsilon|_{\epsilon=0}$ and $\partial\phi/\partial\epsilon|_{\epsilon=0}$ exist and that no operator perturbation is present in the interior of the system (i.e., the interior composition of the system is unchanged as the boundary is perturbed). Then, by differentiating Eqs. (14) and (18) with respect to ϵ and setting $\epsilon = 0$ in the resulting equations, they obtain

$$[L(\mathbf{x}) - \Lambda(0)\,G(\mathbf{x})]\phi_1(\mathbf{x}) = \Lambda'(0)\,G(\mathbf{x})\phi_0(\mathbf{x}), \quad \mathbf{x}\in\mathscr{D}_0, \tag{19}$$

and

$$\phi_1(\mathbf{x}_0) + \gamma(\mathbf{x}_0)\hat{n}(\mathbf{x}_0)\cdot\nabla\phi_1(\mathbf{x}_0) = -\theta(\mathbf{x}_0), \quad \mathbf{x}_0\in\partial\mathscr{D}_0, \tag{20}$$

where

$$\begin{aligned}\theta(\mathbf{x}_0) = {} & X(\mathbf{x}_0)[1 + \hat{n}(\mathbf{x}_0)\cdot\nabla\gamma(\mathbf{x}_0)]\hat{n}(\mathbf{x}_0)\cdot\nabla\phi_0(\mathbf{x}_0) \\ & + \gamma(\mathbf{x}_0)\mathbf{t}(\mathbf{x}_0)\cdot\nabla\phi_0(\mathbf{x}_0) \\ & \gamma(\mathbf{x}_0)X(\mathbf{x}_0)\hat{n}(\mathbf{x}_0)\cdot[\hat{n}(\mathbf{x}_0)\cdot\nabla]\nabla\phi_0(\mathbf{x}_0),\end{aligned} \tag{21}$$

$$\phi_1(\mathbf{x}) = \left.\frac{\partial\phi}{\partial\epsilon}\right|_{\epsilon=0}, \quad \mathbf{x}\in\mathscr{D}_0, \tag{22}$$

and

$$\mathbf{t}(\mathbf{x}_0) = \frac{d}{d\epsilon}\,\hat{n}(\mathbf{x}_\epsilon,\epsilon)|_{\epsilon=0}, \quad \mathbf{x}_0 \in \partial\mathscr{D}_0. \tag{23}$$

To first order, $\epsilon X(\mathbf{x}_0)$ is the distance between $\partial\mathscr{D}_\epsilon$ and $\partial\mathscr{D}_0$ at the point $\mathbf{x}_0$ in the normal direction, and is given by

$$X(\mathbf{x}_0) = -\frac{(\partial F/\partial\epsilon)|_{\epsilon=0}}{\|\nabla F|_{\epsilon=0}\|}, \quad \mathbf{x}_0 \in \partial\mathscr{D}_0. \tag{24}$$

In a straightforward way, $\Lambda'(0)$ is explicitly derived from Eq. (19) as

$$\Lambda'(0)\int_{\mathscr{D}_0} d\mathbf{x}\,\nu(\mathbf{x})\sigma_f(\mathbf{x})\phi_0^2(\mathbf{x}) = -\int_{\partial\mathscr{D}_0} dS\,D(\mathbf{x})\theta(\mathbf{x})\frac{\partial\phi_0(\mathbf{x})}{\partial n}. \tag{25}$$

Finally, one is to note that, with $\Lambda(0)$ and $\Lambda'(0)$ known, we have

$$\Lambda(\epsilon) \approx \Lambda(0) + \epsilon\Lambda'(0), \tag{26}$$

with an error which is formally $O(\epsilon^2)$.

In a completely similar manner, they obtain the change in the eigenvalue for the transport equation as

$$\Lambda'(0)\int_{\mathscr{D}_0} d\mathbf{x}\,\nu(\mathbf{x})\sigma_f(\mathbf{x})v(\mathbf{x}) = -\int_{\partial\mathscr{D}_0} dS\,X(\mathbf{x}_0)[\sigma_s(\mathbf{x})u(\mathbf{x}) + \Lambda(0)\nu(\mathbf{x})\sigma_f(\mathbf{x})v(\mathbf{x})], \tag{27}$$

where

$$u(\mathbf{x}) = (1/4\pi)\int\int d\hat{\Omega}d\hat{\Omega}'\,\psi_0(\mathbf{x}_0, -\Omega)f(\mathbf{x},\hat{\Omega}\cdot\hat{\Omega}')\psi_0(\mathbf{x},\hat{\Omega}'), \tag{28}$$

and

$$v(\mathbf{x}) = \left[(1/4\pi)\int d\hat{\Omega}\,\psi_0(\mathbf{x},\hat{\Omega})\right]^2. \tag{29}$$

In this case, $L(\mathbf{x})$ and $G(\mathbf{x})$ in Eq. (14) describe the one-speed transport operator given by

$$L(\mathbf{x}) = \hat{\Omega}\cdot\nabla + \sigma(\mathbf{x}) - \sigma_s(\mathbf{x})\int d\hat{\Omega}'\,f(\mathbf{x},\hat{\Omega}\cdot\hat{\Omega}')(\cdot), \tag{30}$$

and

$$G(\mathbf{x}) = \frac{\nu(\mathbf{x})\sigma_f(\mathbf{x})}{4\pi}\int d\hat{\Omega}'(\cdot), \tag{31}$$

with the following free surface condition on ψ:

$$\psi(\mathbf{x}_\epsilon,\hat{\Omega},\epsilon) = 0, \quad \hat{\Omega}\cdot n(\mathbf{x}_\epsilon,\epsilon) < 0, \quad \mathbf{x}\in\partial\mathscr{D}_\epsilon. \tag{32}$$

This approach seems to be simpler (in concept) than those suggested by most of the previous authors.[9–15] However, as will be seen shortly, one of our methods is perhaps conceptually even simpler and algebraically less involved than that of Larsen and Pomraning.

III. ANALYSIS IN DIFFUSION THEORY

For simplicity, the discussion in this section is first restricted to the one-speed model in the case of the diffusion approximation. The formalism is then extended to energy-dependent models. We will show that both of our methods yield the same results (formulas) as those of Larsen and Pomraning[3] for the special case of one-group theory.

A. Method 1

Consider the one-speed diffusion eigenvalue equation in a domain $\mathscr{D}_0$ bounded by a smooth surface $\partial\mathscr{D}_0$,

$$-\nabla\cdot D(\mathbf{x})\nabla\phi_0(\mathbf{x}) + \sigma_a(\mathbf{x})\phi_0(\mathbf{x}) = \Lambda_0\nu(\mathbf{x})\sigma_f(\mathbf{x})\phi_0(\mathbf{x}), \quad \mathbf{x}\in\mathscr{D}_0, \tag{33}$$

where $\phi_0(\mathbf{x})$ is the scalar flux and Λ_0 ($= 1/K_{\text{eff}}$) is the eigenvalue of the unperturbed problem. We wish to determine the eigenvalue ($\Lambda_0 + \delta\Lambda$) corresponding to the same diffusion operator with a slightly different boundary $\partial\mathscr{D}$:

$$\nabla\cdot D(\mathbf{x})\nabla\phi(\mathbf{x}) + \sigma_a(\mathbf{x})\phi(\mathbf{x}) = (\Lambda_0 + \delta\Lambda)\nu(\mathbf{x})\sigma_f(\mathbf{x})\phi(\mathbf{x}), \quad \mathbf{x}\in\mathscr{D}, \tag{34}$$

where $\phi(\mathbf{x})$ is the perturbed scalar flux and $\delta\Lambda$ is the perturbation in the eigenvalue due to the change in the boundary. The remaining notation is standard.[21] $D(\mathbf{x})$, $\sigma_a(\mathbf{x})$, and $\nu(\mathbf{x})\sigma_f(\mathbf{x})$ are smooth functions of $\mathbf{x}$ defined in a domain containing $\mathscr{D}_0$ and $\mathscr{D}$ ($\mathscr{D}_0\cup\mathscr{D}$). We note that these functions are assumed to be unchanged due to the boundary perturbation.

Functions $\phi_0(\mathbf{x})$ and $\phi(\mathbf{x})$ are considered to satisfy the Robbin (mixed) type boundary conditions in which linear combinations of the function and its normal derivative vanish on the boundary, i.e.,

$$\phi_0(\mathbf{x}_{s_0}) + \gamma(\mathbf{x}_{s_0})\hat{n}_0\cdot\nabla\phi_0(\mathbf{x}_{s_0}) = 0, \quad \mathbf{x}_{s_0}\in\partial\mathscr{D}_0, \tag{35}$$

and

$$\phi(\mathbf{x}_s) + \gamma(\mathbf{x}_s)\hat{n}\cdot\nabla\phi(\mathbf{x}_s) = 0, \quad \mathbf{x}_s\in\partial\mathscr{D}, \tag{36}$$

where $\mathbf{x}_s$ and $\mathbf{x}_{s_0}$ are the vectors from the origin to the points on the surfaces $\mathscr{D}_0$ and $\mathscr{D}$, respectively; $\hat{n}_0$ and $\hat{n}$ are the unit outward normals to $\partial\mathscr{D}_0$ and $\partial\mathscr{D}$ at $\mathbf{x}_{s_0}$ and $\mathbf{x}_s$, respectively, and $\gamma(\mathbf{x})$ is the linear extrapolation distance.

We assume the eigenvalue Λ_0 and the solution $\phi_0(\mathbf{x})$ are known, and we seek a first-order perturbation formula for the change in the eigenvalue corresponding to the perturbation of the smooth boundary $\partial\mathscr{D}_0$. Following the method of Froehlich,[9] we multiply Eq. (33) by $\phi(\mathbf{x})$, multiply Eq. (34) by $\phi_0(\mathbf{x})$, subtract one result from the other, and integrate over $\mathscr{D}$. This gives

$$\delta\Lambda\int_{\mathscr{D}} dx\,\nu(\mathbf{x})\sigma_f(\mathbf{x})\phi_0(\mathbf{x})\phi(\mathbf{x}) = \int_{\partial\mathscr{D}} dS\,D(\mathbf{x})\left(\phi(\mathbf{x})\frac{\partial\phi_0(\mathbf{x})}{\partial n} - \phi_0(\mathbf{x})\frac{\partial\phi(\mathbf{x})}{\partial n}\right). \tag{37}$$

This result can be written, using the boundary condition on ϕ, Eq. (36), as

$$\delta\Lambda\int_{\mathscr{D}} d\mathbf{x}\,\nu\sigma_f\phi\phi_0 = -\int_{\partial\mathscr{D}} dS\,D\frac{\partial\phi}{\partial n}\left(\phi_0 + \gamma\frac{\partial\phi_0}{\partial n}\right). \tag{38}$$

In the process of obtaining Eq. (38), it was assumed that $\phi_0(\mathbf{x})$ and $\phi(\mathbf{x})$ can be analytically extended into the domain $\mathscr{D}$ and $\mathscr{D}_0$, respectively, while satisfying their respective differential equations.[22] A rigorous determination of the conditions under which analytic continuation is possible is unquestionably a difficult mathematical problem. However, it is clear that a necessary condition is that the material added (or deleted) due to the perturbation must be a smooth extension of (e.g., identical with) the material at the surface of the unperturbed domain $\mathscr{D}_0$.

So far we have made no approximations in obtaining Eq. (38). Now, if we assume the change in the smooth boundary $\partial\mathscr{D}_0$ is small, then one can relate the perturbed surface points $\mathbf{x}$, to the unperturbed surface points $\mathbf{x}_{s_0}$ by

$$\mathbf{x}_s = \mathbf{x}_{s_0} + \epsilon X(\mathbf{x}_{s_0})\hat{n}_0 + O(\epsilon^2), \tag{39}$$

where, to first order, $X(\mathbf{x}_{s_0})$ is the distance between $\partial\mathscr{D}$ and $\partial\mathscr{D}_0$, at the point $\mathbf{x}_{s_0}$, in the normal direction and ϵ is a smallness parameter. Let the unperturbed $(\partial\mathscr{D}_0)$ and the perturbed $(\partial\mathscr{D})$ surface boundaries be given by

$$f(\mathbf{x}_{s_0}) = 0 \tag{40}$$

and

$$f(\mathbf{x}_s) = 0, \tag{41}$$

respectively. Then, using Eq. (39), one can show that the outer normal to $\partial\mathscr{D}$ satisfies the following equation.

$$\hat{n} = \hat{n}_0 + \epsilon X(\mathbf{x}_{s_0})\mathbf{t}_0 + O(\epsilon^2), \tag{42}$$

where $\mathbf{t}_0$ is a vector tangential $(\mathbf{t}_0\cdot\hat{n}_0 = 0)$ to $\partial\mathscr{D}_0$ at $\mathbf{x}_{s_0}$, and is given by

$$\mathbf{t}_0 = \frac{(\hat{n}_0\cdot\nabla)\nabla f(\mathbf{x}_{s_0}) - \hat{n}_0[\hat{n}_0\cdot(\hat{n}_0\cdot\nabla)\nabla f(\mathbf{x}_{s_0})]}{\|\nabla f(\mathbf{x}_{s_0})\|}. \tag{43}$$

In obtaining Eq. (42) we have assumed that $f(\mathbf{x}_s)$ is a smooth function of ϵ. If we further assume that the functions in the integrand on the right-hand side of Eq. (38) are smooth in ϵ, then we can express them as Taylor series expansion in ϵ about the point $\mathbf{x}_{s_0}$. Use of Eqs. (35), (39), and (42) gives

$$\begin{aligned} &D\left\{\frac{\partial\phi}{\partial n}\left(\phi_0 + \gamma\frac{\partial\phi_0}{\partial n}\right)\right\}_{\mathbf{x}=\mathbf{x}_s} \\ &\quad = \epsilon\left\{XD\frac{\partial\phi}{\partial n_0}\left[\left(1+\frac{\partial\gamma}{\partial n_0}\right)\frac{\partial\phi_0}{\partial n_0}\right.\right. \\ &\qquad \left.\left. + \gamma\left(\frac{\partial^2\phi_0}{\partial n_0^2} + |\mathbf{t}_0|\frac{\partial\phi_0}{\partial t_0}\right)\right]\right\}_{\mathbf{x}=\mathbf{x}_{s_0}} + O(\epsilon^2), \end{aligned} \tag{44}$$

where

$$|\mathbf{t}_0|\frac{\partial\phi_0}{\partial t_0} \equiv \mathbf{t}_0\cdot\nabla\phi_0. \tag{45}$$

Thus, Eq. (38) becomes

$$\begin{aligned} &\delta\Lambda\int_{\mathscr{D}} d\mathbf{x}\,\nu\sigma_f\phi\phi_0 \\ &\quad = -\epsilon\int_{\partial\mathscr{D}} dS\,XD\left\{\frac{\partial\phi}{\partial n_0}\left[\left(1+\frac{\partial\gamma}{\partial n_0}\right)\frac{\partial\phi_0}{\partial n_0}\right.\right. \\ &\qquad \left.\left. + \gamma\left(\frac{\partial^2\phi_0}{\partial n_0^2} + |\mathbf{t}_0|\frac{\partial\phi_0}{\partial t_0}\right)\right]\right\}_{\mathbf{x}=\mathbf{x}_{s_0}} + O(\epsilon^2). \end{aligned} \tag{46}$$

We note that the pairs (ϕ_0,ϕ), $(\partial\mathscr{D}_0,\partial\mathscr{D})$, and $(\mathscr{D},\mathscr{D}_0)$ all differ by order ϵ, and hence we can replace ϕ, $\partial\mathscr{D}$, and $\mathscr{D}$ in the above equation by ϕ_0, $\partial\mathscr{D}_0$, and $\mathscr{D}_0$ to obtain

$$\begin{aligned} &\Lambda_1\int_{\mathscr{D}_0} d\mathbf{x}\,\nu\sigma_f\phi_0^2 \\ &\quad = -\int_{\partial\mathscr{D}_0} dS\,XD\frac{\partial\phi_0}{\partial n_0} \\ &\qquad \times\left[\left(1+\frac{\partial\gamma}{\partial n_0}\right)\frac{\partial\phi_0}{\partial n_0} + \gamma\left(\frac{\partial^2\phi_0}{\partial n_0^2} + |\mathbf{t}_0|\frac{\partial\phi_0}{\partial t_0}\right)\right]. \end{aligned} \tag{47}$$

Λ_1 represents the first-order change in the eigenvalue due to the perturbation in the boundary, i.e.,

$$\begin{aligned} \Lambda &= \Lambda_0 + \delta\Lambda \\ &= \Lambda_0 + \epsilon\Lambda_1 + O(\epsilon^2). \end{aligned} \tag{48}$$

It must be mentioned that Eq. (47) is exactly the same as that obtained by Larsen and Pomraning[3] using a different method. Our belief is that the perturbation method presented here is perhaps conceptually simpler (and less involved) than the method developed by LP. Equation (47) makes it possible to estimate the eigenvalue corresponding to the perturbed problem by integrating functions which are either known or defined by the virtue of the perturbation. This formalism is derived for a general mixed boundary condition and a general function f. The results are by no means simple, especially because of the vector $\mathbf{t}_0$, in most geometries. However, in the following we derive a considerably simplified version of the formula for some simple geometries.

If the solution to the unperturbed problem, $\phi_0(\mathbf{x})$ is spherically symmetric with $\partial\mathscr{D}_0$ as the surface of a sphere with radius r_0, then Eq. (47) can be written as

$$\begin{aligned} &\Lambda_1\int_{\mathscr{D}_0} d\mathbf{r}\,\nu(\mathbf{r})\sigma_f(\mathbf{r})\phi_0^2(\mathbf{r}) \\ &\quad = -\delta V D(r_0)\left\{\left(\frac{\partial\phi_0}{\partial r} + \frac{\partial}{\partial r}\gamma\frac{\partial\phi_0}{\partial r}\right)\frac{\partial\phi_0}{\partial r}\right\}_{r=r_0}, \end{aligned} \tag{49}$$

where δV is defined as the difference in volume, to first order, bounded by $\partial\mathscr{D}$ and $\partial\mathscr{D}_0$, respectively.

In order for the perturbation formalism just developed to be useful in the reactor physics community, one must generalize the method to energy-dependent (both continuous and multigroup) reactor theory problems. Since the perturbation corresponds to only the spatial variable $(\mathbf{x})$ and not to the energy variable E, the generalization is rather straightforward.

The energy dependent diffusion eigenvalue problem[23] (the unperturbed problem) is described by

$$\begin{aligned} &-\nabla\cdot D(\mathbf{x},E)\nabla\phi_0(\mathbf{x},E) + \sigma(\mathbf{x},E)\phi_0(\mathbf{x},E) \\ &\quad = \int_{E'} dE'\,[\sigma_s(\mathbf{x};E'\to E) + \Lambda_0\nu(\mathbf{x},E')\sigma_f(\mathbf{x},E'\to E)] \\ &\qquad \times\phi_0(\mathbf{x},E'), \quad \mathbf{x}\in\mathscr{D}_0, \end{aligned} \tag{50}$$

with boundary condition

$$\phi_0(\mathbf{x}_{s_0},E) + \gamma(\mathbf{x}_{s_0},E)\frac{\partial\phi_0(\mathbf{x}_{s_0},E)}{\partial n_0} = 0, \quad \mathbf{x}_{s_0}\in\partial\mathscr{D}_0. \tag{51}$$

The perturbed description corresponds to the same diffusion operator with a slightly different boundary $\partial\mathscr{D}$,

$$\begin{aligned} &-\nabla\cdot D(\mathbf{x},E)\cdot\nabla\phi(\mathbf{x},E) + \sigma(\mathbf{x},E)\phi(\mathbf{x},E) \\ &\quad = \int_{E'} dE'\,[\sigma_s(\mathbf{x};E'\to E) \\ &\qquad + (\Lambda_0+\delta\Lambda)\nu(\mathbf{x},E')\sigma_f(\mathbf{x};E'\to E)]\phi(\mathbf{x},E), \quad \mathbf{x}\in\mathscr{D}, \end{aligned} \tag{52}$$

with boundary condition

$$\phi(\mathbf{x}_s,E) + \gamma(\mathbf{x}_s,E)\frac{\partial\phi(\mathbf{x}_s,E)}{\partial n} = 0, \quad \mathbf{x}_s\in\partial\mathscr{D}. \tag{53}$$

Here, we again want to determine a first-order perturbation formula for the change in the eigenvalue ($\delta\Lambda$) assuming that Λ_0 and $\phi_0(\mathbf{x},E)$ are known from the unperturbed problem.

It is necessary to introduce the reference adjoint problem [the adjoint to Eqs. (50) and (51)]. That is,

$$-\nabla\cdot D(\mathbf{x},E)\nabla\phi_0^*(\mathbf{x},E)+\sigma(\mathbf{x},E)\phi_0^*(\mathbf{x},E)$$
$$=\int_{E'}dE'\,[\sigma_s(\mathbf{x};E\to E')+\Lambda_0\nu(\mathbf{x},E)\sigma_f(\mathbf{x};E\to E')]$$
$$\times\phi_0^*(\mathbf{x},E'),\quad \mathbf{x}\in\mathscr{D}_0, \tag{54}$$

$$\phi_0^*(\mathbf{x}_{s_0},E)+\gamma(\mathbf{x}_{s_0},E)\frac{\partial\phi_0^*(\mathbf{x}_{s_0},E)}{\partial n_0}=0,\quad \mathbf{x}\in\partial\mathscr{D}_0. \tag{55}$$

We now multiply Eqs. (52) and (54) by $\phi_0^*(\mathbf{x},E)$ and $\phi(\mathbf{x},E)$, respectively, and subtract. Using analytic continuation to extend ϕ_0^* and ϕ into $\mathscr{D}$ and $\mathscr{D}_0$, respectively, we integrate the resulting equation over the domain $\mathscr{D}$ and all energies E to obtain

$$\delta\Lambda\int_{\mathscr{D}}d\mathbf{x}\int_E dE\int_{E'}dE'\,\nu(\mathbf{x},E')$$
$$\times\sigma_f(\mathbf{x},E'\to E)\phi(\mathbf{x},E')\phi_0^*(\mathbf{x},E)$$
$$=-\int_{\partial\mathscr{D}}dS\int_E dE\,D(\mathbf{x},E)\frac{\partial\phi(\mathbf{x},E)}{\partial n}$$
$$\times\left[\phi_0^*(\mathbf{x},E)+\gamma(\mathbf{x},E)\frac{\partial\phi_0^*(\mathbf{x},E)}{\partial n}\right], \tag{56}$$

where we have made use of the divergence theorem as well as the boundary condition on ϕ, Eq. (53).

We note that Eq. (56) represents an exact expression for the perturbation in the eigenvalue. However, since ϕ is not known, we use the approximating method described in the previous section to obtain an estimate for $\delta\Lambda$. Assuming that the functions in the integrand on the right-hand side of Eq. (56) are smooth in ϵ, we can then express the integrand as the Taylor series expansion in ϵ about the point $\mathbf{x}_{s_0}$. Using Eqs. (39)–(43) as well as the boundary condition on ϕ_0^*, we get

$$\left\{D\frac{\partial\phi}{\partial n}\left(\phi_0^*+\gamma\frac{\partial\phi_0^*}{\partial n}\right)\right\}_{(\mathbf{x},E)=(\mathbf{x}_{s_0},E)}$$
$$=\epsilon X(\mathbf{x}_{s_0})\left\{D\frac{\partial\phi}{\partial n_0}\left[\left(1+\frac{\partial\gamma}{\partial n_0}\right)\frac{\partial\phi_0^*}{\partial n_0}\right.\right.$$
$$\left.\left.+\gamma\left(\frac{\partial^2\phi_0^*}{\partial n_0^2}+|\mathbf{t}_0|\frac{\partial\phi_0^*}{\partial\mathbf{t}_0}\right)\right]\right\}_{(\mathbf{x}_{s_0},E)}+O(\epsilon^2). \tag{57}$$

Hence, Eq. (56) becomes

$$\Lambda_1\int_{\mathscr{D}_0}d\mathbf{x}\int_E dE\int_{E'}dE'\,\nu(\mathbf{x},E')\sigma_f(\mathbf{x},E'\to E)\phi_0(\mathbf{x},E')\phi_0^*(\mathbf{x},E)$$
$$=-\int_{\partial\mathscr{D}_0}dS\,X(\mathbf{x})\int_E dE\,D\frac{\partial\phi_0}{\partial n_0}\left[\left(1+\frac{\partial\gamma}{\partial n_0}\right)\frac{\partial\phi_0^*}{\partial n_0}+\gamma\left(\frac{\partial^2\phi_0^*}{\partial n_0^2}+|\mathbf{t}_0|\frac{\partial\phi_0^*}{\partial t_0}\right)\right], \tag{58}$$

where we have already noted that the pairs (ϕ_0^*,ϕ), $(\partial\mathscr{D}_0,\partial\mathscr{D})$, and $(\mathscr{D},\mathscr{D}_0)$ all differ by order ϵ and have used Eq. (48).

Equation (58) is indeed the generalized version of Eq. (47). That is, if the problem is energy independent, then the equation given above reduces to the one-speed result, Eq. (47). We again note that this formula is somewhat complicated in its general form. However, for the special case of Dirichlet boundary conditions (vanishing flux condition) Eq. (58) would have the following simple form:

$$\Lambda_1\int_{\mathscr{D}_0}d\mathbf{x}\int_E dE\int_{E'}dE'\,\nu(\mathbf{x},E')$$
$$\times\sigma_f(\mathbf{x},E'\to E)\phi_0(\mathbf{x},E')\phi_0^*(\mathbf{x},E)$$
$$=-\int_{\partial\mathscr{D}_0}dS\,X(\mathbf{x})\int_E dE\,D\frac{\partial\phi_0}{\partial n_0}\frac{\partial\phi_0^*}{\partial n_0}. \tag{59}$$

From Eq. (58), one can also obtain the generalized form of Eq. (49) for the special case of spherical symmetry:

$$\Lambda_1\int_{\mathscr{D}_0}d\mathbf{r}\int_E dE\int_{E'}dE'\,\nu(\mathbf{r},E')\sigma_f(\mathbf{r};E'\to E)\phi_0(\mathbf{r},E)\phi_0^*(\mathbf{r},E)$$
$$=-\delta V\int dE\,D(r_0,E)$$
$$\times\left\{\left(\frac{\partial\phi_0^*}{\partial r}+\frac{\partial}{\partial r}\gamma\frac{\partial\phi_0^*}{\partial r}\right)\frac{\partial\phi_0}{\partial r}\right\}_{r=r_0}, \tag{60}$$

where δV is the same as defined in the previous section.

If a multigroup formulation with G groups were taken as the fundamental diffusion description,[24] the analysis presented here can be repeated within that context, and one obtains the result for the first-order eigenvalue change

$$\Lambda_1\int_{\mathscr{D}_0}d\mathbf{x}\sum_{g=1}^{G}\sum_{g'=1}^{G}\nu_{g'}(\mathbf{x})\sigma_{fg'\to g}(\mathbf{x})\phi_{0g'}(\mathbf{x})\phi_{0g}^*(\mathbf{x})$$
$$=-\int_{\partial\mathscr{D}_0}dS\,X(\mathbf{x})\sum_{g=1}^{G}D_g(\mathbf{x})\frac{\partial\phi_{0g}(\mathbf{x})}{\partial n_0}$$
$$\times\left\{\left(1+\frac{\partial\gamma_g(\mathbf{x})}{\partial n_0}\right)\frac{\partial\phi_{0g}^*(\mathbf{x})}{\partial n_0}\right.$$
$$\left.+\gamma_g(\mathbf{x})\left(\frac{\partial^2\phi_{0g}^*(\mathbf{x})}{\partial n_0^2}+|\mathbf{t}_0|\frac{\partial\phi_{0g}^*(\mathbf{x})}{\partial t_0}\right)\right\}. \tag{61}$$

To obtain Eq. (61) from the multigroup diffusion equations, one follows the procedure outlined in the preceding section with the integrations over energies replaced with summations over all groups (g's).

B. Method 2

In Sec. III A we developed a generalized first-order perturbation method to obtain a formula for the eigenvalue change due to a boundary perturbation. The intent here is to verify that formula by applying a totally different method.

As was mentioned earlier in Sec. II, the subject of boundary perturbation was of some interest in the fields of acoustics and quantum mechanics. One of the previous papers in this field is due to Feshbach.[10] In the analysis to follow, we make an attempt to extend Feshbach's method to the field of reactor physics, in particular diffusion theory in this section. The discussion is first restricted to the one-speed model for simplicity. The generalization of the formalism, to energy-dependent theory, is then outlined to complete the discussion.

We rewrite Eq. (33) as

$$H_0\phi_0(\mathbf{x}) = \Lambda_0 F_0\phi_0(\mathbf{x}), \quad \mathbf{x}\in\mathscr{D}_0, \tag{62}$$

with the same boundary condition as given by Eq. (35). We want to determine the eigenvalue Λ corresponding to the same diffusion operator defined in a domain $\mathscr{D}$ bounded by a slightly different boundary $\partial\mathscr{D}$:

$$H_0\phi(\mathbf{x}) = \Lambda F_0\phi(\mathbf{x}), \quad \mathbf{x}\in\mathscr{D}, \tag{63}$$

where the operators H_0 and F_0 are defined as

$$H_0 \equiv -\nabla\cdot D(\mathbf{x})\nabla + \sigma_a(\mathbf{x}), \tag{64}$$

$$F_0 \equiv \nu(\mathbf{x})\sigma_f(\mathbf{x}), \tag{65}$$

and Eq. (36) is to be satisfied as the boundary condition on $\phi(\mathbf{x})$.

We suppose the eigenvalue Λ_0 and the solution $\phi_0(\mathbf{x})$ are known, and we seek to find Λ to first order. Following the method of Feshbach, we let T be some operator, to be determined later, and define a new function $\eta(\mathbf{x})$ by

$$\eta(\mathbf{x}) = T\phi(\mathbf{x}). \tag{66}$$

T must be such an operator that $\eta(\mathbf{x})$ satisfies the same boundary condition (as ϕ_0) on $\partial\mathscr{D}_0$:

$$\eta(\mathbf{x}_{s_0}) + \gamma(\mathbf{x}_{s_0})\frac{\partial\eta(\mathbf{x}_{s_0})}{\partial n_0} = 0, \quad \mathbf{x}_{s_0}\in\partial\mathscr{D}_0. \tag{67}$$

Suppose that T can be written as

$$T = I + \epsilon P_1 + O(\epsilon^2), \tag{68}$$

where I is the identity operator, ϵ is the smallness parameter, and P_1 is some operator which is unknown at this point. P_1 will indirectly be determined from the boundary conditions. From Eqs. (63) and (66), the equation satisfied by $\eta(\mathbf{x})$ is given by

$$[H_0 + \epsilon H_1 + O(\epsilon^2)]\eta(\mathbf{x}) = \Lambda\,[F_0 + \epsilon F_1 + O(\epsilon^2)]\eta(\mathbf{x}), \quad \mathbf{x}\in\mathscr{D}_0, \tag{69}$$

where

$$H_1 = P_1H_0 - H_0P_1 \tag{70}$$

and

$$F_1 = P_1F_0 - F_0P_1. \tag{71}$$

Equations (67) and (69) constitute the perturbed problem in which ordinary perturbation methods may now be applied. One obtains the formula

$$\Lambda_1\int_{\mathscr{D}_0} d\mathbf{x}\,\nu(\mathbf{x})\sigma_f(\mathbf{x})\phi_0^2(\mathbf{x}) = -\int_{\mathscr{D}_0} d\mathbf{x}\,\phi_0(H_1 - \Lambda_0F_0)P_1\phi_0, \tag{72}$$

in which we have used the boundary conditions on η and ϕ_0. We have also assumed that the following can be expressed by series expansion in ϵ:

$$\Lambda = \Lambda_0 + \epsilon\Lambda_1 + O(\epsilon^2) \tag{73}$$

and

$$\eta(\mathbf{x}) = \phi_0(\mathbf{x}) + \epsilon\phi_1(\mathbf{x}) + O(\epsilon^2). \tag{74}$$

Equation (72) can further be reduced by using Eqs. (35) and (62) and the divergence theorem. We obtain

$$\Lambda_1\int_{\mathscr{D}_0} d\mathbf{x}\,\nu(\mathbf{x})\sigma_f(\mathbf{x})\phi_0^2(\mathbf{x}) = -\int_{\partial\mathscr{D}_0} dS\,D(\mathbf{x})\frac{\partial\phi_0(\mathbf{x})}{\partial n_0}\left[P_1\phi_0 + \gamma\frac{\partial}{\partial n_0}(P_1\phi_0)\right]. \tag{75}$$

In order to determine Λ_1, one needs to find P_0 (the only unknown in the equation) from the boundary condition on $\phi(\mathbf{x})$, which has not yet been used.

Define $X_i(x_i)$ such that if x_i represents the coordinate of a point on $\partial\mathscr{D}_0$, then to first order $x_i + \epsilon X_i(x_i)$ gives the coordinates of a point on $\partial\mathscr{D}$. Define an operator R (displacement operator) such that

$$\begin{aligned} Rg(\mathbf{x}) &= g(\mathbf{x} + \epsilon\mathbf{X}) \\ &= g(\mathbf{x}) + \epsilon\mathbf{X}\cdot\nabla g(\mathbf{x}) + O(\epsilon^2) \\ &= [I + \epsilon\mathbf{X}\cdot\nabla + O(\epsilon^2)]g(\mathbf{x}), \end{aligned} \tag{76}$$

where $g(\mathbf{x})$ is some smooth function. Hence, R can formally be written as

$$R = I + \epsilon\alpha + O(\epsilon^2), \tag{77}$$

where

$$\alpha \equiv \mathbf{X}\cdot\nabla. \tag{78}$$

As in Eq. (39), we choose $X(\mathbf{x})$ as

$$\mathbf{X}(\mathbf{x}) = X(\mathbf{x})\hat{n}_0. \tag{79}$$

Now, consider the boundary condition on ϕ, Eq. (36),

$$\phi(\mathbf{x}) + \gamma(\mathbf{x})\frac{\partial\phi(\mathbf{x})}{\partial n} = 0, \quad \mathbf{x}\in\partial\mathscr{D}. \tag{80}$$

From the definition on R, we have

$$R\phi(\mathbf{x}) + R\left[\gamma(\mathbf{x})\frac{\partial\phi(\mathbf{x})}{\partial n}\right] = 0, \quad \mathbf{x}\in\partial\mathscr{D}_0, \tag{81}$$

or

$$RT^{-1}\eta(\mathbf{x}) + R\left\{\gamma\frac{\partial}{\partial n}[T^{-1}\eta(\mathbf{x})]\right\} = 0, \quad \mathbf{x}\in\partial\mathscr{D}_0, \tag{82}$$

where

$$T^{-1}T = I. \tag{83}$$

Using Eqs. (42) and (74) and separating into orders, we find, in zeroth order,

$$\phi_0(\mathbf{x}) + \gamma(\mathbf{x})\frac{\partial\phi_0(\mathbf{x})}{\partial n_0} = 0, \quad \mathbf{x}\in\partial\mathscr{D}_0, \tag{84}$$

as expected, and, in first order,

$$\begin{aligned} &P_1\phi_0(\mathbf{x}) + \gamma(\mathbf{x})\frac{\partial}{\partial n_0}[P_1\phi_0(\mathbf{x})] \\ &= X(\mathbf{x})\left\{\left(1 + \frac{\partial\gamma(\mathbf{x})}{\partial n_0}\right)\frac{\partial\phi_0(\mathbf{x})}{\partial n_0} + \gamma(\mathbf{x})\left[\frac{\partial^2\phi_0(\mathbf{x})}{\partial n_0^2} + |\mathbf{t}_0|\frac{\partial\phi_0(\mathbf{x})}{\partial t_0}\right]\right\}, \quad \mathbf{x}\in\partial\mathscr{D}_0, \end{aligned} \tag{85}$$

where $\mathbf{t}_0$ is given by Eq. (43).

Substitution of Eq. (85) in (75) yields the desired formula,

$$\Lambda_1 \int_{\mathscr{D}_0} d\mathbf{x}\, \nu(\mathbf{x})\sigma_f(\mathbf{x})\phi_0^2(\mathbf{x}) = -\int_{\partial\mathscr{D}_0} dS\, X(\mathbf{x})D(\mathbf{x})\frac{\partial\phi_0(\mathbf{x})}{\partial n_0} \times\left\{\left(1+\frac{\partial\gamma(\mathbf{x})}{\partial n_0}\right)\frac{\partial\phi_0(\mathbf{x})}{\partial n_0}+\gamma(\mathbf{x}) \times\left[\frac{\partial^2\phi_0(\mathbf{x})}{\partial n_0^2}+|\mathbf{t}_0|\frac{\partial\phi_0(\mathbf{x})}{\partial t_0}\right]\right\}. \tag{86}$$

A comparison of this equation with that given in Eq. (47) reveals that both Methods 1 and 2 yield the same result for the eigenvalue change Λ_1 due to a boundary perturbation. Both methods seem to apply with equal ease to the diffusion equation. However, the method just developed (modified Feshbach and Clogston's method) is perhaps more involved algebraically but not conceptually. The advantage of this method is that it is less restrictive in that it only requires the analytic extension of the solution ϕ into the unperturbed domain $\mathscr{D}_0$, Eq. (66). The method presented in Sec. III A requires the analytic extension of the unperturbed solution ϕ_0 into the perturbed domain $\mathscr{D}$ as well.

The generalization of this method to energy-dependent theory is obvious. Starting with the energy dependent diffusion operator, as in Eq. (50) and its adjoint as in Eq. (54), one obtains the generalized version of Eq. (75)

$$\Lambda_1 \int_{\mathscr{D}_0} d\mathbf{x}\int_E dE\int_{E'} dE' \times\nu(\mathbf{x},E')\sigma_f(\mathbf{x};E'\to E)\phi_0^*(\mathbf{x},E)\phi_0(\mathbf{x},E') = -\int_{\partial\mathscr{D}_0} dS\int_E dE\, D(\mathbf{x},E)\frac{\partial\phi_0(\mathbf{x},E)}{\partial n_0} \times\left\{P_1\phi_0^*(\mathbf{x},E)+\gamma(\mathbf{x},E)\frac{\partial}{\partial n_0}[P_1\phi_0^*(\mathbf{x},E)]\right\}. \tag{87}$$

Using the displacement operator R and the boundary condition on ϕ^*, one can show, following the procedure just outlined,

$$P_1\phi_0^*(\mathbf{x},E)+\gamma(\mathbf{x},E)\frac{\partial}{\partial n_0}[P_1\phi_0^*(\mathbf{x},E)] = X(\mathbf{x})\left\{\left(1+\frac{\partial\gamma}{\partial n_0}\right)\frac{\partial\phi_0^*}{\partial n_0} +\gamma\left(\frac{\partial^2\phi_0^*}{\partial n_0^2}+|\mathbf{t}_0|\frac{\partial\phi_0^*}{\partial t_0}\right)\right\}_{(\mathbf{x},E)} + O(\epsilon^2),\quad \mathbf{x}\in\partial\mathscr{D}_0. \tag{88}$$

Hence,

$$\Lambda_1\int_{\mathscr{D}_0} d\mathbf{x}\int_E dE\int_{E'} dE' \times\nu(\mathbf{x},E')\sigma_f(\mathbf{x};E'\to E)\phi_0(\mathbf{x},E')\phi_0^*(\mathbf{x},E) = -\int_{\partial\mathscr{D}_0} dS\, X(\mathbf{x})\int_E dE\, D\frac{\partial\phi_0}{\partial n_0} \times\left[\left(1+\frac{\partial\gamma}{\partial n_0}\right)\frac{\partial\phi_0^*}{\partial n_0}+\gamma\left(\frac{\partial^2\phi_0^*}{\partial n_0^2}+|\mathbf{t}_0|\frac{\partial\phi_0^*}{\partial t_0}\right)\right]. \tag{89}$$

Comparing this equation with Eq. (58), we note that both Methods 1 and 2 yield the same result for the eigenvalue change. This is obviously what one would expect if the methods are to be correct. Using this method, we find that one would again obtain Eq. (61) in multigroup diffusion theory.

Thus far in Secs. III A and III B, we have presented two different methods with which one can obtain a first-order perturbation formula, in diffusion theory, for the change in the eigenvalue due to a boundary perturbation. In Sec. III C, we demonstrate the accuracy of the formalism by considering some examples. We must mention that Examples 1 and 2 have already been considered by Larsen and Pomraning.[3] However, here, we utilize the simple perturbation formula (49) and demonstrate its simplicity over the more complicated equation (47).

C. Examples

Consider the equations

$$\nabla^2\phi_0(\mathbf{x})+\Lambda_0\phi_0(\mathbf{x})=0,\quad \mathbf{x}\in\mathscr{D}_0, \tag{90}$$

and

$$\nabla^2\phi(\mathbf{x})+(\Lambda_0+\delta\Lambda)\phi(\mathbf{x})=0,\quad \mathbf{x}\in\mathscr{D}, \tag{91}$$

with Dirichlet boundary conditions

$$\phi_0(\mathbf{x})=0,\quad \mathbf{x}\in\partial\mathscr{D}_0, \tag{92}$$

and

$$\phi(\mathbf{x})=0,\quad \mathbf{x}\in\partial\mathscr{D}. \tag{93}$$

In this case, from Eq. (47), the perturbation formula becomes

$$\Lambda_1=\frac{\int_{\partial\mathscr{D}_0} dS\, X(\mathbf{x})(\partial\phi_0/\partial n_0)^2}{\int_{\mathscr{D}_0} d\mathbf{x}\,\phi_0^2}, \tag{94}$$

or, from Eq. (49),

$$\Lambda_1=\frac{-\delta V(\partial\phi_0/\partial n_0)^2_{r=r_0}}{\int_{\mathscr{D}_0} d\mathbf{r}\,\phi_0^2} \tag{95}$$

when ϕ_0 is spherically symmetric.

Example 1: In this example, we estimate the eigenvalue of a square of side 2 by taking a circle of radius r_0 as the unperturbed problem. In this case, δV as well as $X(\mathbf{x})$ are functions of r_0.

The unperturbed solution is

$$\phi_0(r)=J_0(Br/r_0), \tag{96}$$

with

$$\Lambda_0=(B/r_0)^2, \tag{97}$$

where $B=2.4048$ is the first zero of the zeroth-order Bessel function J_0. Using Eq. (94) or (95), one gets

$$\Lambda=(B/r_0)^2[3-(8/\pi r_0)\ln(1+\sqrt{2})]+\cdots \tag{98}$$

or

$$\Lambda=(B/r_0)^2(2-4/\pi r_0^2)+\cdots, \tag{99}$$

respectively. One might choose r_0 such that $\Lambda_1=0$. According to Eqs. (94) or (95), this leads to

TABLE I. Eigenvalue for a square, diffusion theory.

Unperturbed circle radius r_0	Λ_e (exact) = 4.9348 Λ_p (perturbation) Eq. (98)	Eq. (99)	$(\Lambda_e - \Lambda_p)/\Lambda_e$ Eq. (98)	Eq. (99)	$\lvert X(r_{s_0})\rvert_{max}/r_0$
1.0	4.3697	4.2029	0.12	0.18	0.17
1.1	4.5865	4.5296	0.07	0.08	0.08
1.1222	4.5922	4.5414	0.07	0.08	0.07
1.1284	4.5918	4.5420	0.07	0.08	0.06
1.2	4.5368	4.4811	0.08	0.09	0.04
1.2071	4.5272	4.4697	0.08	0.09	0.03
1.3	4.3580	4.2658	0.12	0.14	0.05

$$r_0 = (4/\pi)\ln(1+\sqrt{2}) = 1.1222 \tag{100}$$

or

$$r_0 = 2/\sqrt{\pi} = 1.1284. \tag{101}$$

The choice $r_0 = 2/\sqrt{\pi}$ corresponds to conserving area (the Wigner–Seitz approach, $\delta V = 0$). Other choices are: (1) minimize the maximum perturbation, $|X(r_{s_0})|_{max}$,

$$r_0 = (1+\sqrt{2})/2) = 1.2071, \tag{102}$$

and (2) set $d\Lambda/dr_0 = 0$ in Eqs. (98) or (99) which leads to Eqs. (100) or (101), respectively. Λ for various values of r_0 is given in Table I. The values for Λ_p in the first column are the same as those obtained by Larsen and Pomraning.[3]

Comparison with the exact eigenvalue shows that the best estimate is obtained with r_0 given by Eq. (100). The Wigner–Seitz results are almost as good, although those of Larsen and Pomraning are somewhat more accurate. However, neither of these results are especially good. This is due to the fact that the perturbation (circle to a square) is quite large, and any treatment only to first order in this perturbation is not likely to be very accurate. In the last three columns of Table I, we note that the relative error in Λ, $(\Lambda_e - \Lambda_p)/\Lambda_e$, is of the same order of magnitude as the relative maximum boundary perturbation, $|X(r_{s_0})|_{max}/r_0$.

Example 2: A more interesting example would be to consider estimating the eigenvalues of an ellipse. We take an ellipse with a semimajor axis of unity and an eccentricity e. The semiminor axis is then given by $(1-e^2)^{1/2}$. As the unperturbed system we choose a circle with radius r_0 in the range $(1-e^2)^{1/2} \leqslant r_0 \leqslant 1$. The endpoints of this inequality correspond to the circle completely enclosing the ellipse or completely enclosed by the ellipse. The unperturbed solution is given by Eqs. (96) and (97). The perturbation follows from Eq. (95), since the unperturbed solution is spherically symmetric. We have

$$\delta V = \pi[(1-e^2)^{1/2} - r_0^2]. \tag{103}$$

Hence,

$$\Lambda = (B/r_0)^2(2 - (1-e^2)^{1/2}/r_0^2 + \cdots). \tag{104}$$

If we determine r_0 by setting $\Lambda_1 = 0$, we obtain

$$r_0^2 = (1-e^2)^{1/2}. \tag{105}$$

Thus, Eq. (104) becomes

$$\Lambda = B^2/(1-e^2)^{1/2} + \cdots. \tag{106}$$

A comparison of this perturbation result with exact numerical results[3] is given in Table II. We see, as expected, that the accuracy of the estimate becomes worse as the perturbation gets larger. The results, however, are quite accurate for relatively large perturbations. Finally, in the last column of Table II, we have presented the results that Larsen and Pomraning obtained by using Eq. (94), which does not make use of the symmetry property of the unperturbed solution. In fact, Larsen and Pomraning used Eq. (94) and obtained the following, as shown in the table:

$$\Lambda = (B/r_0)^2[3 - (4/\pi r_0)K(e)(1-e^2)^{1/2}] + \cdots, \tag{107}$$

where $K(e)$ is the complete elliptic integral of the first kind, defined by

$$K(e) \equiv \int_0^{\pi/2} (1 - e^2\cos^2\theta)^{-1/2}\, d\theta. \tag{108}$$

The results obtained by Larsen and Pomraning are less accurate (for this particular problem) than those obtained here. However, one can easily show that Eq. (107) is equivalent to Eq. (106) to $O(\epsilon^2)$.

Example 3: In this example, we consider a more realistic problem concerned with the gas cooled fast reactor (GCFR)[25] in one-dimensional slab geometry. The unperturbed problem corresponds to a 112 cm thick core surrounded by a 5 cm thick axial blanket. The perturbed problem corresponds to adding (or deleting) a layer of blanket material of thickness ϵ. We must mention that the designed

TABLE II. Eigenvalues for an ellipse, diffusion theory.

Eccentricity e	$\Lambda_p^{1/2}$ (perturbation) Eq. (106)	$\Lambda_e^{1/2}$ (exact)	$\Lambda_p^{1/2}$ (perturbation) Eq. (107)
0	2.405	2.405	2.405
0.1	2.411	2.411	2.411
0.2	2.430	2.430	2.430
0.3	2.462	2.463	2.464
0.4	2.512	2.513	2.516
0.5	2.584	2.588	2.593
0.6	2.689	2.697	2.720
0.7	2.846	2.866	2.921
0.8	3.105	3.156	3.293
0.85	3.313	3.394	3.628
0.90	3.642	3.800	4.214
0.95	4.117	4.417	5.183

(optimum) thickness of the blanket is 45 cm (which would have led to very small eigenvalue perturbations if taken as the reference blanket thickness). The core is a homogeneous mixture of plutonium and uranium oxide, stainless steel, and boron carbide (as the control material), and the blankets are composed of natural uranium oxide, stainless steel, and helium. For this system, Eq. (61) gives

$$\Lambda_1 N = -\epsilon \sum_{g=1}^{3} \times \left\{ D_g \frac{d\phi_{0g}}{dx} \left(\frac{d\phi^*_{0g}}{dx} + \gamma_g \frac{d^2\phi^*_{0g}}{dx^2} \right) \right\}_{\substack{\text{unperturbed} \\ \text{boundary } (x=61)}}, \quad (109)$$

where

$$N \equiv \int_0^{61} dx \sum_{g=1}^{3} \chi_g \phi^*_{0g} \sum_{g'=1}^{3} \nu_{g'} \sigma_{fg'} \phi_{0g'} \quad (110)$$

and

$$\gamma_g = \sqrt{3} D_g. \quad (111)$$

The three group constants are taken from ANL-5800,[26] Table 7.5. A comparison of the "exact" discrete ordinates[27] S-2 (with isotropic scattering) eigenvalues with Eq. (109) is given in Table III. [We must emphasize that in slab geometry the S-2 calculation is equivalent to the diffusion approximation with Mark boundary condition (111).] It is seen that the perturbation formula has predicted quite good results (see column 3 in Table III).

IV. ANALYSIS IN TRANSPORT THEORY

In this section, we consider the energy dependent transport eigenvalue equation and modify Methods 1 and 2, as described in Sec. III, to obtain a first-order perturbation formula for the change in the eigenvalue due to a boundary perturbation. We then show that the formalism suggested by Larsen and Pomraning is a special case of our formula (the one-speed model). The analysis is concluded by considering some numerical examples in Sec. III C.

TABLE III. Eigenvalue for the reflected GCFR,[a] diffusion theory.

Additional blanket ϵ	Λ_p (perturbation)	Λ_e (exact)	% error $[100(\Lambda_e - \Lambda_p)/\Lambda_e]$
−5	1.0417	1.0531	1.08
−4	1.0332	1.0402	0.67
−3	1.0247	1.0285	0.37
−2	1.0162	1.0178	0.16
−1	1.0077	1.0081	0.04
−0.5	1.0035	1.0035	—
0	0.9992	0.9992	—
0.5	0.9950	0.9950	—
1	0.9907	0.9911	0.04
2	0.9822	0.9837	0.15
3	0.9737	0.9769	0.33
4	0.9652	0.9707	0.57
5	0.9567	0.9651	0.87
7	0.9397	0.9551	1.61

[a] The number density of B-10 in this reactor is taken as 7.3×10^{20} cm^{-3} to obtain a roughly critical unperturbed reactor.

A. Method 1

Consider the transport eigenvalue equation[28,29] in a domain $\mathscr{D}_0$ bounded by a smooth surface $\partial\mathscr{D}_0$:

$$\hat{\Omega}\cdot\nabla\psi_0(\mathbf{x},E,\hat{\Omega}) + \sigma(\mathbf{x},E)\psi_0(\mathbf{x},E,\hat{\Omega}) = \int_{E'} dE' \int_{4\pi} d\hat{\Omega}' \left[\sigma_s(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega}) + \Lambda_0\nu(\mathbf{x},E')\sigma_f(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega})\right]\psi_0(\mathbf{x},E',\hat{\Omega}'), \quad \mathbf{x}\in\mathscr{D}_0, \quad (112)$$

where $\psi_0(\mathbf{x},E,\hat{\Omega})$ is the angular flux, $\mathbf{x}$ is the neutron position vector, E is the energy variable, and $\hat{\Omega}$ is the unit vector in the direction of neutron motion. We now wish to determine the eigenvalue $(\Lambda + \delta\Lambda)$ corresponding to the same transport operator with a slightly different boundary $\partial\mathscr{D}$:

$$\hat{\Omega}\cdot\nabla\psi(\mathbf{x},E,\hat{\Omega}) + \sigma(\mathbf{x},E)\psi(\mathbf{x},E,\hat{\Omega}) = \int_{E'} dE' \int_{4\pi} d\hat{\Omega}' \left[\sigma_s(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega}) + (\Lambda_0+\delta\Lambda)\nu(\mathbf{x},E') \times \sigma_f(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega})\right]\psi(\mathbf{x},E',\hat{\Omega}'), \quad \mathbf{x}\in\mathscr{D}, \quad (113)$$

where $\psi(\mathbf{x},E,\hat{\Omega})$ is the perturbed angular flux and $\delta\Lambda$ is the perturbation in the eigenvalue due to the slight change in the boundary. $\sigma(\mathbf{x},E)$, $\sigma_s(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega})$, and $\nu(\mathbf{x},E')\sigma_f(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega})$ are smooth functions of $\mathbf{x}$ defined in a domain containing both $\mathscr{D}_0$ and $\mathscr{D}$ $(\mathscr{D}_0\cup\mathscr{D})$. These functions are assumed to be unchanged due to the boundary perturbation.

As boundary conditions, we demand that there be no incoming neutrons for all positions $\mathbf{x}$ on the boundary surfaces $\partial\mathscr{D}_0$ and $\partial\mathscr{D}$, respectively.

$$\psi_0(\mathbf{x}_{s_0},E,\hat{\Omega}) = 0 \quad \text{for all } \mathbf{x}_{s_0}\in\partial\mathscr{D}_0 \text{ and } \hat{n}_0\cdot\hat{\Omega} < 0 \quad (114)$$

and

$$\psi(\mathbf{x}_s,E,\hat{\Omega}) = 0 \quad \text{for all } \mathbf{x}_s\in\partial\mathscr{D} \text{ and } \hat{n}\cdot\hat{\Omega} < 0. \quad (115)$$

We assume the eigenvalue Λ_0 and the solution $\psi_0(\mathbf{x},E,\hat{\Omega})$ are known, and seek to find $\delta\Lambda$ to first order. Since the analysis here is similar to that in Sec. III, we shall present it with a minimum of discussion. Multiply the adjoint to the unperturbed equation and Eq. (113) by $\psi(\mathbf{x},E,\hat{\Omega})$ and $\psi^*_0(\mathbf{x},E,\hat{\Omega})$, the adjoint angular flux, respectively, and subtract. Integrate the resulting equation over all $\mathbf{x}\in\mathscr{D}$, E, and $\hat{\Omega}$ and then use the divergence theorem and the boundary condition on $\psi(\mathbf{x},E,\hat{\Omega})$ to get the following exact formula for $\delta\Lambda$:

$$\delta\Lambda \int_{\mathscr{D}} d\mathbf{x} \int_E dE \int_{E'} dE' \int_{4\pi} d\hat{\Omega}' \int_{4\pi} d\hat{\Omega} \times \nu(\mathbf{x},E')\sigma_f(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega})\psi^*_0(\mathbf{x},E,\hat{\Omega})\psi(\mathbf{x},E',\hat{\Omega}') = \int_{\partial\mathscr{D}} dS \int_E dE \int_{n\cdot\hat{\Omega}>0} d\hat{\Omega}\, \hat{n}\cdot\hat{\Omega}\psi^*_0(\mathbf{x},E,\hat{\Omega})\psi(\mathbf{x},E,\hat{\Omega}), \quad (116)$$

where $\psi^*_0(\mathbf{x},E,\hat{\Omega})$ satisfies

$$-\hat{\Omega}\cdot\nabla\psi^*_0(\mathbf{x},E,\hat{\Omega}) + \sigma(\mathbf{x},E)\psi^*_0(\mathbf{x},E,\hat{\Omega}) = \int_{E'} dE' \int_{4\pi} d\hat{\Omega}' \left[\sigma_s(\mathbf{x};E,\hat{\Omega}\to E',\hat{\Omega}') + \Lambda_0\nu(\mathbf{x},E)\sigma_f(\mathbf{x};E,\hat{\Omega}\to E',\hat{\Omega}')\right]\psi^*_0(\mathbf{x},E',\hat{\Omega}'), \quad \mathbf{x}\in\mathscr{D}_0, \quad (117)$$

and

$$\psi_0^*(\mathbf{x}_{s_0},E,\hat{\Omega})=0 \quad \text{for all } \mathbf{x}_{s_0}\in\partial\mathscr{D}_0 \text{ and } \hat{n}_0\cdot\hat{\Omega}>0. \tag{118}$$

In view of the arguments given in Sec. III, we assume that $\psi_0^*(\mathbf{x},E,\hat{\Omega})$ and $\psi(\mathbf{x},E,\hat{\Omega})$ can be analytically extended into $\mathscr{D}$ and $\mathscr{D}_0$, respectively, while satisfying their respective integro-differential equations.

Using similar arguments as in Sec. III A, Eqs. (39)–(43), one can show that

$$\begin{aligned}[\hat{n}\cdot\hat{\Omega}\psi_0^*\psi]_{(\mathbf{x},E,\hat{\Omega})=(\mathbf{x}_s,E,\hat{\Omega})}&=\{\hat{n}_0\cdot\Omega\psi_0^*\psi+\epsilon X(\mathbf{x}_{s_0})[\hat{n}_0\cdot\hat{\Omega}(\psi_0^* n_0\cdot\nabla\psi+\psi\hat{n}_0\cdot\nabla\psi_0^*)\\&\quad+\mathbf{t}_0\cdot\hat{\Omega}\psi_0^*\psi]\}_{(\mathbf{x},E,\hat{\Omega})=(\mathbf{x}_{s_0},E,\hat{\Omega})}+O(\epsilon^2).\end{aligned} \tag{119}$$

Substitute Eq. (119) in Eq. (116) to get

$$\begin{aligned}&\delta\Lambda\int_{\mathscr{D}}d\mathbf{x}\int_E dE\int_{E'}dE'\int_{4\pi}d\hat{\Omega}\int_{4\pi}d\hat{\Omega}'\\&\times\nu(\mathbf{x},E)\sigma_f(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega})\psi_0^*(\mathbf{x},E,\hat{\Omega})\psi(\mathbf{x},E',\hat{\Omega}')\\&=\int_{\partial\mathscr{D}}dS\int_E dE\int_{\hat{n}\cdot\hat{\Omega}>0}d\hat{\Omega}\\&\quad\times\{\hat{n}_0\cdot\hat{\Omega}\psi_0^*\psi+\epsilon X(\mathbf{x}_{s_0})[\hat{n}_0\cdot\hat{\Omega}(\psi_0^*\hat{n}_0\cdot\nabla\psi+\psi\hat{n}_0\cdot\nabla\psi_0^*)\\&\quad+\mathbf{t}_0\cdot\hat{\Omega}\psi_0^*]\}_{(\mathbf{x}_{s_0},E,\hat{\Omega})}+O(\epsilon^2).\end{aligned} \tag{120}$$

We note that the pairs (ψ_0,ψ), (n_0,n), $(\partial\mathscr{D},\partial\mathscr{D}_0)$, and $(\mathscr{D}_0,\mathscr{D})$ all differ by order ϵ, and hence we can replace ψ, n, $\partial\mathscr{D}$, and $\mathscr{D}$, in the above equation by ψ_0, n_0, $\partial\mathscr{D}_0$, and $\mathscr{D}_0$. This leads to, after making use of the boundary condition on ψ_0^*, Eq. (118),

$$\begin{aligned}&\Lambda_1\int_{\mathscr{D}_0}d\mathbf{x}\int_E dE\int_{E'}dE'\int_{4\pi}d\hat{\Omega}'\int_{4\pi}d\hat{\Omega}\\&\quad\times\nu(\mathbf{x},E')\sigma_f(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega})\psi_0^*(\mathbf{x},E,\hat{\Omega})\psi_0(\mathbf{x},E',\hat{\Omega}')\\&=\int_{\partial\mathscr{D}_0}dS\,X(\mathbf{x})\int_E dE\int_{\hat{n}_0\cdot\hat{\Omega}>0}d\hat{\Omega}\\&\quad\times\hat{n}_0\cdot\hat{\Omega}\psi_0(\mathbf{x},E,\hat{\Omega})\hat{n}_0\cdot\nabla\psi_0^*(\mathbf{x},E,\hat{\Omega}).\end{aligned} \tag{121}$$

Λ_1 represents the first-order change in the eigenvalue due to the perturbation in the boundary [see Eq. (48)]. This equation can further be reduced by noting that, for all $\mathbf{x}\in\partial\mathscr{D}_0$ and $\hat{n}_0\cdot\hat{\Omega}>0$, Eqs. (117) and (118) imply

$$\begin{aligned}(\hat{n}_0\cdot\hat{\Omega})\hat{n}_0\cdot\nabla\psi_0^*(\mathbf{x},E,\hat{\Omega})&=\hat{\Omega}\cdot\nabla\psi_0^*(\mathbf{x},E,\hat{\Omega})\\&=-\int_{E'}dE'\int_E dE\,[\sigma_s(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega})\\&\quad+\Lambda_0\nu(\mathbf{x},E)\sigma_f(\mathbf{x};E,\hat{\Omega}\to E',\hat{\Omega}')\psi_0^*(\mathbf{x},E',\hat{\Omega}'),\\&\mathbf{x}\in\partial\mathscr{D}_0,\ \hat{n}_0\cdot\hat{\Omega}>0.\end{aligned} \tag{122}$$

Substitution of Eq. (122) in Eq. (121) yields the end result

$$\begin{aligned}&\Lambda_1\int_{\mathscr{D}_0}d\mathbf{x}\int_E dE\int_{E'}dE'\,W_f(\mathbf{x};E'\to E)\\&=-\int_{\partial\mathscr{D}_0}dS\,X(\mathbf{x})\int_E dE\int_{E'}dE'\,[W_s(\mathbf{x};E'\to E)\\&\quad+\Lambda_0 W_f(\mathbf{x};E'\to E)],\end{aligned} \tag{123}$$

where

$$\begin{aligned}W_s(\mathbf{x};E'\to E)\equiv\int_{4\pi}d\hat{\Omega}\,\psi_0^*(\mathbf{x},E,\hat{\Omega})\int_{4\pi}d\hat{\Omega}'\\\times\sigma_s(\mathbf{x},E',\hat{\Omega}'\to E,\hat{\Omega})\psi_0(\mathbf{x},E',\hat{\Omega}')\end{aligned} \tag{124}$$

and

$$\begin{aligned}W_f(\mathbf{x};E'\to E)\equiv\int_{4\pi}d\hat{\Omega}\,\psi_0^*(\mathbf{x},E,\hat{\Omega})\int_{4\pi}d\hat{\Omega}'\\\times\nu(\mathbf{x},E')\sigma_f(\mathbf{x};E'\hat{\Omega}'\to E,\hat{\Omega})\psi_0(\mathbf{x},E',\hat{\Omega}').\end{aligned} \tag{125}$$

The analysis presented here can be repeated within the multigroup formulation context, with G groups, to obtain

$$\begin{aligned}&\Lambda_1\int_{\mathscr{D}_0}d\mathbf{x}\sum_{g=1}^{G}\sum_{g'=1}^{G}W_{f_{g'\to g}}(\mathbf{x})\\&=-\int_{\partial\mathscr{D}_0}dS\,X(\mathbf{x})\sum_{g=1}^{G}\sum_{g'=1}^{G}\\&\quad\times[W_{s_{g'\to g}}(\mathbf{x})+\Lambda_0 W_{f_{g'\to g}}(\mathbf{x})],\end{aligned} \tag{126}$$

where

$$\begin{aligned}W_{s_{g'\to g}}(\mathbf{x})&\equiv\int_{4\pi}d\hat{\Omega}\,\psi_{0g}^*(\mathbf{x},\hat{\Omega})\int_{4\pi}d\hat{\Omega}'\\&\quad\times\sigma_{s_{g'\to g}}(\mathbf{x};\hat{\Omega}'\to\hat{\Omega})\psi_{0g'}(\mathbf{x},\hat{\Omega}')\end{aligned} \tag{127}$$

and

$$\begin{aligned}W_{f_{g'\to g}}(\mathbf{x})&\equiv\int_{4\pi}d\hat{\Omega}\,\psi_{0g}^*(\mathbf{x},\hat{\Omega})\int_{4\pi}d\hat{\Omega}\,\nu_{g'}(\mathbf{x})\\&\quad\times\sigma_{f_{g'\to g}}(\mathbf{x};\hat{\Omega}'\to\hat{\Omega})\psi_{0g'}(\mathbf{x},\hat{\Omega}').\end{aligned} \tag{128}$$

To derive this equation from the multigroup transport equation,[30] one follows the procedure just outlined with the integrations over energies replaced with summations over all groups (g's).

In the case of one-speed (one-group) transport theory ψ and ψ^* are energy-independent and hence both the continuous and the multigroup formulation for Λ_1, Eqs. (123) and (126) reduce to the following equation:

$$\Lambda_1\int_{\mathscr{D}_0}d\mathbf{x}\,W_f(\mathbf{x})=-\int_{\partial\mathscr{D}_0}dS\,X(\mathbf{x})[W_s(\mathbf{x})+\Lambda_0 W_f(\mathbf{x})], \tag{129}$$

where

$$W_s(\mathbf{x})\equiv\int_{4\pi}d\hat{\Omega}\,\psi_0^*(\mathbf{x},\hat{\Omega})\int_{4\pi}d\hat{\Omega}'\,\sigma_s(\mathbf{x};\hat{\Omega}'\to\hat{\Omega})\psi_0(\mathbf{x},\hat{\Omega}') \tag{130}$$

and

$$W_f(\mathbf{x})\equiv\int_{4\pi}d\hat{\Omega}\,\psi_0^*(\mathbf{x},\hat{\Omega})\int_{4\pi}d\hat{\Omega}'\,\nu(\mathbf{x})\sigma_f(\mathbf{x};\hat{\Omega}'\to\hat{\Omega})\psi_0(\mathbf{x},\hat{\Omega}'). \tag{131}$$

Equation (129) is exactly the same result as that obtained by Larsen and Pomraning, who employed a quite different perturbation method.

B. Method 2

In this section we want to rederive Eq. (123) by extending Method 2 of Sec. III B to transport theory. The analysis here is very similar to that in Sec. III B, and shall be presented with a minimum of discussion.

We rewrite Eqs. (117) and (118), the adjoint unperturbed problem, as

$$A_0^*\psi_0^*(\mathbf{x},E,\hat{\Omega}) = \Lambda_0 B_0^*\psi_0^*(\mathbf{x},E,\hat{\Omega}), \quad \mathbf{x}\in\mathscr{D}_0, \tag{132}$$

and

$$\psi_0^*(\mathbf{x}_{s_0},E,\hat{\Omega}) = 0 \quad \text{for all } \mathbf{x}\in\partial\mathscr{D}_0 \text{ and } \hat{n}_0\cdot\hat{\Omega} > 0, \tag{133}$$

where

$$A_0^* \equiv -\hat{\Omega}\cdot\nabla + \sigma(\mathbf{x},E) - \int_{E'} dE' \int_{4\pi} d\hat{\Omega}'\, \sigma_s(\mathbf{x};E,\hat{\Omega}\to E',\hat{\Omega}')(\cdot), \tag{134}$$

and

$$B_0^* \equiv \int_{E'} dE' \int_{4\pi} d\hat{\Omega}'\, \nu(\mathbf{x},E)\sigma_f(\mathbf{x};E,\hat{\Omega}\to E',\hat{\Omega}')(\cdot). \tag{135}$$

Similarly, the perturbed problem is obtained by rewriting Eqs. (113) and (115)

$$A_0\psi(\mathbf{x},E,\hat{\Omega}) = \Lambda B_0\psi(\mathbf{x},E,\hat{\Omega}), \quad \mathbf{x}\in\mathscr{D}, \tag{136}$$

and

$$\psi(\mathbf{x}_s,E,\hat{\Omega}) = 0 \quad \text{for all } \mathbf{x}_s\in\partial\mathscr{D} \text{ and } \hat{n}\cdot\hat{\Omega} < 0, \tag{137}$$

where

$$A_0 \equiv \hat{\Omega}\cdot\nabla + \sigma(\mathbf{x},E) - \int_{E'} dE' \int_{4\pi} d\hat{\Omega}'\, \sigma_s(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega})(\cdot) \tag{138}$$

and

$$B_0 \equiv \int_{E'} dE' \int_{4\pi} d\hat{\Omega}'\, \nu(\mathbf{x},E')\sigma_f(\mathbf{x};E'\hat{\Omega}'\to E,\hat{\Omega})(\cdot). \tag{139}$$

We suppose that Λ_0 and $\psi_0^*(\mathbf{x},E,\hat{\Omega})$ are known, and we seek to find Λ to first order. Following the analysis given in Sec. III B, we define $\zeta(\mathbf{x},E,\hat{\Omega})$ by

$$\zeta(\mathbf{x},E,\hat{\Omega}) = \kappa\psi(\mathbf{x},E,\hat{\Omega}), \tag{140}$$

where the operator κ is such that

$$\zeta(\mathbf{x}_{s_0},E,\hat{\Omega}) = 0 \quad \text{for all } \mathbf{x}_{s_0}\in\partial\mathscr{D}_0 \text{ and } \hat{n}_0\cdot\hat{\Omega} < 0. \tag{141}$$

Assume that κ can be written as

$$\kappa = I + \epsilon Q_1 + O(\epsilon^2); \tag{142}$$

Q_1 is some operator which is unknown at this time. Q_1 will indirectly be determined from the boundary condition on $\psi(\mathbf{x},E,\hat{\Omega})$. Substituting Eq. (140) in (136), we get

$$(A_0 + \epsilon A_1)\zeta(\mathbf{x},E,\hat{\Omega}) + O(\epsilon^2) = \Lambda(B_0 + \epsilon B_1)\zeta(\mathbf{x},E,\hat{\Omega}) + O(\epsilon^2), \quad \mathbf{x}\in\mathscr{D}_0, \tag{143}$$

where

$$A_1 = Q_1A_0 - A_0Q_1 \tag{144}$$

and

$$B_1 = Q_1B_0 - B_0Q_1, \tag{145}$$

which constitute the perturbed problem in terms of $\zeta(\mathbf{x},E,\hat{\Omega})$. Applying ordinary perturbation methods, one obtains the following formula for Λ_1:

$$\Lambda_1 \int_{\mathscr{D}_0} d\mathbf{x} \int_E dE \int_{E'} dE' \int_{4\pi} d\hat{\Omega} \int_{4\pi} d\hat{\Omega}' \times \nu(\mathbf{x},E')\sigma_f(\mathbf{x};E'\hat{\Omega}'\to E,\hat{\Omega})\psi_0^*(\mathbf{x},E,\hat{\Omega})\psi_0(\mathbf{x},E',\hat{\Omega}') = -\int_{\mathscr{D}_0} d\mathbf{x} \int_E dE \int_{4\pi} d\hat{\Omega}\, \psi_0^*(\mathbf{x},E,\hat{\Omega}) \times (A_0 - \Lambda_0B_0)Q_1\psi_0(\mathbf{x},E,\hat{\Omega}), \tag{146}$$

in which we have used the boundary condition on ζ and ψ_0^* and have assumed that Λ and ζ can be expressed by series expansion in ϵ,

$$\Lambda = \Lambda_0 + \epsilon\Lambda_1 + O(\epsilon^2) \tag{147}$$

and

$$\zeta(\mathbf{x},E,\hat{\Omega}) = \psi_0(\mathbf{x},E,\hat{\Omega}) + \epsilon\psi_1(\mathbf{x},E,\hat{\Omega}) + O(\epsilon^2). \tag{148}$$

Equation (146) can further be reduced by using Eqs. (132) and Gauss theorem, i.e.,

$$\Lambda_1 \int_{\mathscr{D}_0} d\mathbf{x} \int_E dE \int_{E'} dE' \int_{4\pi} d\hat{\Omega} \int_{4\pi} d\hat{\Omega}' \times \nu(\mathbf{x},E')\sigma_f(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega})\psi_0^*(\mathbf{x},E,\hat{\Omega})\psi_0(\mathbf{x},E',\hat{\Omega}') = -\int_{\partial\mathscr{D}_0} dS \int_E dE \int_{4\pi} d\hat{\Omega}\, n_0\cdot\hat{\Omega} \times \psi_0^*(\mathbf{x},E,\hat{\Omega})Q_1\psi_0(\mathbf{x},E,\hat{\Omega}). \tag{149}$$

We now must find $Q_1\psi_0(\mathbf{x},E,\hat{\Omega})$ on $\partial\mathscr{D}_0$ in order to obtain the end result. This is done by considering the boundary condition on ψ, which is as yet unused,

$$\psi(\mathbf{x},E,\hat{\Omega}) = 0 \quad \text{for all } \mathbf{x}\in\partial\mathscr{D} \text{ and } \hat{n}\cdot\hat{\Omega} < 0. \tag{150}$$

But from the definition of the displacement operator R, Sec. III B, we have

$$R\psi(\mathbf{x},E,\hat{\Omega}) = 0 \quad \text{for all } \mathbf{x}\in\partial\mathscr{D}_0 \text{ and } \hat{n}_0\cdot\hat{\Omega} < 0 \tag{151}$$

or

$$R\kappa^{-1}\zeta(\mathbf{x},E,\hat{\Omega}) = 0 \quad \text{for all } \mathbf{x}\in\partial\mathscr{D}_0 \text{ and } \hat{n}_0\cdot\hat{\Omega} < 0, \tag{152}$$

where

$$\kappa\kappa^{-1} = I. \tag{153}$$

Using Eqs. (148) and (152), we find upon separating into orders, in zeroth order,

$$\psi_0(\mathbf{x},E,\hat{\Omega}) = 0 \quad \text{for all } \mathbf{x}\in\partial\mathscr{D}_0 \text{ and } \hat{n}_0\cdot\hat{\Omega} < 0, \tag{154}$$

as expected, and the first-order result

$$Q_1\psi_0(\mathbf{x},E,\hat{\Omega}) = X(\mathbf{x})\hat{n}_0\cdot\nabla\psi_0(\mathbf{x},E,\hat{\Omega}) \quad \text{for } \mathbf{x}\in\partial\mathscr{D}_0 \text{ and } \hat{n}_0\cdot\hat{\Omega} < 0. \tag{155}$$

Substituting Eq. (155) in Eq. (149), we get

$$\Lambda_1 \int_{\mathscr{D}_0} d\mathbf{x} \int_E dE \int_{E'} dE' \int_{4\pi} d\hat{\Omega} \int_{4\pi} d\hat{\Omega}' \times \nu(\mathbf{x},E')\sigma_f(\mathbf{x};E',\hat{\Omega}'\to E,\hat{\Omega})\psi_0^*(\mathbf{x},E,\hat{\Omega})\psi(\mathbf{x},E',\hat{\Omega}') = -\int_{\partial\mathscr{D}_0} dS \int_E dE \int_{\hat{n}_0\cdot\hat{\Omega}<0} \times d\hat{\Omega}\, X(\mathbf{x})\hat{n}_0\cdot\hat{\Omega}\psi_0^*(\mathbf{x},E,\hat{\Omega})\hat{n}_0\cdot\nabla\psi_0(\mathbf{x},E,\hat{\Omega}). \tag{156}$$

But from Eqs. (112) and (114) we have

$$(\hat{n}_0\cdot\hat{\Omega})[\hat{n}_0\cdot\nabla\psi_0(\mathbf{x},E,\hat{\Omega})] = \hat{\Omega}\cdot\nabla\psi_0(\mathbf{x},E,\hat{\Omega}) = \int_E dE \int_{4\pi} d\hat{\Omega}'\,[\sigma_s(\mathbf{x};E'\hat{\Omega}'\to E,\hat{\Omega}) + \Lambda_0\nu(\mathbf{x},E')\sigma_f(\mathbf{x};E'\hat{\Omega}'\to E,\hat{\Omega})] \times \psi_0(\mathbf{x},E',\hat{\Omega}') \quad \text{for all } \mathbf{x}\in\partial\mathscr{D}_0 \text{ and } \hat{n}_0\cdot\hat{\Omega} < 0. \tag{157}$$

Thus, Eq. (156) becomes

$$\Lambda_1 \int_{\mathscr{D}_0} d\mathbf{x} \int_E dE \int_{E'} dE'\, W_f(\mathbf{x};E'\to E)$$
$$= -\int_{\partial\mathscr{D}_0} dS\, X(\mathbf{x}) \int_E dE \int_{E'} dE'$$
$$\times [W_s(\mathbf{x};E'\to E) + \Lambda_0 W_f(\mathbf{x};E'\to E)], \quad (158)$$

where $W_s(\mathbf{x};E'\to E)$ and $W_f(\mathbf{x};E'\to E)$ are defined by Eqs. (124) and (125), respectively. As expected, Eq. (158) is identical with Eq. (123) which verifies that Method 2 yields the same results for Λ_1 as that obtained through Method 1. We also found the same agreement in the multigroup transport formulation.

It might be of interest to note that in the special case of isotropic scattering in one-speed theory both Eqs. (123) and (126) reduce to

$$\Lambda_1 \int_{\mathscr{D}_0} d\mathbf{x}\, \nu(\mathbf{x})\sigma_f(\mathbf{x})\phi_0^2(\mathbf{x})$$
$$= -\int_{\partial\mathscr{D}_0} dS\, X(\mathbf{x})[\sigma_s(\mathbf{x}) + \Lambda_0 \nu(\mathbf{x})\sigma_f(\mathbf{x})]\phi_0^2(\mathbf{x}), \quad (159)$$

where $\phi_0(\mathbf{x})$ is the unperturbed scalar flux,

$$\phi_0(\mathbf{x}) \equiv \int_{4\pi} d\hat{\Omega}\, \psi_0(\mathbf{x},\hat{\Omega}), \quad \mathbf{x}\in\mathscr{D}_0. \quad (160)$$

That is, the formula contains only the scalar flux and not the angular flux.

C. Examples

In this section we would like to present two transport theory example problems in which a discrete-ordinates (S-N)[31] approximation is assumed to perform the numerical computations. Since it is easily seen that the perturbation formulas (123) and (159) apply to the S-N equations as well as the exact transport equation, the eigenvalues reported here actually correspond to the S-N equations rather than the exact transport equation. We note that the S-N eigenvalues are good approximations to the exact eigenvalues for $N\gg 1$.

Example 1: We consider a two-dimensional problem in x–y geometry with isotropic scattering. The unperturbed problem is a bare square core that occupies $|x|\leqslant 1$ and $|y|\leqslant 1$ with $\sigma_a = 0.5$, $\nu\sigma_f = 1$, and $\sigma_s = 0.5$. We take this square with a thin layer of the same material added to one side as the perturbed problem. The layer occupies $|x|\leqslant \frac{1}{2}$ and $1\leqslant y < 1+\epsilon$. Since the boundary conditions are not to be perturbed, we let, in the perturbed square, an infinite absorber (i.e., σ_a is infinitely large) occupy $\frac{1}{2}\leqslant|x|\leqslant 1$ and $1\leqslant y\leqslant 1+\epsilon$ for $\epsilon\geqslant 0$ and $|x|\leqslant\frac{1}{2}$ and $1\leqslant y\leqslant 1+\epsilon$ for $\epsilon\leqslant 0$. The addition of the absorber leads to the unperturbed nonreentrant boundary condition in the perturbed problem. The "exact" as well as the perturbed eigenvalues were computed with the discrete-ordinates S-2 as well as S-4 approximations.[32] For this problem Eq. (159) gives

$$\Lambda = \Lambda_0 \left[1 - \frac{\epsilon \int_{-1/2}^{1/2} dx\, \phi_0^2(x,1)}{\int_{-1}^{1} dx \int_{-1}^{1} dy\, \phi_0^2(x,y)} \right] + \cdots \quad (161)$$

as the perturbed eigenvalue. In Table IV we give the "exact" as well as the perturbed eigenvalues for some values of ϵ, using S-2 as well as S-4 approximations. In the last column of this table, we explicitly show that the % error in the perturbed eigenvalue in the S-2 approximation is the same as that in the S-4 approximation. That is, within the numerical approximation (e.g., S-N), the error involved in the perturbation formula is second order.

Example 2: Finally, we conclude this section by considering a more realistic energy-dependent (three groups) example problem. In fact, we reconsider the one-dimensional (slab geometry) GCFR problem described in Sec. III C, Ex. 3. For this reactor Eq. (126) gives

$$\Lambda_1 N = -\epsilon \left\{ \sum_{g=1}^{3} \sum_{g'=1}^{3} \times (\sigma_{s_{g'\to g}} + \Lambda_0 \nu\sigma_{f_{g'\to g}}) \phi_{0g}^{*} \phi_{0g'} \right\}_{\substack{\text{unperturbed} \\ \text{boundary } (x=61)}}, \quad (162)$$

where

$$N = \int_0^{61} dx \sum_{g=1}^{3} \chi_g \phi_{0g}^{*} \sum_{g'=1}^{3} \nu_{g'}\sigma_{fg'}\phi_{0g'}. \quad (163)$$

A comparison of the exact discrete ordinates S-4 and S-2, both with isotropic scattering, eigenvalues[27] with the pertur-

TABLE IV. Eigenvalue for a perturbed square, transport theory.

Additional core ϵ	Λ_p (perturbation) S − 2	Λ_p (perturbation) S − 4	Λ_e (exact) S − 2	Λ_e (exact) S − 4	% error $[100(\Lambda_e-\Lambda_p)/\Lambda_e]$ S − 2	% error $[100(\Lambda_e-\Lambda_p)/\Lambda_e]$ S − 4
− 1.0	1.1950	1.1616	1.4569	1.3609	17.9	14.6
− 0.5	1.1390	1.1033	1.1848	1.1455	3.8	3.7
− 0.3	1.1166	1.0799	1.1335	1.0955	1.5	1.4
− 0.2	1.1054	1.0682	1.1128	1.0754	0.7	0.7
− 0.1	1.0942	1.0566	1.0961	1.0584	0.2	0.2
− 0.05	1.0886	1.0507	1.0891	1.0513	—	—
− 0.02	1.0852	1.0472	1.0853	1.0473	—	—
0	1.0830	1.0449	1.0830	1.0449	—	—
0.02	1.0808	1.0426	1.0809	1.0426	—	—
0.05	1.0774	1.0391	1.0777	1.0400	—	—
0.1	1.0718	1.0332	1.0730	1.0346	0.1	0.1
0.2	1.0606	1.0216	1.0651	1.0264	0.4	0.5
0.3	1.0494	1.0099	1.0590	1.0199	0.9	1.0
0.5	1.0270	0.9866	1.0510	1.0107	2.3	2.3
1.0	0.9710	0.9282	1.0446	0.9989	7.1	7.1

TABLE V. Eigenvalue for the reflected GCFR, transport theory.

Additional blanket ϵ	Λ_p (perturbation) S-2[a]	Λ_p (perturbation) S-4	Λ_e (exact) S-2[a]	Λ_e (exact) S-4	% error $[100(\Lambda_e - \Lambda_p)/\Lambda_e]$ S-2	% error $[100(\Lambda_e - \Lambda_p)/\Lambda_e]$ S-4
−5	1.0417	1.0231	1.0531	1.0337	1.08	1.03
−4	1.0332	1.0156	1.0402	1.0222	0.67	0.65
−3	1.0247	1.0081	1.0285	1.0117	0.37	0.36
−2	1.0162	1.0006	1.0178	1.0022	0.16	0.16
−1	1.0077	0.9931	1.0081	0.9935	0.04	0.04
−0.5	1.0035	0.9894	1.0035	0.9895	—	—
0	0.9992	0.9856	0.9992	0.9856	—	—
0.5	0.9950	0.9819	0.9950	0.9819	—	—
1	0.9907	0.9781	0.9911	0.9785	0.04	0.04
2	0.9822	0.9706	0.9837	0.9719	0.15	0.13
3	0.9737	0.9631	0.9769	0.9659	0.33	0.29
4	0.9652	0.9556	0.9707	0.9604	0.57	0.50
5	0.9567	0.9481	0.9651	0.9555	0.87	0.78
7	0.9397	0.9331	0.9551	0.9467	1.61	1.44

[a] Diffusion theory prediction with Mark boundary conditions.

bation results, Eq. (162), is given in Table V. The perturbation results are seen to be very accurate even for large values of ϵ. This is, however, misleading, since a comparison in terms of reactivity (Table VI) indicates that the error in estimating the reactivity is orders of magnitude larger than that in approximating the eigenvalue. In the third and sixth columns of both Tables V and VI, we explicitly show that the % error in the perturbation estimate in S-2 calculation is nearly the same as that in S-4 calculation.

V. DISCUSSION

The primary purpose of this paper has been to develop a first-order perturbation formula for the change in the eigenvalue due to a perturbation in the exterior boundary of a system. The motivation for this work was the recognized lack of energy-dependent first-order perturbation theory that can handle eigenvalue problems of the type mentioned above. The perturbation formalism was obtained in transport theory as well as in the diffusion approximation.

The major effort has been directed towards developing two different formalisms that would lead to the same perturbation formula for the eigenvalue change. In the first method, we have shown that the standard first-order perturbation theory can be extended to deal with changes in the boundary (exterior) of a system as well as with perturbation in the operator (reactor core composition). The second method is an extension of Feshbach and Clogston's idea,[10] in which they suggested the use of an inverse operator to convert boundary changes into a fictitious perturbation in the interior operator of the system. Using this inverse operator, it was shown that one obtains the same formula as that resulting from Method 1.

When applicable (i.e., in one-group theory) comparisons were made with Larsen and Pomraning's method,[3] which is the only other existing method in the field of reactor physics. It was concluded that Method 1, which is a natural extension of the first-order standard perturbation theory, is conceptually and algebraically less involved than the other

TABLE VI. Reactivity for the reflected GCFR, transport theory.

Additional blanket ϵ	Reactivity R_p in S[a] (perturbation) S-2[b]	Reactivity R_p in S[a] (perturbation) S-4	Reactivity R_e in S[a] (exact) S-2[b]	Reactivity R_e in S[a] (exact) S-4	% error $[100\lvert R_e - R_p\rvert/R_e]$ S-2	% error $[100\lvert R_e - R_p\rvert/R_e]$ S-4
−5	−10.79	−9.52	−13.67	−12.21	21.1	22.0
−4	−8.62	−7.62	−10.39	−9.28	17.0	17.9
−3	−6.47	−5.73	−7.43	−6.63	12.9	13.6
−2	−4.30	−3.81	−4.72	−4.21	8.9	9.5
−1	−2.15	−1.89	−2.25	−2.03	4.4	6.9
−0.5	−1.09	−0.97	−1.09	−0.99	—	2.0
0	0	0	0	0	—	—
0.5	1.06	0.93	1.06	0.93	—	—
1	2.16	1.91	2.06	1.81	4.9	5.5
2	4.31	3.81	3.94	3.48	9.4	9.5
3	6.47	5.71	5.65	5.00	14.5	14.2
4	8.64	7.63	7.24	6.39	19.3	19.4
5	10.80	9.51	8.66	7.65	24.7	24.3
7	15.11	13.33	11.19	9.88	35.0	34.9

[a] $R = (\Lambda_0 - \Lambda)/\beta$, $\beta = 0.003\,94$, and Λ_0 = unperturbed eigenvalue.
[b] Diffusion theory prediction with Mark boundary conditions.

two methods. Its only conceptual disadvantage (over the other methods) is that in the course of the derivation it is required that the unperturbed and perturbed solutions be analytically continued into the perturbed and unperturbed domains, respectively. However, in the final formula such a continuation is not needed because the formula requires only the information about the unperturbed problem (of course, the boundary perturbation has to be known).

The accuracy of the formulas was assessed by considering several numerical examples both in transport theory and the diffusion approximation. The examples all indicated that good eigenvalue estimates (in particular, in transport theory) can be obtained by using the perturbation formalism developed in this paper.

There appear to be a number of useful and significant extensions of the work contained in this paper. Of perhaps most direct interest would be for one to consider a generalization of the perturbation theory developed in this paper to compute second- and higher-order eigenvalue changes due to boundary perturbations. These changes can be obtained by generalizing either Method 1 or 2. To determine, for example, the second-order change in the eigenvalue, one needs find the first-order change in the perturbed solution ϕ_1 in terms of the unperturbed solution ϕ_0. In this respect, Method 2 in its present form seems more attractive since the higher-order changes in the perturbed solution $(\phi_1,\phi_2,\ldots,\phi_n)$ can be explicitly obtained in terms of ϕ_0. On the other hand, Method 1 needs to be modified to yield these changes. Regardless of the method (existing ones) used, the problem of finding the eigenfunctions for each order will be reduced to solving inhomogeneous differential equations (integro-differential equations in the case of transport theory) subject to inhomogeneous boundary conditions. The complication of these methods is that the solutions to these equations can be, in all but simple cases, very difficult to obtain. Several attempts have already been made in solving these equations for the simple case when the system operator is of the Helmholtz type.[10,14,15]

It might also be of interest to consider alternate approaches to the problem of boundary perturbation. For example, one can modify Brillouin's or Wassermann's method (Sec. II) so as to be applicable to the transport or diffusion equation. Or as suggested by Larsen and Pomraning, one may use the well-known fixed domain perturbation formula (i.e., Rayleigh quotient) by completely enclosing the perturbed domain within the smallest smooth domain possible. The unperturbed problem would then correspond to adding material to this additional region surrounding the perturbed domain. The perturbation, now a volume (or operator) perturbation, would consist of replacing the added material by void (vacuum). It must be emphasized that this treatment of the boundary change via a volume perturbation cannot work in the context of diffusion theory. Replacing material by vacuum would imply an infinite perturbation in the diffusion coefficient, making use of the Rayleigh quotient impossible. This approach has been discussed, in detail, by Larsen and Pomraning.[3]

Another interesting subject to pursue is the computation of other functionals (than the eigenvalue) in an eigenvalue problem. For example, one may be interested in estimating the change in the breeding ratio in a criticality problem when the boundary of the reactor is perturbed. In this case, Method 2 seems to be the ideal tool with which one may approach this problem, since in this method boundary changes are converted into fictitious perturbations in the interior operator of the system. Once this (operator) perturbation is known, one can then use the existing perturbation (or variational) methods to estimate the change in the desired functional.

Finally, in contrast to suggestions made above, it may be of interest to consider fixed source (noneigenvalue) problems in which one is to obtain an estimate of the change in an arbitrary functional of the solution when the boundary of the system is perturbed. Again, as suggested above, by using the idea of an inverse operator (Method 2), one can reduce the problem to one in which ordinary perturbation (or variational) methods may be applied to obtain the change in the functional (e.g., breeding ratios, reaction rates, etc.).

ACKNOWLEDGMENT

This work was supported partially by the National Science Foundation and partially by the General Atomic Co.

[1] J. J. Duderstadt and W. R. Martin, *Transport Theory* (Wiley, New York, 1979), Chap. 7.
[2] G. I. Bell and S. Glasstone, *Nuclear Reactor Theory* (Van Nostrand Reinhold, New York, 1970), Chap. 6.
[3] E. W. Larsen and G. C. Pomraning, Nucl. Sci. Eng. **77**, 415 (1981).
[4] L. Rayleigh, *The Theory of Sound* (McMillan, New York, 1877; Dover, New York, 1945), Vol. 1, Sec. 209.
[5] G. Poyla and S. Szego, *Isotropic Inequalities in Mathematical Physics* (Princeton U. P., Princeton, New Jersey, 1951).
[6] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[7] M. L. Brillouin, C. R. Acad. Sci., Paris **204**, 1863 (1937).
[8] M. N. Cabrera, C. R. Acad. Sci., Paris **207**, 1175 (1938).
[9] H. Froehlich, Phys. Rev. **54**, 945 (1938).
[10] H. Feshbach and A. M. Clogston, Phys. Rev. **59**, 189 (1940).
[11] R. Fuchs, Proc. R. Soc. (London) **176**, 214 (1941).
[12] H. Feshbach, Phys. Rev. **65**, 307 (1940).
[13] R. H. Bolt, H. Feshbach, and A. M. Clogston, J. Acoust. Soc. Am. **13**, 65 (1942).
[14] G. D. Wassermann, Phil. Mag. **37**, 536 (1946).
[15] G. D. Wassermann, Phil. Mag. **59**, 251 (1948).
[16] H. Feshbach and C. M. Harris, J. Acoust. Soc. Am. **18**, 472 (1946).
[17] H. B. Keller and J. B. Keller, Quart. Appl. Math. **12**, 141 (1954).
[18] G. Strang and A. E. Berger, *AMS Summer Proceedings on PDE's* (Univ. of California Press, Berkeley, 1971).
[19] J. J. Blair, J. Appl. Math. **24**, 277 (1973).
[20] J. W. Helton and J. V. Ralston, J. Diff. Eqs. **21**, 378 (1976).
[21] J. J. Duderstadt and L. J. Hamilton, *Nuclear Reactor Analysis* (Wiley, New York, 1976), p. 216.
[22] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. 1, p. 395.
[23] J. J. Duderstadt and L. J. Hamilton, in Ref. 21, p. 138.
[24] J. J. Duderstadt and L. J. Hamilton, in Ref. 21, p. 288.
[25] F. Rahnema, Ph.D. thesis (University of California, Los Angeles, 1981).
[26] Argonne National Laboratory, *Reactor Physics Constants*, ANL-5800, 1963.
[27] W. W. Engle, Jr., CCC-82/ANISN Radiation Shielding Information Center, ORNL, Oak Ridge, Tenn., 1967.
[28] G. I. Bell and S. Glasstone, in Ref. 2, p. 99.
[29] J. J. Duderstadt and W. R. Martin, in Ref. 1, p. 20.
[30] J. J. Duderstadt and W. R. Martin, in Ref. 1, p. 67.
[31] G. I. Bell and S. Glasstone, in Ref. 21, Chap. 5.
[32] W. A. Rhodes, *et al.*, ORNL/TM-6529, Oak Ridge National Laboratory, 1978.

On massive electromagnetism in the null-plane: their limit and "dual" invariance

S. Salamó
Universidad Simón Bolivar, Departamento de Física, Apartado Postal 80659, Caracas 1080, Venezuela

(Received 18 August 1982; accepted for publication 24 September 1982)

The limiting process is developed, stressing the simplicity of the null-plane formulation. A SU(2) "dual" invariance for the dynamical variables is made explicit. The symmetry is not satisfied by the field strengths.

PACS numbers: 41.10. − j, 03.50.De

1. INTRODUCTION

The study of the $m \to 0$ limiting process for massive electromagnetism has been carried out by several authors.[1] In these studies the 3 + 1 decomposition of the theory has been exploited in order to isolate the dynamical variables. To our knowledge there is no treatment of this problem in the framework of a 2 + 2 decomposition,[2,3] that is within the null-plane formalism. In the case of electromagnetism, the null-plane formalism has several advantages over the 3 + 1 decomposition. Let us mention a few of them: the constraints are easily solved. Dual invariance is made explicit at the level of the Lagrangian[4] and finally it led to a covariant theory of monopoles with two degrees of freedom.[5,6]

This paper is organized as follows. We first discuss the massive electromagnetism and the limit of the theory when the mass goes to zero. Second, we analyze how dual invariance is implemented in the theory, at the level of field strengths or at the level of the dynamical variables of the theory.

2. MASSIVE ELECTROMAGNETISM IN THE NULL-PLANE

Let us denote by $(\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3)$ the usual space-time coordinates. The infinite momentum frame is defined by a new set of variables $(x^{(+)}, x^1, x^2, x^{(-)})$, related to the usual ones by

$$x^{(+)} = \frac{\hat{x}^0 - \hat{x}^3}{\sqrt{2}}, \quad x^{(-)} = \frac{\hat{x}^0 + \hat{x}^3}{\sqrt{2}}, \quad x^1 = \hat{x}^1, \quad x^2 = \hat{x}^2. \tag{1}$$

The matrix that relates both set of coordinates is then given by

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}, \tag{2}$$

so that the null components of any given tensor in the 3 + 1 representation can be easily obtained by means of the Q matrix given above.

In order to have an unconstrained action for massive electromagnetism, we start with the following first-order action, depending on ten independent variables A_μ and $F_{\mu\nu}$:

$$dI = \{\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + A_{\mu,\nu}F^{\mu\nu} - (m^2/2)A_\mu A^\mu + A_\mu J^\mu\} d^4x, \tag{3}$$

where $F_{\mu\nu}$ is an arbitrary antisymmetric tensor and J_μ the external current. Variation on the ten independent variables provides us with the following equations of motion:

$$\begin{aligned} &\text{(a)}\ F_{\mu\nu} = A_{\mu,\nu} - A_{\nu,\mu}, \\ &\text{(b)}\ \partial_\nu F^{\mu\nu} = J^\mu - m^2 A^\mu. \end{aligned} \tag{4}$$

We look for the evolution of the system along the $x^{(+)}$ variable, which is a null-plane. Therefore in the null-plane decomposition of the equations of motion, the constraints are those where the derivatives with respect to the $x^{(+)}$ variable do not appear. In our case these are

$$\begin{aligned} &\text{(a)}\ F_{ab} = A_{b,a} - A_{a,b}, \\ &\text{(b)}\ F_{a-} = A_{-,a} - A_{a,-}, \quad a, b = 1, 2. \\ &\text{(c)}\ F_{+-,-} - F_{a-,a} = J_- - m^2 A_-, \end{aligned} \tag{5}$$

as one can easily verify from the equations of motion (4).

In order to solve the constraints, it is convenient to decompose the two vectors A_a and F_{a-} into their longitudinal and transverse parts, i.e.,: $A_a = A^{\mathrm{L}}_{,a} + \epsilon_{ab}A^{\mathrm{T}}_{,b}$ and a similar expression for F_{a-}. Using the above decomposition in Eq. (5), we obtain for F_{ab}, A_-, and F_{a-} :

$$\begin{aligned} &\text{(a)}\ F_{ab} = A^{\mathrm{T}}_{b,a} - A^{\mathrm{T}}_{a,b}, \\ &\text{(b)}\ A_- = D^{-1}[F_{+-,-} + \nabla A^{\mathrm{L}}_{,-} - J_-], \\ &\text{(c)}\ F_{a-} = \epsilon_{ab}\nabla^{-1}F_{12,b-} + D^{-1}[F_{+-,-a} \\ &\qquad + \nabla A^{\mathrm{L}}_{,-a} - J_{-,a}] - A^{\mathrm{L}}_{,-a}, \end{aligned} \tag{6}$$

where the operators ∇ and D are defined as follows: $\nabla \equiv \partial_1^2 + \partial_2^2$ and $D \equiv \nabla - m^2$.[7] It is worth pointing out that the four constraints solved [Eqs. (6)] are expressed in terms of the variables A^{T}_a, F_{+-}, A^{L} and F_{ab} and also that we are left with six independent variables. After substituting Eqs. (6) into the action (3), a straightforward calculation gives us

$$\begin{aligned} dI = \{&\tfrac{1}{2}F_{+-}F_{-+} - F_{+-,+}D^{-1}F_{+-,-} \\ &+ F_{+-}D^{-1}(J_{+,-} - J_{-,+}) + J_+D^{-1}J_- \\ &+ \tfrac{1}{2}A_{a,b}F_{ab} - A^{\mathrm{T}}F_{12,+-} + A^{\mathrm{T}}_{,i}J^{\mathrm{T}}_{,i} \\ &- (m^2/2)A^{\mathrm{T}}_{,i}A^{\mathrm{T}}_{,i} + A^{\mathrm{L}}\nabla A^{\mathrm{L}}_{,-+} - A^{\mathrm{L}}D^{-1}\nabla^2 A^{\mathrm{L}}_{,-+} \\ &+ A^{\mathrm{L}}D^{-1}\nabla[J_{-,+} + J_{+,-}] + A^{\mathrm{L}}_{,i}J^{\mathrm{L}}_{,i} \\ &- (m^2/2)A^{\mathrm{L}}_{,i}A^{\mathrm{L}}_{,i}\} dx\, dx^{(+)}\, dx^{(-)}. \end{aligned} \tag{7}$$

Notice that the variables A_+ and F_{a+} do not occur in the action, thus the number of degrees of freedom has been reduced to three, a well-known result for the Proca fields.

Due to the structure of the reduced action (7), and also from our experience with the null-plane treatment of electromagnetism,[5] it is convenient at this point to introduce a new

 0022-2488/83/030702-03$02.50

set of variables, defined as follows:

$$A^{(1)} \equiv (-\nabla)^{1/2} A^{T}, \quad A^{(2)} \equiv (-D)^{-1/2} F_{+-},$$
$$A^{(3)} \equiv (-D)^{-1/2} (-\nabla)^{1/2} A^{L}. \tag{8}$$

Then, the action (7) in terms of the new variables reads as

$$\begin{aligned} dI = \{ & A^{(1)}_{,+} A^{(1)}_{,-} - \tfrac{1}{2}[(-D)^{1/2} A^{(1)}]^2 \\ & + A^{(1)} (-\nabla)^{1/2} J^{T} + A^{(2)}_{,+} A^{(2)}_{,-} - \tfrac{1}{2}[(-D)^{1/2} A^{(2)}]^2 \\ & + A^{(2)} (-D)^{-1/2} [J_{-,+} - J_{+,-}] + m^2 (A^{(3)}_{,+} A^{(3)}_{,-} \\ & - \tfrac{1}{2}[(-D)^{1/2} A^{(3)}]^2) + J_+ D^{-1} J_- \\ & + A^{(3)} ((-\nabla)^{1/2} (-D)^{-1/2} [J_{-,+} + J_{+,-}] \\ & + (-\nabla)^{1/2} (-D)^{1/2} J^{L}) \} dx\, dx^{(+)}\, dx^{(-)}. \end{aligned} \tag{9}$$

If we consider the limit of the action when $m \to 0$, we easily see that the term $A^{(3)}[(-\nabla)^{1/2}(-D)^{-1/2}(J_{-,+} + J_{+,-}) + (-\nabla)^{1/2}(-D)^{1/2} J^{L}] \to A^{(3)} \partial_\mu J^\mu$. So if we further assume that J_μ is a conserved current, the $A^{(3)}$ field becomes a Lagrange multiplier of the theory when $m \to 0$. In this case we are left with the usual unconstrained Lagrangian[4,5] containing two degrees of freedom, namely:

$$\begin{aligned} dI_0 = \{ & A^{(1)}_{,+} A^{(1)}_{,-} - \tfrac{1}{2}[(-\nabla)^{1/2} A^{(1)}]^2 + A^{(1)} (-\nabla)^{1/2} J^{T} \\ & + A^{(2)}_{,+} A^{(2)}_{,-} - \tfrac{1}{2}[(-\nabla)^{1/2} A^{(2)}]^2 \\ & + A^{(2)} (-\nabla)^{-1/2} (J_{-,+} - J_{+,-}) \\ & + J_+ \nabla^{-1} J_- \} dx\, dx^{(+)}\, dx^{(-)}. \end{aligned} \tag{10}$$

The equations of motion for massive electromagnetism are obtained from the action (9) after variation of the action with respect to the independent fields $A^{(i)}$ $(i = 1,2,3)$, to obtain

$$\text{(a)} \quad [2\partial_+ \partial_- - \nabla + m^2] A^{(1)} = (-\nabla)^{1/2} J^{T},$$

$$\begin{aligned} \text{(b)} \quad & [2\partial_+ \partial_- - \nabla + m^2] A^{(2)} \\ & = -(-D)^{-1/2} [J_{+,-} - J_{-,+}], \end{aligned} \tag{11}$$

$$\begin{aligned} \text{(c)} \quad & [2\partial_+ \partial_- - \nabla + m^2] A^{(3)} \\ & = (1/m^2)(-\nabla)^{1/2} [(-D)^{-1/2} (J_{-,+} \\ & \quad + J_{+,-}) + (-D)^{1/2} J^{L}]. \end{aligned}$$

Thus, the three degrees of freedom satisfy inhomogeneous massive Klein–Gordon equations, whereas for massless electromagnetism, the two degrees of freedom satisfy inhomogeneous massless Klein–Gordon equations.[4]

From the above equations we see that the propagator of the degrees of freedom for massive electromagnetism have the desired limit if $(1/m^2)(-\nabla)^{1/2}[(-D)^{-1/2}(J_{-,+} + J_{+,-}) + (-D)^{1/2} J^{L}] \to 0$, or in other words, when $(1/m^2) P_\mu J^\mu \to 0$, a well-known result.[8]

3. DUAL INVARIANCE

In order to discuss the invariance of the theory under dual transformations, it is convenient to write the fields strengths **E** and **B** in terms of the dynamical variables. First we need the null-plane components of the four-potential A_μ in terms of the above variables. These are derived from Eqs. (5) and from the equations of motion for the $A^{(i)}$ fields [Eqs. (11)]; we obtain

$$\begin{aligned} A_a &= \epsilon_{ab} (-\nabla)^{-1/2} A^{(1)}_{,b} + (-D)^{1/2} (-\nabla)^{-1/2} A^{(3)}_{,a}, \\ A_- &= -(-D)^{-1/2} A^{(2)}_{,-} + (-D)^{-1/2} (-\nabla)^{1/2} A^{(3)}_{,-} \\ &\quad + (-D)^{-1} J_-, \\ A_+ &= (-D)^{-1/2} A^{(2)}_{,+} + (-D)^{-1/2} (-\nabla)^{1/2} A^{(3)}_{,+} \\ &\quad + (-D)^{-1} J_+. \end{aligned} \tag{12}$$

Since the field strengths **E** and **B** are defined in terms of the null components of $F_{\mu\nu}$ as

$$\begin{aligned} E_i &= F_{0i} = (1/\sqrt{2})(F_{+i} - F_{-i}), \\ E_3 &= F_{03} = F_{+-}, \\ B_i &= \epsilon_{ij} F_{3j} - (\epsilon_{ij}/\sqrt{2})(F_{-j} - F_{+j}), \\ B_3 &= -F_{12}, \end{aligned} \tag{13}$$

we obtain for **E** and **B**,

$$\begin{aligned} E_i &= \epsilon_{ij} (-\nabla)^{-1/2} A^{(1)}_{,0j} + (-D)^{-1/2} A^{(2)}_{,3i} \\ &\quad + [(-D)^{1/2} (-\nabla)^{-1/2} - (-D)^{-1/2} (-\nabla)^{1/2}] A^{(3)}_{,0i} \\ &\quad - (-D)^{-1} J_{0,i}, \\ E_3 &= (-D)^{1/2} A^{(2)}, \\ B_i &= -(-\nabla)^{-1/2} A^{(1)}_{,3i} + \epsilon_{ij} (-D)^{-1/2} A^{(2)}_{,0j} \\ &\quad + [(-D)^{1/2} (-\nabla)^{-1/2} - (-D)^{-1/2} (-\nabla)^{1/2}] \epsilon_{ij} \\ &\quad \times A^{(3)}_{,3j} - (-D)^{-1} \epsilon_{ij} J_{3,j} \\ B_3 &= -(-\nabla)^{1/2} A^{(1)}. \end{aligned} \tag{14}$$

With these results, we are now in a position to discuss the zero current case and furthermore, try to generalize the dual invariance occurring in free electromagnetism. First, we would like to point out three important facts that appear in free electromagnetism. (a) When the action is expressed in terms of the field strengths **E** and **B**, the action is not dual invariant, only the equations of motion share the above symmetry.[9] (b) If one writes the unconstrained action, the action is indeed invariant under U(1), and finally (c) the U(1) symmetry for the dynamical variables implies the same symmetry for the field strengths **E** and **B**, and vice versa.[5] The second point can be easily seen from Eq. (10) (after setting the external current equal to zero). The resulting action can be written as

$$dI' = [\partial_+ \mathbf{A} \cdot \partial_- \mathbf{A} - \tfrac{1}{2} \partial_i \mathbf{A} \cdot \partial_i \mathbf{A}]\, dx\, dx^{(+)}\, dx^{(-)}, \tag{15}$$

where **A** is defined as a two component vector: $\mathbf{A} = \begin{pmatrix} A^{(1)} \\ A^{(2)} \end{pmatrix}$, thus the U(1) invariance is exhibit in an explicit way.[10]

The reduced action for free massive electromagnetism can be obtained from Eq. (9); after setting the external current to zero, we have

$$\begin{aligned} dI_{J=0} = \{ & \partial_+ \mathbf{C} \cdot \partial_- \mathbf{C} - \tfrac{1}{2} \partial_i \mathbf{C} \cdot \partial_i \mathbf{C} \\ & - \tfrac{1}{2} m^2 \mathbf{C} \cdot \mathbf{C} \} dx\, dx^{(+)}\, dx^{(-)}, \end{aligned} \tag{16}$$

where **C** is defined as

$$\mathbf{C} = \begin{pmatrix} A^{(1)} \\ A^{(2)} \\ A^{(3)\prime} \end{pmatrix}$$

and

$$A^{(3)\prime} = m A^{(3)}.$$

Therefore, the symmetry group of the theory is SU(2). We

can say that this SU(2) "dual" symmetry generalizes the U(1) invariance occurring in free massless electromagnetism. It is convenient to stress that the SU(2) symmetry is not shared by the field **E** and **B**, as one can see from the equations for the fields in terms of the dynamical variables $A^{(i)}$ [Eq. (14)]. This situation does not occur in free massless electromagnetism, as we mentioned before.

The corresponding conserved current associated to this "dual" invariance is obviously

$$\mathbf{J}^{\mu} = \mathbf{C} \times \partial^{\mu} \mathbf{C}. \tag{17}$$

The expression of this current in terms of the **E** and **B** fields is complicated and not very illuminating. Only J_3^{μ} has a simple expression,

$$J_3^{\mu} = -[(-D)^{-1/2} E_3 (-\nabla)^{-1/2} B_3^{\mu} - (-\nabla)^{-1/2} B_3 (-D)^{-1/2} E_3^{\mu}], \tag{18}$$

so that the conserved Q_3 charge is

$$Q_3 = -\int d^3x [(-D)^{-1/2} E_3 (-\nabla)^{-1/2} B_{3,0} - (-\nabla)^{-1/2} B_3 (-D)^{-1/2} E_{3,0}]. \tag{19}$$

As one expects, this charge reduces to the known dual charge[4,10] when $m \to 0$, that is to say

$$\begin{aligned} Q_3 \underset{m=0}{\to} Q &= -\int d^3x [(-\nabla)^{-1/2} E_3 (-\nabla)^{-1/2} B_{3,0} - (-\nabla)^{-1/2} B_3 (-\nabla)^{-1/2} E_{3,0}] \\ &= \frac{1}{2} \int d^3x [B_i (\nabla + \partial_3^2)^{-1} E_{i,0} - E_i (\nabla + \partial_3^2)^{-1} B_{i,0}]. \end{aligned} \tag{20}$$

To conclude, a few remarks seem pertinent. First, the simplicity of the null-plane treatment provides us with explicit dual invariance for the unconstrained Lagrangian for free massless and massive electromagnetism. Also the limit when $m \to 0$ for massive electromagnetism is very transparent in this formulation.

Finally, one could try to apply the results given here to a massive Cabibbo–Ferrari theory for monopoles.[11] One would expect that a Lagrangian formulation of this theory would have similar features to those of the massless case[5]: (a) too many degrees of freedom (likely 6) and (b)physical observables which appear to depend on three degrees of freedom. Work in this direction is now in progress.

ACKNOWLEDGMENT

I would like to thank Professor R. Gambini for several stimulating discussions.

[1] S. Deser, Ann. Inst. Henri Poincaré **16**, 79 (1972) and references therein.
[2] A complete treatment for massive QED in the null-plane is given by D. S. Soper, Phys. Rev. D **4**, 1620 (1971).
[3] C. Aragone and A. Restuccia, Phys. Rev. D **13**, 207 (1976). The authors have worked out the gravitating massive vector field in the light cone. The flat case also is studied but the interaction with external currents is not considered.
[4] R. Gambini and S. Hojman, Ann. Phys. **105**, 407 (1977).
[5] R. Gambini and S. Salamó, Nuovo Cimento A **49**, 276 (1979).
[6] R. Gambini, S. Salamó, and A. Trías, Lett. Nuovo Cimento **27**, 385 (1980).
[7] Properties of Rietz and Bessel potentials are found in: E. M. Stein, *Singular and Differentiability Properties of Functions* (Princeton U.P., Princeton, NJ, 1970).
[8] C. Fronsdal, Nucl. Phys. B **167**, 237 (1980).
[9] Recently a general treatment of dual symmetry for the field strengths has been given by M. K. Gaillard and B. Zumino, Nucl. Phys. B **193**, 221 (1981). See also: V. I. Fushchich and A. G. Nikitin, J. Phys. A: Math. Gen. **12**, 747 (1979) and S. Hojman, in "Relativity and Gravitation," *Proceedings of the Third Latin-American Symposium*, edited by S. Hojman, M. Rosenbaum, and M. P. Ryan (Univ. Nac. Aut. de México, México, 1982).
[10] S. Deser and C. Teitelboim, Phys. Rev. D **13**, 1592 (1976), S. Deser, J. Phys. A: Math. Gen. **15**, 1053 (1982). Their treatment in 3 + 1 does not exhibit explicitly the dual symmetry for the unconstrained Lagrangian, in contrast with our formulation, Eq. (15).
[11] E. Kyriakopoulos, Phys. Rev. D **16**, 517 (1977).

Charged particles in an intense, plane electromagnetic wave: Limitations of the nonrelativistic theory

Kai Drühl and J. K. McIver
Institute for Modern Optics, Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131

(Received 20 July 1982; accepted for publication 30 September 1982)

We discuss the classical and quantum theory of a charged particle moving in a plane wave electromagnetic field. In the nonrelativistic theory forbidden regions in the momentum spectrum of the particle occur, while no such regions exist in the relativistic theory. The gaps in the momentum spectrum are shown to be an artifact of the nonrelativistic approximation and are caused by the possibility of transforming to a Galilean frame of reference moving at the speed of light in which the particle moves in a time independent periodic potential. A quantitative estimate of the momentum gaps for optical frequencies shows that they are proportional to a very large power $n \simeq 10^5$ of the ratio of particle velocity to the speed of light, and hence, beyond the limits of accuracy of the nonrelativistc theory. For this result to hold in the low frequency limit we have to employ a conjecture about the radius of convergence of the perturbation series expansion for eigenvalues of Mathieu's equation. While no rigorous proof seems to exist at present, this conjecture is well supported both by numerical studies and by our WKB calculations.

PACS numbers: 41.70. + t, 03.65.Sq, 11.10.Qr

1. INTRODUCTION

The interaction of intense classical radiation fields with matter is of considerable practical and theoretical interest. In fact applications in the field of laser damage to optical components are so numerous that a separate review paper would be necessary to do justice to the amount of effort involved here. The following references are therefore meant to mark two typical approaches, and do by no means exhaust the set of representative papers.[1,2] From a theoretical point of view the solution of the external field problem has led to advances in the fully quantum mechanical treatment, and served to clarify important issues in quantum electrodynamics.[3,4]

For the case of a spinless particle interacting with a classical plane wave an exact solution of the relativistic quantum mechanical problem was given by Gordon[5] and later for the spin 1/2-particle by Volkov.[6] Furthermore, both the relativistic and nonrelativistic classical problem can be reduced to a one dimensional problem, and hence all solutions can be obtained in terms of a simple quadrature.[4,5]

The solution of the relativistic problem applies to all situations where the radiation field can be treated as an external classical quantity, and it is valid in particular in the nonrelativistic regime, where the velocity v of the electron is much smaller than the velocity c of light. Nevertheless, it is important to have an independent formulation in terms of the nonrelativistic Schrödinger equation. From a theoretical point of view nonrelativistic quantum mechanics should be considered a theory in its own right, and hence an independent solution of the problem be sought within its framework (we shall discuss the necessary precautions for doing so in this paper). From a practical point of view one would like to apply the theory to the effects of intense radiation fields on electrons in a solid, where the influence of the electronic band structure has to be taken into account.[7,8] Since band structure is calculated in the framework of nonrelativistic quantum mechanics it is important to maintain a clear understanding of how intense classical radiation fields are to be incorporated into this framework. Recently an exact solution for the nonrelativistic Schrödinger equation in the presence of an external plane wave field has been obtained.[9] On the basis of this solution the authors predict an optically induced band structure for free electrons, with relative energy separation in the microwave region for the highest laser intensities presently available. No such band structure is found in the known relativistic solutions.

In this paper we analyze the apparent discrepancies between the relativistic and the nonrelativistic theories. Since the WKB solution is exact for the relativistic quantum problem, the analogy between classical and quantum theory is a very close one.

In Sec. 2 we discuss both the relativistic and the nonrelativistic classical theory. The classical relativistic problem has been treated before[4,5] and we present it here for the sake of completeness and in order to establish contact to the nonrelativistic treatment. In the nonrelativistic theory it is possible to transform to a Galilean frame of reference in which the phase of the external wave is constant. In this frame the particle moves in a time independent, periodic potential. It is precisely this aspect of the classical motion which leads to forbidden energy bands in the quantum theory. No such transformation is possible in the relativistic formulation.

In Sec. 3 we interpret the parameters of the exact relativistic wavefunction for a spinless particle in terms of the classical constants of motion and discuss their classical limit. Only the wavefunction aspects of the nonrelativistic limit are discussed, leaving aside the well known procedure of obtaining the appropriate limiting form of the scalar product.[10] The nonrelativistic quantum theory is then developed in close analogy to the discussion of the classical theory. The bound state problem is discussed, and the WKB solution is

 0022-2488/83/030705-07$02.50

shown to approximate the nonrelativistic limit of the exact solution.

In Sec. 4 we obtain more quantitative results for a circularly polarized plane wave. The resulting Schrödinger equation is equivalent to Mathieu's equation, and we determine the appropriate region in the stability chart, in which solutions must be sought. The WKB approximation is employed to show that:

(a) No forbidden regions exist to leading order in the quasiclassical approximation in the valid asymptotic region.

(b) Asymptotic formulas employed before[9] belong to an inappropriate region of the stability chart, which corresponds to the unphysical bound state situation of the nonrelativistic theory.

The exact solutions of Mathieu's equation do show forbidden regions for the momentum spectrum of the particle. An estimate of their size is obtained from recent results of Harrell[11] and Avron and Simon.[12] These results show that the spectrum of parallel (to the direction of propagation of the wave) electron momentum has forbidden regions of size

$$\Delta r^3 \simeq \hbar k \theta^n,$$

where $\hbar k$ is the photon momentum, $\theta = v/c$ and n is an integer labeling the region

$$n = 2mc/\hbar k.$$

For optical photons $n \simeq 10^5$, showing that the width of the forbidden region is much smaller than the lowest order relativistic corrections.

We note that for this result to apply in the low frequency limit we have to conjecture that the perturbation series for the energy eigenvalues converges in a region where the coupling is proportional to the eigenvalue considered.[13] No mathematical proof for this conjecture seems to exist at the present, and the available proofs are restricted to regions where the coupling is at most proportional to the square root of the eigenvalue. However our conjecture is well supported by numerical results for the low lying eigenvalues,[13] and by WKB results in the region of large eigenvalues.

2. CLASSICAL MOTION

The motion of a relativistic classical particle in a plane wave field is obtained in the fully covariant formulation from a Hamiltonian

$$\mathscr{H}(x,p) = \frac{1}{2} g^{\mu\nu}\left(p_\mu - \frac{e}{c}A_\mu\right)\left(p_\nu - \frac{e}{c}A_\nu\right),$$

where

$$A_\mu = A_\mu(\phi), \quad \phi = (k\cdot x), \quad (k\cdot k) = 0; \quad x^0 = ct$$

is the vector potential of the field with wave vector k and phase ϕ. For two four vectors a and b their scalar product will be denoted by

$$(a\cdot b) = g^{\mu\nu}a_\mu b_\nu = a^\nu b_\nu = a_0 b_0 - \sum_{\alpha=1}^{3} a_\alpha b_\alpha.$$

The equations of motion for trajectories $x^\nu = x^\nu(\tau)$, $p_\mu = p_\mu(\tau)$ are

$$\frac{d}{d\tau}x^\nu = \frac{\partial \mathscr{H}}{\partial p_\nu} = p^\nu - \frac{e}{c}A^\nu, \tag{1a}$$

$$\frac{d}{d\tau}p_\mu = -\frac{\partial \mathscr{H}}{\partial x^\mu} = k_\mu \frac{e}{c}\left(p^\nu - \frac{e}{c}A^\nu\right)\frac{d}{d\phi}A_\nu. \tag{1b}$$

For a particle of mass m the Hamiltonian $\mathscr{H}$ is subject to the constraint

$$\mathscr{H} = \tfrac{1}{2}m^2c^2, \tag{2}$$

showing that

$$w^\mu = \frac{d}{d\tau}x^\mu/mc, \quad (w\cdot w) = 1$$

is the relativistic four velocity and the integration parameter τ is proportional to proper time. Since $A = A(\phi)$ we have

$$(k\cdot A) = 0 \tag{3}$$

and the phase changes at constant rate along any trajectory

$$\frac{d}{d\tau}\phi = \left(k\cdot\frac{d}{d\tau}x\right) = (k\cdot p) = \text{const},$$

$$\phi(\tau) = \phi_0 + (k\cdot p)\cdot\tau. \tag{4}$$

It is a then straightforward consequence of Eqs. (1), (3), and (4) that the functions $\rho_\nu(x,p)$ below are constants of motion (i.e., constant along any trajectory):

$$\rho_\nu(x,p) = p_\nu - \frac{k_\nu}{(k\cdot p)}\left[\frac{e}{c}(p\cdot A) - \frac{1}{2}\frac{e^2}{c^2}(A\cdot A)\right], \tag{5a}$$

$$\rho_\nu(x(\tau), p(\tau)) = \rho_\nu = \text{const}. \tag{5b}$$

They form a four vector ρ which is constrained to the "mass shell"

$$(\rho\cdot\rho) = m^2c^2. \tag{6}$$

For any given constant four vector ρ on the mass shell Eq. (5b) can be solved to give p as a function of ρ:

$$p_\nu = \rho_\nu + \frac{k_\nu}{(k\cdot\rho)}\left[\frac{e}{c}(\rho\cdot A) - \frac{1}{2}\frac{e^2}{c^2}(A\cdot A)\right]$$

$$= \rho_\nu + \frac{k_\nu}{(k\cdot\rho)}B(\rho,\phi); \quad (k\cdot p) = (k\cdot\rho). \tag{7}$$

Together with Eq. (4) this gives the momentum p as an explict function of τ. The position vector x can then be obtained from Eq. (1a) by direct integration. As an alternative procedure we note that a solution of the Hamilton–Jacobi equation may be constructed from Eq. (7):

$$p_\nu = \frac{\partial}{\partial x_\nu}W(x,\rho); \quad W(x,\rho)$$

$$= (\rho\cdot x) + \frac{1}{(k\cdot\rho)}\int_{\phi_0}^{(kx)} d\phi\, B(\rho,\phi), \tag{8a}$$

$$g^{\mu\nu}\left(\frac{\partial W}{\partial x^\mu} - \frac{e}{c}A_\mu\right)\left(\frac{\partial W}{\partial x^\nu} - \frac{e}{c}A_\nu\right) = (\rho\cdot\rho). \tag{8b}$$

A canonical transformation from constant initial coordinates x' and $p' = \rho$ is generated by Hamilton's principal

function S (see Ref. 14, Chap. 9)

$$S(x,\rho,\tau) = W(x,\rho) - \tfrac{1}{2}(\rho\cdot\rho)\cdot\tau, \tag{9a}$$

$$x'^{\mu} = \frac{\partial S}{\partial \rho^{\mu}} = x^{\mu} - \frac{k^{\mu}}{(k\cdot\rho)^2}\int_{\phi_0}^{\phi(\tau)} d\phi\, B(\rho,\phi) + \frac{1}{(k\rho)}\int_{\phi_0}^{\phi(\tau)} d\phi\, \frac{e}{c}A^{\mu} - \rho^{\mu}\cdot\tau,$$

$$\phi(\tau) = (k\cdot x) = (k\cdot x') + (k\cdot\rho)\tau = \phi_0 + (k\cdot\rho)\tau. \tag{9b}$$

This shows that the parameters ρ_ν may assume all values compatible with Eq. (6), and that every trajectory is uniquely characterized by these values and the initial position vector x'. The physical interpretation of the constants of motion ρ_ν is most easily found for a situation in which $A_\nu = 0$ in some regions of space-time. In these regions $\rho_\nu = p_\nu$. For wavefields of infinite extension like a monochromatic plane wave ρ is equal to the average particle momentum (up to an additive constant).

Let us now turn to the nonrelativistic limit of our theory. It is convenient to start from the equivalent, six-dimensional canonical formualtion, in which the total energy $E = cp_0$ is used as a Hamiltonian. E is obtained by solving Eq. (2) for p_0 in terms of the remaining variables. In a gauge where $A_0 = 0$ we obtain

$$E = (m^2c^4 + 2mc^2H)^{1/2},$$

$$H = \frac{1}{2m}\sum_{\alpha=1}^{3}\left(p^{\alpha} - \frac{e}{c}A^{\alpha}\right)^2. \tag{10}$$

The integration variable associated with E is just the canonically conjugate variable $t = x^0/c$, and the resulting equations of motion are equivalent to Eq. (1).

Note that for the Hamiltonian equations of motion to have their usual form momenta $p^{\alpha} = -p_{\alpha}$; $\alpha = 1,2,3$; must be used in the six-dimensional formulation. This is consistent with Eqs. (1).

In the nonrelativistic limit H is much smaller than the rest energy mc^2 on the whole trajectory. By expanding E to lowest order, H becomes the nonrelativistic Hamiltonian and is equal to the kinetic energy

$$\sum_{\alpha=1}^{3}\left(p^{\alpha} - \frac{e}{c}A^{\alpha}\right)^2 = m^2c^2\sum_{\alpha=1}^{3}(w^{\alpha})^2 = m^2c^2, \tag{11a}$$

$$\theta = v/c \ll 1; \quad E \simeq mc^2 + H. \tag{11b}$$

Condition (11b) puts restrictions on the values of parameters $\rho_\alpha, \alpha = 1,2,3$ and on the magnitude of the vector potential A. Let us assume that A is periodic without constant component, and fix our frame of reference and gauge such that

$$A_0 = A_3 = 0; \quad \phi = k^0(ct - x^3).$$

In this case $p_1 = \rho_1$ and $p_2 = \rho_2$ are constants, and from Eqs. (5) and (11) we obtain

$$(\rho)^2 = (\rho^1)^2 + (\rho^2)^2 \leqslant \theta^2 m^2c^2, \tag{12a}$$

$$(e/c)^2A^2 = (e/c)^2[(A^1)^2 + (A^2)^2] \leqslant \theta^2 m^2c^2, \tag{12b}$$

$$|B(\rho,\phi)| \leqslant \theta^2 m^2c^2, \tag{12c}$$

$$(\rho^3)^2 \leqslant \theta^2 m^2c^2. \tag{12d}$$

Equation (12d) is valid up to higher orders of θ^2. If any of these conditions is violated the particle will assume relativistic velocities at some points on the corresponding trajectories.

As in the relativistic theory the momenta p^1 and p^2 are constants of the motion, and we shall assume then to be fixed throughout the following discussion

$$p^{\beta} = r^{\beta}, \quad \beta = 1,2. \tag{13}$$

In order to maintain a clear distinction we have denoted the nonrelativistic constants, although equal in magnitude to their relativistic counterparts by the new letter r^{β}. Under the above assumption the nonrelativistic Hamiltonian H describes the motion of a one-dimensional mass point in the time dependent potential $-B(r,\phi)/m$,

$$H = \frac{1}{2m}\sum_{\beta=1,2}(r^{\beta})^2 + \frac{1}{2m}(p^3)^2 - \frac{1}{m}B(r,\phi). \tag{14}$$

The rate of change of the phase ϕ along the trajectory is given by

$$\dot{\phi} = k^0(c - \dot{x}) = k^0(mc^2 - p^3c)/mc. \tag{15a}$$

Note that $\dot{\phi}$ is no longer a constant of the motion. A constant of the motion is

$$C_0 = H - p^3c, \tag{15b}$$

which may be obtained from the relativistic constant of motion $(k\cdot p)$ by expanding the latter to first order in θ^2. This situation occurs in all orders of approximation. If the relativistic Hamiltonian is expanded to order $2n$ in θ, the resulting equations of motion are correct to order $2n - 1$, while the constants of motion coincide with the relativistic constants of motion up to order $2n$. This shows that it is not possible to approximate the relativistic dynamics consistently in any finite order of expansion. In fact one can argue that the additional term H in Eq. (15b) represents a second order relativistic correction to the term pc, which should be dropped since the equations of motion are correct only to first order. While this is not a self-consistent procedure within the framework of the nonrelativistic theory, it is consistent from the point of view of the relativistic theory, and is justified as a valid approximation for $\theta \ll 1$, which is the range of validity of the nonrelativistic formulation. It is instructive at this point to push the nonrelativistic theory beyond its limits, and change to a comoving frame of reference, in which the phase of the radiation field is constant, corresponding to an observer moving at the speed of light. While this is not possible in the relativistic theory (unless we consider a dielectric medium of refractive index larger than unity, see Ref. 14), it is a perfectly consistent formal possibility in the nonrelativistic theory. The corresponding time dependent canonical transformation from variables x^3 and p^3 to new variables x' and p' is generated by the function F (Ref. 15, Chap. 8)

$$F(p',x^3) = p'(x^3 - ct) + mcx^3, \tag{16a}$$

$$x' = \frac{\partial F}{\partial p'} = x^3 - ct, \quad p^3 = \frac{\partial F}{\partial x^3} = p' + mc. \tag{16b}$$

The new time independent Hamiltonian H' is given by

$$H' = H + \frac{\partial F}{\partial t} = H - p'c = \frac{1}{2}mc^2 + \frac{1}{2m}\sum_{\beta=1,2}(r^{\beta})^2 + \frac{1}{2m}(p') - \frac{1}{m}B(r, -k_0x')$$

and is just equal (up to a constant) to the constant of motion C_0 above.

In the nonrelativistic regime proper we have

$$(p')^2/2m \simeq mc^2; \quad (1/m)B \sim \theta^2 mc^2; \quad \theta \ll 1.$$

Hence, the kinetic energy term will always dominate the potential energy, leading to unbounded motion of the variable x' (which is proportional to the phase ϕ). If we allow the kinetic energy in the comoving frame to become small however, bounded, oscillatory motion in the regions of negative potential energy will occur, corresponding to bound states in the quantum mechanical system. Of course this would correspond to relativistic velocities in the fixed frame of reference, and hence violate the nonrelativistic approximation. Let us finally introduce a momentum like constant of motion r^3 in analogy with the relativistic case by putting

$$C_0 = H - p^3 c = \frac{1}{2m} \sum_{\alpha=1}^{3} (r^\alpha)^2 - r^3 c. \tag{17}$$

The momentum p^3 can be expressed in terms of r^3

$$\begin{aligned} p^3 &= mc - ((mc - r^3)^2 + 2B(r,\phi))^{1/2} \\ &\simeq r^3 - (1/mc)B(r,p). \end{aligned} \tag{18}$$

The last term of Eq. (18) gives p^3 to lowest order in θ^2, and is identical with the corresponding expansion of the relativistic expression [Eq. (7)]. This shows that r^3 coincides with the relativistic constant p^3 to lowest order in θ^2.

Equations (15a) and (18) may be integrated to give the phase ϕ and hence x^3 and p^3 as an explicit function of t.

3. QUANTUM THEORY

The Klein–Gordon equation for a scalar particle in the presence of the plane wave field A_μ is

$$\hat{H}\psi = g^{\mu\nu}(\hat{p}_\mu - (e/c)A_\mu)(\hat{p}_\nu - (e/c)A_\nu)\psi = m^2c^2\psi,$$

$$\hat{p}_\mu \psi = +i\hbar \frac{\partial}{\partial x^\mu}\psi. \tag{19}$$

The solution W of the Hamilton–Jacobi equation [Eq. (8a)] also satisfies the wave equation

$$g^{\mu\nu}\frac{\partial}{\partial x^\mu}\frac{\partial}{\partial x^\nu}W(x,\rho) = 0. \tag{20}$$

In this case the WKB approximation gives an exact solution, and the vector parameter ρ is to be interpreted in terms of the asymptotic or average particle momentum as in the classical case:

$$\psi(x,\rho) = e^{-iW(x,\rho)/\hbar}. \tag{21}$$

Furthermore, if the wavelength of the external radiation field is large compared to the Compton wavelength of the particle,

$$k_0^{-1} \gg \hbar/mc,$$

wavepackets can be constructed from the wavefunctions [Eq. (21)] for which the expectation value of the position operator[16] moves along the classical trajectories defined by putting the classical constants of motion equal to their quantum mechanical expectation values.

In the lowest order of the nonrelativistic approximation the parameter ρ_0 is expanded

$$\rho_0 = (m^2c^2 + \rho^2)^{1/2} \simeq mc + \frac{1}{2mc}\rho^2; \quad \rho^2 = \sum_\alpha (\rho_\alpha)^2.$$

This gives for the phase $W(x,\rho)$ of the wavefunction [Eq. (22)] (in a gauge where $A_0 = 0$)

$$\begin{aligned} W(x,\rho) \simeq W_{(1)}(x,\rho) &= \left(mc + \frac{1}{2mc}\rho^2\right)x^0 \\ &+ \sum_\alpha \rho_\alpha x^\alpha + \frac{1}{mck_0}\int_{\phi_0}^{(kx)} d\phi\, B(\rho,\phi). \end{aligned} \tag{22}$$

After discarding the rest energy term mcx^0 we obtain the nonrelativistic wavefunction

$$\begin{aligned} \psi(x^\alpha,t,\rho) = \exp\Bigg[-i\hbar\Bigg(&\frac{1}{2m}\rho^2 t + \sum_\alpha \rho_\alpha x^\alpha \\ &+ \frac{1}{mck_0}\int_{\phi_0}^{(kv)} d\phi\, B(\rho,\phi)\Bigg)\Bigg]. \end{aligned} \tag{23}$$

We shall show below that Eq. (23) is a valid approximation to an exact solution[7] of the nonrelativistic Schrödinger equation under the conditions of Eqs. (12a)–(12d).

Let us now consider the nonrelativistic Schrödinger equation. Our construction of an exact solution will closely parallel the procedure adopted for the classical problem. The Schrödinger equation is

$$\begin{aligned} &i\hbar\frac{\partial}{\partial t}\psi = H\psi, \\ &H = \frac{1}{2m}(\hat{p}^3)^2 + \frac{1}{2m}\sum_{\alpha=1,2}\left(\hat{p}^\alpha - \frac{e}{c}A^\alpha(\phi)\right)^2, \\ &\phi = k_0(ct - x^3). \end{aligned}$$

Since the vector potential A^α does not depend on x^1 and x^2, we can obtain solutions of the form

$$\psi(x^\alpha,t) = \exp\left[i\hbar\left(\sum_{\beta=1,2} r^\beta x^\beta\right)\right]\chi(x^3,t), \tag{24}$$

where χ satisfies the Schrödinger equation with Hamiltonian H_0,

$$\begin{aligned} &i\hbar\frac{\partial}{\partial t}\chi = H_0\chi, \\ &H_0 = (1/2m)(\hat{p}^3)^2 + (1/2m)\sum_{\beta=1,2}(r^\beta)^2 - (1/m)B(r,\phi). \end{aligned} \tag{25}$$

Note that our momentum operators are defined in analogy with the classical momenta

$$\hat{p}^\alpha = -\hat{p}_\alpha = -i\hbar\frac{\partial}{\partial x^\alpha}, \quad \alpha = 1,2,3.$$

The potential $(1/m)B$ depends on t only via ϕ, hence a solution can be obtained by a time dependent, unitary transformation [compare Eq. (16)]

$$\begin{aligned} &\chi(x,t) = e^{i\hbar^{-1}mcx}\chi'(x - ct,t) = [U(t)\chi'](x,t), \\ &U(t) = e^{i\hbar^{-1}mc\hat{x}}e^{-i\hbar^{-1}ct\hat{p}}; \quad \hat{p} = \hat{p}^3, \quad \hat{x} = \hat{x}^3. \end{aligned} \tag{26}$$

$$\begin{aligned} i\hbar\frac{\partial}{\partial t}\chi &= i\hbar\frac{\partial}{\partial t}U(t)\chi' \\ &= U(t)\left[c\hat{p} + i\hbar\frac{\partial}{\partial t}\right]\chi' = HU\chi'. \end{aligned}$$

It follows that χ' obeys the time independent Schrödinger

equation

$$i\hbar\frac{\partial}{\partial t}\chi' = [U^{-1}HU - c\hat{p}]\chi' = H'\chi', \tag{27a}$$

$$H' = \frac{1}{2m}\hat{p}^2 - \frac{1}{m}B(r, -k^0x') + \frac{1}{2m}\sum_\beta (r^\beta)^2 + \frac{1}{2}mc^2. \tag{27b}$$

Solutions of Eq. (27) are of the form

$$\chi'(x',t) = e^{-i\hbar^{-1}E't}\chi''(x'),$$

$$H'\chi'' = E'\chi''. \tag{28}$$

The full wavefunction ψ is then given by

$$\psi(x,t) = \exp\left[i\hbar^{-1}\left(\sum_{\beta=1,2} r^\beta x^\beta + mcx^3 - E't\right)\right]\cdot\chi''(x^3 - ct). \tag{29}$$

For this wavefunction to describe a nonrelativistic particle, an appropriate range of energy eigenvalues E' for the Hamiltonian H' must be chosen. Let us remember that in order to cancel the term linear in $\hat{p}$ in Eq. (24a) we had to boost the momentum to relativistic vlaues by $\hat{p}\to\hat{p} + mc$ [Eq. (26)]. Hence the energy E' and the distribution of momentum associated with the wavefunction χ'' will be at relativistic values, also. It follows that the kinetic energy contribution to the Hamiltonian (27b) is much larger than the potential energy, resulting in unbounded states. Also in this situation the WKB approximation is valid. In this approximation the wave function ψ'' is given by

$$\chi''_{\mathrm{WKB}}(x') = (k(x'))^{-1/2}e^{-i\hbar^{-1}K(x')},$$

$$K(x') = \int_0^{x'} d\xi\, k(\xi) = \int_0^{x'} d\xi \left(2mE' - \sum_\beta (r^\beta)^2 - m^2c^2 - 2B(r, -k^0\xi)\right)^{1/2}. \tag{30a}$$

In view of transformation (26) we define a parameter r^3 by

$$2mE' - \sum_\beta (r^\beta)^2 - m^2c^2 = (mc - r^3)^2; \quad r^3 \ll mc, \tag{30b}$$

$$K(x') \simeq (mc - r^3)x' - \frac{1}{mc - r^3}\int_0^{x'} d\xi\, B(r, -k^0\xi) \simeq (mc - r^3)x' + (1/mck^0)\int_0^{(kx)} d\phi\, B(r,\phi). \tag{30c}$$

For the wavefunction ψ we obtain finally the approximation:

$$\psi_{\mathrm{WKB}}(x^\alpha,t) = \exp\left[i\hbar^{-1}\left(\sum_{\alpha=1}^{3} r^\alpha x^\alpha - Et - \frac{1}{mck^0}\int_0^{(kx)} d\phi\, B(r,\phi)\right)\right], \tag{30d}$$

where $E = (1/2m)\,\Sigma_{\alpha=1}^{3}(r^\alpha)^2$. This is identical to the nonrelativistic limit [Eq. (24)] of the exact relativistic wavefunction [Eq. (23)] if we put $r^\alpha = -p_\alpha$; $\alpha = 1,2,3$.

An analysis of Eq. (28) which is not based on the WKB approximation can be given along the following lines for a periodic external field. In this case solutions of Eq. (29) can be obtained in the form

$$\chi''(x') = e^{-i\hbar^{-1}p'x'}u(x'), \tag{31a}$$

where u is a periodic function. The wavefunction ψ is then given by

$$\psi(x^\alpha,t) = \exp\left[i^{-1}\left(\sum_{\beta=1,2} r^\beta x^\beta + (mc - p')x^3 - (E' - cp')t\right)\right]u(x^3 - ct). \tag{31b}$$

The parameters $(mc - p')$ and $(E' - cp')$ are the average three-component of momentum and energy of the particle in the state defined by wavefunction (31b). The conditions

$$(mc - p') \ll mc, \quad (E' - cp') \ll mc \tag{31c}$$

have then to be taken into account in order to select the appropriate values of parameters p' and E' for the solutions of Eq. (28). In the next section we discuss this procedure in detail for the case of a monochromatic plane wave of circular polarization, which leads to the well studied Mathieu equation.

4. CIRCULARLY POLARIZED RADIATION AND MATHIEU'S EQUATION

For circularly polarized radiation the vector potential in Eq. (22) is given by

$$A^1(\phi) = A\cos\phi, \quad A^2(\phi) = A\sin\phi, \quad A_3 = 0. \tag{32a}$$

With $r^1 = r\cos\alpha$ and $r^2 = r\sin\alpha$ we obtain for the potential $(1/m)B$ in Eq. (27b):

$$-(1/m)B(r,\phi) = (1/m)\left[-\frac{e}{c}rA\cos(\phi - \alpha) + (e^2/2c^2)A^2\right]. \tag{32b}$$

Finally introducing the new variable $z = (\phi - \alpha)/2$, we obtain Mathieu's equation from Eq. (28).

$$\frac{d^2}{dz^2}g(z) + (a - 2q\cos 2z)g(z) = 0, \tag{33a}$$

$$g(z) = g(-[k^0x' + \alpha]/2) = \chi''(x'), \tag{33b}$$

$$a = \frac{8m}{(\hbar k^0)^2}\left[E' - \frac{1}{2}mc^2 - \frac{1}{2m}r^2 - \frac{1}{2m}\frac{e^2}{c^2}A^2\right] = \frac{4}{(\hbar k^0)^2}\left[(mc - r^3)^2 - \frac{e^2}{c^2}A^2\right], \tag{33c}$$

$$q = -\frac{4}{(\hbar k^0)^2}\frac{e}{c}rA. \tag{33d}$$

Equations (30a)–(30d) agree with Eqs. (7a)–(7e) of Ref. 7 if identify their quantity $mv_\parallel$ with our parameter r^3. Note that r^3 has been defined by Eq. (30b).

As is well known, solutions of Mathieu's equation are either stable, (i.e., bounded) or unstable, (i.e., unbounded), depending on the values of a and q. Figure 1 shows the regions of stability and instability in the q-a plane, known as the stability chart. Since unstable solutions are not acceptable as wavefunctions, the corresponding values of a (for given q) and hence of r^3 [Eq. (33c)] are "forbidden" for a quantum particle. For radiation at optical wavelengths ($\lambda = 10^{-4}$

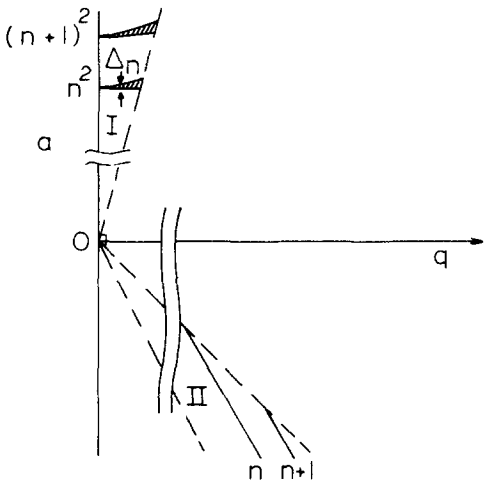


FIG. 1. Stability diagram for Mathieu's equation showing the two regions of interest: $q/a \gg 1$ (region II) and $q/a \ll 1$ (region I). Region I corresponds to the nonrelativistic case. A blowup of the region where $q, a \approx 1$ is shown in Fig. 2.

cm) and a particle of electronic mass we have

$$a \simeq 1.7 \times 10^{11} \gg 1, \tag{34a}$$

showing that we are in the asymptotic region $a \gg 1$. On the other hand we find for the ratio of q to a:

$$\epsilon = q/a \simeq -\frac{erA}{m^2c^3} \leqslant \theta^2 \ll 1, \tag{34b}$$

where we have used the condition $\theta^2 \ll 1$. This shows that even for nonrelativistic velocities both a and q can be much larger than unity, their ratio however being fixed by Eq. (34b). The region in the stability chart of interest to us as given by Eqs. (34a)–(34b) is denoted by I in Fig. 1.

As mentioned in Sec. 3 the WKB approximation is valid under these conditions. Since the WKB solutions are bounded for all values of the parameters a and q satisfying Eq. (34b) the forbidden regions have zero width in this approximation. On the other hand the expressions for the eigenvalues $a_n(q)$ for periodic and antiperiodic solutions derived from the WKB approximation do reproduce the asymptotic form (for $n \to \infty$) of the known power series expansions[13,17] of $a_n(q)$:

$$\int_0^\pi d_z(a_n(q) - 2q\cos 2z)^{1/2} = n\pi; \quad a_n(0) = n^2. \tag{35}$$

In view of the asymptotic conditions Eq. (34a) and (34b) it is convenient to express a_n as a function of the ratio ϵ:

$$\tilde{a}_n(\epsilon) = n^2\pi^2 / \left[\int_0^\pi dz(1 - 2\epsilon\cos 2z)^{1/2}\right]^2. \tag{36}$$

Note that $\tilde{a}_n(t)$ is an analytic function of ϵ for $|\epsilon| < 1/2$.

In order to obtain an estimate on the order of magnitude of the "forbidden" regions we have to go beyond the WKB approximation. In fact the problem of energy gaps for Mathieu's equation is an area of current research.[11,12]

In particular it has been shown[11,12] that for $n \to \infty$ the width Δ_n of the nth "forbidden" interval of the eigenvalue a leading to unstable solutions is given by

$$\Delta_n(q) = c_0 n\left[\frac{q}{4n^2}\right]^n, \tag{37}$$

where c_0 is a numerical constant. On the other hand $\Delta_n(q)$ is the difference between eigenvalues $a_n(q)$ and $b_n(q)$ belonging to even and odd eigenfunctions, and Eq. (37) also gives the correct first term in the power series expansion of $\Delta_n(q)$. For values of ϵ such that q is within the radius of convergence of the corresponding power series we obtain the leading term of a power series expansion $\tilde{\Delta}_n(\epsilon)$ with respect to ϵ:

$$\tilde{\Delta}_n(\epsilon) = c_0 \cdot n[\epsilon/4]^n + \cdots. \tag{38}$$

From Eqs. (38) and (33c) we obtain for the width of the corresponding forbidden region for r^3

$$\Delta r^3 = \tfrac{1}{4}\hbar k(\epsilon/4)^n, \quad n = \sqrt{a}.$$

Since $n \simeq 10^5$ for optical frequencies this width is proportional to a very large power of the small parameter θ^2, and hence beyond the limit of accuracy of the nonrelativistic approximation.

Let us mention that Eq. (37) was proven under the condition

$$|q| < \tfrac{1}{2}n.$$

Since $a_n \simeq n^2$ this limits the value of ϵ to

$$|\epsilon| \leqslant \frac{\sqrt{a}}{2}.$$

For optical frequencies this gives $|\epsilon| < 10^{-5}$ which is an acceptable restriction. In the limit of low frequencies however we would have to require that $|\epsilon| \to 0$, which appears to be an implausible result on physical grounds. We therefore conjecture that the power series for $a_n(q)$ and $b_n(q)$ and hence for $\Delta_n(q)$ have a radius of convergence proportional to n^2. This conjecture has been made before[13] as a result of numerical studies. It implies that the corresponding power series expansions in ϵ have a radius of convergence larger than some constant independent of n. This is supported by the WKB result [Eq. (36)] showing that for large n the radius of convergence is $|\epsilon| = 1/2$.

Let us finally identify the region of validity of the asymptotic formulas for a employed in Ref. 7. This region is obtained for

$$a < 2|q|; \quad |\epsilon| > 1/2, \tag{39a}$$

corresponding to states bound in the wells of the periodic

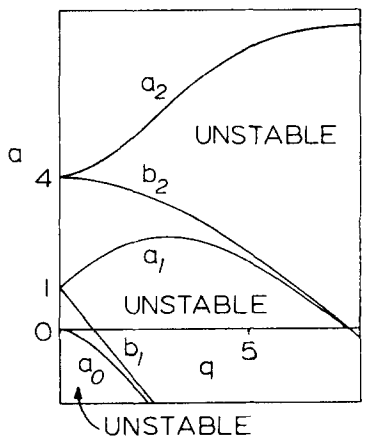


FIG. 2. Stability diagram for Mathieu's equation when a and q are small.

potential $2q \cos 2z$. For the low lying states the WKB approximation again gives good results and one obtains for the first terms

$$a_n(q) = -2|q| + 2(2n+1)\sqrt{|q|}, \tag{39b}$$

which is Eq. (10) of Ref. 7.

Equation (39b) is just the energy spectrum of the harmonic oscillator corresponding to an expansion of the cosine in Eq. (33a) to lowest order.

This shows that the region of validity of asymptotic formula (39b) (see Fig. 1) is given by

$$|a/2q + 1| \ll 1,$$

which is clearly inconsistent with Eq. (34b).

5. SUMMARY AND CONCLUSION

The nonrelativistic theory of a spinless quantum particle moving in a plane wave electromagnetic field does predict forbidden regions for the eigenvalues of some momentum like observable. No such region exists in the relativistic theory.

The occurrence of such regions was shown to be an artifact of the nonrelativistic approximation. This artifact is introduced by the formal possibility of transforming to frames moving at the speed of light in which the phase of the plane wave does not change in time. In these frames bound states of the particle are possible both in the classical and quantum theory. Since these particles would move at the speed of light in the original frame of reference, the corresponding states are clearly unphysical and violate the conditions of validity of the nonrelativistic approximation.

The case of a monochromatic wave of circular polarization leads to the well studied Mathieu equation. Recent results on the gap asymptotics for this equation show that the size of the forbidden regions in the momentum spectrum of the particle vanishes asymptotically like a very large power of the ratio of the particle velocity to the velocity of light. This result holds for those regions of the stability plane in which the conditions of validity of the nonrelativistic approximation are satisfied. Asymptotic expressions for the spectrum considered before in the literature are shown to be restricted to regions where these conditions are violated and which correspond to unphysical bound states.

ACKNOWLEDGMENT

This work was supported by the U.S. Air Force Office of Scientific Research.

[1] L. U. Keldysh, Zh. Eksp. Teor. Fiz. **67**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1965)].
[2] H. D. Jones and H. R. Reiss, Phys. Rev. B **16**, 2466 (1977).
[3] J. Bergou and S. Varro, J. Phys. A **14**, 2281 (1981).
[4] L. S. Brown and T. W. B. Kibble, Phys. Rev. **133**, A705 (1964).
[5] W. Gordon, Z. Phys. **40**, 117 (1926).
[6] D. M. Volkov, Z. Phys. **94**, 250 (1935).
[7] N. Tzoar and J. I. Gersten, Phys. Rev. B **12**, 1132 (1975).
[8] M. Wantelet and L. D. Laude, Phys. Status Solidi B **89**, 275 (1978).
[9] J. Bergou and S. Varro, J. Phys. A **13**, 3553 (1980).
[10] V. B. Berestetskii, E. M. Liftshitz, and L. P. Pitaevskii, *Relativistic Quantum Theory* (Pergamon, Oxford, 1971), §33.
[11] E. Harrell, Am. J. Math. (to be published), see also Ref. 12.
[12] J. Avron and B. Simon, Ann. Phys. (N.Y.) **134**, 76 (1981).
[13] J. Meixner, F. W. Schäfke, and G. Wolf, *Mathieu Functions and Spheroidal Functions and their Mathematical Foundations, Further Studies* (Springer, New York, 1980).
[14] W. Becker, Physica A **87**, 601 (1977).
[15] H. Goldstein, *Classical Mechanics* (Addison Wesley, Reading, MA, 1965).
[16] S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper and Row, New York, 1962), Chap. 3.c.
[17] R. E. Langer, Am. Math. Soc. Trans. **36**, 637 (1934).

Strong fluctuations in the spinor wave optics. Mutual coherence function

P. Hillion
Institut Henri Poincaré, 75231-Paris, France

S. Quinnez
Faculté Sciences Mirande, 21000-Dijon, France

(Received 3 September 1981; accepted for publication 18 December 1981)

In this paper, we apply the spinor wave optics to light beam propagation in a strong turbulent medium. Using the Markov technique, we give the equations for the mutual coherence 4-vector $\Gamma_\mu(z;\rho_1,\rho_2)$, $\mu = 0,1,2,3$, whose third component $\Gamma_3(z;\rho_1,\rho_2)$ is the usual mutual coherence function (the three other components have to do with the Poynting vector). Then, scaling transverse and longitudinal distances by the beam waist ω_0 and diffraction length l, respectively, an expansion procedure in powers of $\omega_0 l^{-1}$ is developed. The zeroth-order approximation is discussed, and it is shown that this approximation gives the same mutual coherence function as the paraxial optics. The case of a plane wave is also considered.

PACS numbers: 42.20.Ee

I. INTRODUCTION

We recently developed,[1,2] for light propagation in isotropic media, a spinor theory simpler than Maxwell's theory and taking into account polarization contrary to the scalar theory.

Here, in the framework of this theory and using the Markov method,[3] we consider the mutual coherence function for a light beam in a strong turbulent medium while, in a second paper, we discuss the intensity covariance function.

We shall see that one is lead to a mutual coherence 4-vector, and we give the system of equations for this 4-vector. A zeroth-order approximation to the solution of this system is obtained using the parameter $\omega_0 l^{-1}$, where ω_0 is the beam waist and l the diffraction length. For the mutual coherence function this approximate solution is also the solution of the usual parabolic equation. This result should not come as a surprise since Lax *et al.*,[4] using an expansion procedure in powers of $\omega_0 l^{-1}$, proved that the parabolic equation is a zeroth order approximation to Maxwell's equations. We also discuss the case of plane waves.

We first sum up the main points of the spinor wave formalism, and we start with both spinor equations

$$[i\sigma^j \vec{\partial}_j + K_0 n(\bar{x})]\psi(\bar{x}) = 0, \tag{1a}$$

$$\psi^+(\bar{x})[-i\sigma^j \overleftarrow{\partial}_j + K_0 n(\bar{x})] = 0, \quad i = \sqrt{-1}. \tag{1b}$$

$\psi(\bar{x})$ and its Hermitian conjugate $\psi^+(\bar{x})$ are two component spinors, K_0 is the wavenumber, $n(\bar{x})$ the refractive index, ∂_j, $j = 1,2,3$, the derivative with respect to x_j, the σ_j's are the Pauli matrices, $\bar{x}$ denotes an arbitrary point in $\mathbb{R}^3$, and we use the summation convention.

Then the intensity $I(\bar{x})$ and the Poynting vector $P_j(\bar{x})$, $j = 1,2,3$, are defined by the relations

$$I(\bar{x}) = \psi^+(\bar{x})\psi(\bar{x}), \quad P_j(\bar{x}) = \psi^+(\bar{x})\sigma_j\psi(\bar{x}), \quad j = 1,2,3, \tag{2a}$$

and a simple calculation gives

$$P_j(\bar{x})P^j(\bar{x}) = I^2(\bar{x}). \tag{2b}$$

Now we would like to obtain the average intensity $\langle I(\bar{x})\rangle$, where the angle brackets $\langle\ \rangle$ denote an ensemble average for a light beam propagating along the 0z direction in a turbulent medium characterized by a refractive index $n(\bar{x})$ with

$$n(\bar{x}) = 1 + \mu(\bar{x}), \quad |\mu(\bar{x})| \ll 1, \tag{3a}$$

where $\mu(\bar{x})$ is a Gaussian random field with zero mean $\langle \mu(\bar{x})\rangle = 0$ and with correlation function (for a beam propagating along 0z)

$$B_n(\bar{x},\bar{x}') = \langle \mu(\bar{x})\mu(\bar{x}')\rangle = A_n(\rho - \rho')\delta(z - z'), \quad \bar{x} = (\rho,z). \tag{3b}$$

Here $\delta(z)$ is the Dirac distribution while ρ denotes the transverse coordinates (x, y). According to (2a), one has

$$\langle I(z,\rho_1)\rangle = \langle \psi^+(z,\rho_1)\psi(z,\rho_1)\rangle = \lim_{\rho_2\to\rho_1} \langle \psi^+(z,\rho_2)\psi(z,\rho_1)\rangle,$$

$$\langle P_j(z,\rho_1)\rangle = \langle \psi^+(z,\rho_1)\sigma_j\psi(z,\rho_1)\rangle$$

$$= \lim_{\rho_2\to\rho_1} \langle \psi^+(z,\rho_2)\sigma_j\psi(z,\rho_1)\rangle, \quad j = 1,2,3$$

where $\langle \psi^+(z,\rho_2)\psi(z,\rho_1)\rangle$ and $\langle \psi^+(z,\rho_2)\sigma_j\psi(z,\rho)\rangle$ are, respectively, the scalar and vector mutual coherence functions.

As previously stated, we give in this paper the system of equations for the 4-vector $\langle \psi^+(z,\rho_2)\sigma_\mu\psi(z,\rho_1)\rangle$, $\mu = 1,2,3$ (σ_0 is the 2×2 identity matrix), and we then discuss an approximate solution of these equations; but we first start with the calculation of the coherent intensity that is, according to (2a),

$$I_c(\rho,z) = \langle \psi^+(z,\rho)\rangle\ \langle \psi(z,\rho)\rangle \tag{4}$$

in order to illustrate the Markov technique in the case of the spinor formalism.

II. COHERENT INTENSITY

The functional (or variational derivative)[5] is an essential tool of the Markov technique; we shall assume that the functional derivative has all the properties of the ordinary derivative, in particular with respect to its definition for an infinite interval and for differentiation under the integral sign. We shall also assume that all the functionals we used have all the

 0022-2488/83/030712-09$02.50

properties of differentiation needed to make our calculations valid. We intend to discuss these points in a forthcoming paper.

Let us then take the ensemble average of Eq. (1a) [one has similar computations with Eq. (1b)]; we get

$$(i\sigma^j\vec{\partial}_j + K_0)\langle\psi(z,\rho)\rangle + K_0\langle\mu(z,\rho)\psi(z,\rho)\rangle = 0, \tag{5}$$

where we used (3a). Since $\mu(\bar{x})$ is Gaussian, we may use the Novikov–Furutsu formula[6]

$$\begin{aligned}\langle\mu(z,\rho)\psi(z,\rho)\rangle &= \int dz'\,d\rho'\,\langle\mu(z,\rho)\mu(z,\rho')\rangle\left\langle\frac{\delta\psi(z,\rho)}{\delta\mu(z',\rho')}\right\rangle \\ &= \int dz'\,\delta(z-z')\int d\rho'\,A_n(\rho-\rho')\left\langle\frac{\delta\psi(z,\rho)}{\delta\mu(z',\rho')}\right\rangle \\ &= \int d\rho'\,A_n(\rho-\rho')\left\langle\frac{\delta\psi(z,\rho)}{\delta\mu(z,\rho')}\right\rangle,\end{aligned} \tag{6}$$

where $\delta\psi/\delta\mu$ is a variational derivative and Eq. (5) becomes

$$(i\sigma^j\vec{\partial}_j + K_0)\langle\psi(\bar{x})\rangle + K_0\int d\rho'\,A_n(\rho-\rho')\left\langle\frac{\delta\psi(z,\rho)}{\delta\mu(z,\rho')}\right\rangle = 0. \tag{5'}$$

To compute the last term on the left-hand side of (5'), we write (1a) in the following form where the Greek indices take the values 1,2 while the summation convention is used:

$$i\sigma_3\partial_3\psi(\bar{x}) + (i\sigma^\alpha\partial_\alpha + K_0)\psi(\bar{x}) + K_0\mu(\bar{x})\psi(\bar{x}) = 0, \quad \partial_3 = \frac{\partial}{\partial z},$$

and by integrating with respect to z from $z=0$ to z, it becomes

$$i\sigma_3\{\psi(z,\rho)-\psi(0,\rho)\} + (i\sigma^\alpha\partial_\alpha + K_0)\int_0^z\psi(\zeta,\rho)\,d\zeta + K_0\int_0^z\mu(\zeta,\rho)\psi(\zeta,\rho)\,d\zeta = 0$$

Now we take the variational derivative of this last relation,

$$i\sigma_3\frac{\delta\psi(z,\rho)}{\delta\mu(z,\rho')} + (i\sigma^\alpha\partial_\alpha + K_0)\frac{\delta}{\delta\mu(z,\rho')}\int_0^z\psi(\zeta,\rho)\,d\zeta + K_0\frac{\delta}{\delta\mu(z,\rho')}\int_0^z\mu(\zeta,\rho)\psi(\zeta,\rho)\,d\zeta = 0 \tag{7}$$

and let $z'\to z$. Neglecting the backscattering from $\mu(z',\rho')$ in the region $z'>z$, one has

$$\frac{\delta\psi(z,\rho)}{\delta\mu(z',\rho')} = 0 \quad \text{if } z'>z; \tag{8}$$

then the second term on the left-hand side of (7) becomes

$$\frac{\delta}{\delta\mu(z',\rho')}\int_0^z\psi(\zeta,\rho)\,d\zeta = \int_{z'}^z\frac{\delta\psi(\zeta,\rho)}{\delta\mu(z',\rho')}\,d\zeta,$$

which gives

$$\lim_{z\to z'}\frac{\delta}{\delta\mu(z',\rho')}\int_0^z\psi(\zeta,\rho)\,d\zeta = 0. \tag{7'}$$

Let us now consider the last term on the left-hand side of (7) and let $L(z)$ be:

$$L(z) = \int_0^z\mu(\zeta,\rho)\psi(\zeta,\rho)\,d\zeta. \tag{9}$$

Using the Heaviside function $H(z-\zeta)$, we write this expression

$$L(z) = \int_0^\infty d\zeta\int_{-\infty}^\infty d\rho''\,[H(z-\zeta)\delta(\rho-\rho'')\mu(\zeta,\rho'')\psi(\zeta,\rho'')],$$

where $\delta(\rho)$ is the Dirac distribution. Now we use the following result[6]: If one has

$$L(z) = \int_a^b F(z,\xi,\mu(\xi))\,d\xi,$$

then the variational derivative of L with respect to μ is

$$\frac{\delta L(z)}{\delta\mu(z')} = \frac{\delta F(z,z',\mu(z'))}{\delta\mu(z')}. \tag{10}$$

Using this last result, (9) becomes

$$\frac{\delta L(z)}{\delta\mu(z',\rho')} = H(z-z')\delta(\rho-\rho')\psi(z',\rho') + \int_0^\infty d\xi\times\int_{-\infty}^\infty d\rho''\,[H(z-\xi)\delta(\rho-\rho'')\mu(\xi,\rho'')\times\frac{\delta\psi(\xi,\rho'')}{\delta\mu(z',\rho')}],$$

and using (8) together with $H(0)=\frac{1}{2}$, one obtains

$$\lim_{z\to z'}\frac{\delta L(z)}{\delta\mu(z',\rho')} = \tfrac{1}{2}\delta(\rho-\rho')\psi(z,\rho'). \tag{7''}$$

Substituting (7') and (7'') into (7) gives

$$\frac{\delta\psi(z,\rho)}{\delta\mu(z',\rho')} = \tfrac{1}{2}iK_0\sigma_3\delta(\rho-\rho')\psi(z,\rho'), \tag{11}$$

which implies

$$\left\langle\frac{\delta\psi(z,\rho)}{\delta\mu(z,\rho')}\right\rangle = \tfrac{1}{2}iK_0\sigma_3\delta(\rho-\rho')\langle\psi(z,\rho')\rangle. \tag{11'}$$

Taking (11') into account, Eq. (5') becomes

$$(i\sigma^j\partial_j + K_0)\langle\psi(\bar{x})\rangle + \tfrac{1}{2}iK_0^2\sigma_3\times\int d\rho'\,\delta(\rho-\rho')A_n(\rho-\rho')\langle\psi(z,\rho')\rangle = 0;$$

that is,

$$[\sigma^j\partial_j - iK_0 + \tfrac{1}{2}K_0^2\sigma_3A_n(0)]\,\langle\psi(z,\rho)\rangle = 0. \tag{12}$$

Together with the boundary condition at $z=0$,

$$\langle\psi(0,\rho)\rangle = \psi_0(\rho), \tag{12'}$$

Eq. (12) completely determines the coherent field $\langle\psi(z,\rho)\rangle$. The solution of (12) can be obtained by writing

$$\langle\psi(z,\rho)\rangle = \Phi(z,\rho)e^{-\alpha_0 z} \tag{13}$$

where the attenuation constant α_0 is

$$\alpha_0 = \tfrac{1}{2}K_0^2A_n(0). \tag{13'}$$

Substituting (13) into (12) shows that $\Phi(z,\rho)$ is the field in free space satisfying

$$(\sigma^j\partial_j - iK_0)\Phi(z,\rho) = 0. \tag{13''}$$

Remark: Starting from Eq. (1b), one obtains similar results for $\psi^+(z,\rho)$:

$$\langle\psi^+(z,\rho)\rangle = \Phi^+(z,\rho)e^{-\alpha_0 z}, \quad \Phi^+(z,\rho)(\sigma^j\overleftarrow{\partial}_j + iK_0) = 0.$$

We now have to discuss the fact that the light beam is propagating along the 0z axis, which implies in free space

$$P_\alpha(\bar{x}) = 0, \quad \alpha = 1,2, \quad I(\bar{x}) = |P_3(\bar{x})| \tag{14}$$

Now let ψ_1, ψ_2 be the two components of the spinor $\psi,\tilde{\psi}_1$, $\tilde{\psi}_2$ their complex conjugate. From the representation of the Pauli matrices!

$$\sigma_1=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix},\quad \sigma_2=\begin{bmatrix}0 & -i\\ i & 0\end{bmatrix},\quad \sigma_3=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}, \tag{15}$$

we get

$$\begin{aligned}
&P_1(\bar{x})=\tilde{\psi}_1(\bar{x})\psi_2(\bar{x})+\tilde{\psi}_2(\bar{x})\psi_1(\bar{x}),\\
&P_2(\bar{x})=-i(\tilde{\psi}_1(\bar{x})\psi_2(\bar{x})-\tilde{\psi}_2(\bar{x})\psi_1(\bar{x})),\\
&P_3(\bar{x})=\tilde{\psi}_1(\bar{x})\psi_1(\bar{x})-\tilde{\psi}_2(\bar{x})\psi_2(\bar{x}),\\
&I(\bar{x})=\tilde{\psi}_1(\bar{x})\psi_1(\bar{x})+\tilde{\psi}_2(\bar{x})\psi_2(\bar{x}),
\end{aligned} \tag{15'}$$

and the relations (14) imply $\psi_2(\bar{x})=\tilde{\psi}_2(\bar{x})=0$.

The generalization of (14) to the coherent field in a random medium is

$$P_{\alpha,c}(\bar{x})=0,\quad \alpha=1,2,\quad I_c(\bar{x})=|P_{3,c}(\bar{x})|, \tag{16}$$

which leads to $\langle\psi_2(\bar{x})\rangle=\langle\tilde{\psi}_2(\bar{x})\rangle=0$; that is, from (13), $\varphi_2(\bar{x})=\bar{\varphi}_2(\bar{x})=0$, where φ_1 and φ_2 are the two components of the spinor Φ.

Using this last result together with Eqs. (13) and (14), one obtains for the coherent intensity

$$I_c(\rho,z)=|\varphi_1(\rho,z)|^2e^{-2\alpha_0 z}, \tag{17}$$

which is similar to the expression supplied by the parabolic equation[6]

$$I_c(\rho,z)=|U(\rho,z)|^2e^{-2\alpha_0 z},$$

where $U(\rho,z)$ is a solution of the scalar equation $(2iK_0\,\partial/\partial z+\partial^\alpha\partial_\alpha)U(\rho,z)=0$.

This result should not come as a surprise since both coherent fields Φ and U are solutions of a first-order differential equation (with respect to z) which is charcteristic of a first order Markov process. Thus the coherent intensity can be considered as the autocorrelation function of a stationary Gauss–Markov process with power spectral density $p(\omega)$,

$$p(\omega)=4\alpha_0\sigma^2/(\omega^2+4\alpha_0^2),\quad \sigma^2=|\varphi_1(\rho,z)|^2$$
$$[\text{resp. } \sigma^2=|U(\rho,z)|^2];$$

the random medium acts as a band pass filter with bandwidth $2\alpha_0$.

Remark: Many people use, instead of $B_n(\bar{x},\bar{x}')$, the correlation function

$$B_\epsilon(\bar{x},\bar{x}')=\langle\epsilon(\bar{x}),\epsilon(\bar{x}')\rangle=\delta(z-z')A_\epsilon(\rho,\rho'),$$

where $\epsilon(\bar{x})$ is the dielectric constant and one has

$$B_\epsilon(\bar{x},\bar{x}')=4B_n(\bar{x},\bar{x}'),\quad \alpha_0=\tfrac{1}{2}K_0^2A_n(0)=\tfrac{1}{8}K_0^2A_\epsilon(0).$$

We shall now consider the mutual coherence 4-vector.

III. EQUATION FOR THE MUTUAL COHERENCE 4-VECTOR

A. Notation

In order to simplify later computations, we use the following notation: starting with the current 4-vector,

$$Z_\mu(z;\rho_1,\rho_2)=\psi^+(z,\rho_2)\sigma_3\sigma_\mu\psi(z,\rho_1),\quad \mu=0,1,2,3, \tag{18}$$

we define the functionals

$$\Gamma_\mu(z;\rho_1,\rho_2)=\langle\psi^+(z,\rho_2)\sigma_3\sigma_\mu\psi(z,\rho_1)\rangle,\quad \mu=0,1,2,3, \tag{19}$$

with

$$\sigma_\mu=\{\sigma_0,\sigma_j\},\quad \mu=0,1,2,3,\quad j=1,2,3.$$

According to these conventions, the scalar mutual coherence function is

$$\Gamma_3(z;\rho_1,\rho_2)=\langle\psi^+(z,\rho_2)\psi(z,\rho_1)\rangle \tag{19'}$$

[the usual notation[3,6] is $\Gamma(z;\rho_1,\rho_2)$ while the transverse components of the mutual coherence vector are $\Gamma_\alpha(z;\rho_1,\rho_2)$, $\alpha=1,2$, and the longitudinal component $\Gamma_0(z;\rho_1,\rho_2)$].

B. Fundamental equations

We start with Eq. (1a) written for $\bar{x}=\bar{x}_1$ and Eq. (1b) written for $\bar{x}=\bar{x}_2$:

$$[i\sigma^j\vec{\partial}_j^1+K_0n(\bar{x}_1)]\psi(\bar{x}_1)=0, \tag{20a}$$

$$\psi^+(\bar{x}_2)[-i\sigma^j\overleftarrow{\partial}_j^2+K_0n(\bar{x}_2)]=0; \tag{20b}$$

the notations $\vec{\partial}_j^1$, $\overleftarrow{\partial}_j^2$ are obvious.

Let us first multiply (20a) on the left by $\psi^+(\bar{x}_2)\sigma_3$ and (20b) on the right by $\sigma_3\psi(\bar{x}_1)$; one has

$$i\psi^+(\bar{x}_2)\sigma_3\sigma^j\vec{\partial}_j^1\psi(\bar{x}_1)+K_0n(\bar{x}_1)\psi^+(\bar{x}_2)\sigma_3\psi(\bar{x}_1)=0$$

and

$$-i\overleftarrow{\partial}_j^2\psi^+(\bar{x}_2)\sigma^j\sigma_3\psi(\bar{x}_1)+K_0n(\bar{x}_2)\psi^+(\bar{x}_2)\sigma_3\psi(\bar{x}_1)=0.$$

for $z=z_1=z_2$, these equations become

$$\begin{aligned}
&i\psi^+(z,\rho_2)\left(\frac{\partial}{\partial z}\psi(z,\rho_1)\right)+i\psi^+(z,\rho_2)\sigma_3\sigma^\alpha\vec{\partial}_\alpha^1\psi(z,\rho_1)\\
&\quad+K_0n(z,\rho_1)Z_0(z;\rho_1,\rho_2)=0,\\
&\left(-i\frac{\partial}{\partial z}\psi^+(z,\rho_2)\right)\psi(z,\rho_1)-i\overleftarrow{\partial}_\alpha^2\psi^+(z,\rho_2)\sigma^\alpha\sigma_3\psi(z,\rho_1)\\
&\quad+K_0n(z,\rho_2)Z_0(z;\rho_1,\rho_2)=0;
\end{aligned}$$

substracting the second equation from the first one and averaging, we get

$$\begin{aligned}
&i\frac{\partial}{\partial z}\Gamma_3(z;\rho_1,\rho_2)+i(\partial_\alpha^1-\partial_\alpha^2)\Gamma^\alpha(z;\rho_1,\rho_2)\\
&\quad+K_0\langle[\mu(z,\rho_1)-\mu(z,\rho_2)]Z_0(z;\rho_1,\rho_2)\rangle=0,
\end{aligned} \tag{21}$$

where we used the relation $\sigma_\alpha\sigma_3+\sigma_3\sigma_\alpha=0$, $\alpha=1,2$. In Eq. (21) ∂_α^1 and ∂_α^2 denote the derivatives with respect to x_1, y_1 and x_2, y_2, respectively. Now, to compute the last term on the left-hand side of (21), we still use the Novikov–Furutsu formula

$$\langle\mu(z,\rho_1)Z_0(z;\rho_1,\rho_2)\rangle$$
$$=\int d\rho_1'\,A_n(\rho_1-\rho_1')\left\langle\frac{\delta Z_0(z;\rho_1,\rho_2)}{\delta\mu(z,\rho_1')}\right\rangle \tag{22}$$

with

$$\begin{aligned}
\frac{\delta Z_0(z;\rho_1,\rho_2)}{\delta\mu(z,\rho_1')}&=\frac{\delta\psi^+(z,\rho_2)}{\delta\mu(z,\rho_1')}\sigma_3\psi(z,\rho_1)\\
&\quad+\psi^+(z,\rho_2)\sigma_3\frac{\delta\psi(z,\rho_1)}{\delta\mu(z,\rho_1')}
\end{aligned}$$

from (18); then using (11) and averaging,

$$\left\langle \frac{\delta Z_0(z;\rho_1,\rho_2)}{\delta\mu(z,\rho_1')} \right\rangle = -\tfrac{1}{2}iK_0\delta(\rho_2-\rho_1')\Gamma_3(z;\rho_1,\rho_1') + \tfrac{1}{2}iK_0\delta(\rho_1-\rho_1')\Gamma_3(z;\rho_2,\rho_1) \quad (23)$$

Substituting (23) into (22) gives:

$$\langle \mu(z,\rho_1)Z_0(z;\rho_1,\rho_2)\rangle = \tfrac{1}{2}iK_0[A_n(0)-A_n(\rho_1-\rho_2)]\Gamma_3(z;\rho_1,\rho_2). \quad (24a)$$

By interchanging ρ_1 and ρ_2 in (22) and taking the Hermitian conjugate, we get

$$\langle \mu(z,\rho_2)Z_0^{+}(z;\rho_2,\rho_1)\rangle = \int d\rho_1'\, A_n(\rho_2-\rho_1')\left\langle \frac{\delta Z_0^{+}(z;\rho_2,\rho_1)}{\delta\mu(z,\rho_1')}\right\rangle,$$

and, since, according to (18), one has $Z_0^{+}(z;\rho_2,\rho_1) = Z_0(z;\rho_1,\rho_2)$, the last relation becomes

$$\langle \mu(z,\rho_2)Z_0(z;\rho_1,\rho_2)\rangle = \int d\rho_1'\, A_n(\rho_2-\rho_1')\left\langle \frac{\delta Z_0(z;\rho_1,\rho_2)}{\delta\mu(z,\rho_1')}\right\rangle. \quad (22')$$

Substituting (23) into (22'), we get

$$\langle \mu(z,\rho_2)Z_0(z;\rho_1,\rho_2)\rangle = -\tfrac{1}{2}iK_0[A_n(0)-A_n(\rho_2-\rho_1)]\Gamma_3(z;\rho_1,\rho_2)$$

and, assuming $A_n(\rho_1-\rho_2)=A_n(\rho_2-\rho_1)$,

$$\langle \mu(z,\rho_2)Z_0(z;\rho_1,\rho_2)\rangle = -\tfrac{1}{2}iK_0[A_n(0)-A_n(\rho_1-\rho_2)]\Gamma_3(z;\rho_1,\rho_2). \quad (24b)$$

Using (24a) and (24b), Eq. (21) becomes

$$\left(\frac{\partial}{\partial z}+K_0^2[A_n(0)-A_n(\rho_1-\rho_2)]\right)\Gamma_3(z;\rho_1,\rho_2) + (\partial_\alpha^1-\partial_\alpha^2)\Gamma^\alpha(z;\rho_1,\rho_2)=0. \quad (25)$$

This is the first equation for the mutual coherence 4-vector. We now multiply (20a) on the left by $\psi(\bar{x}_2)$ and (20b) on the right by $\psi(\bar{x}_1)$. For $z=z_1=z_2$, one has

$$i\psi^{+}(z,\rho_2)\sigma_3\left(\frac{\partial}{\partial z}\psi(z,\rho_1)\right)+i\psi^{+}(z,\rho_2)\sigma^\alpha\partial_\alpha^1\psi(z,\rho_1) + K_0n(z,\rho_1)\psi^{+}(z,\rho_2)\psi(z,\rho_1)=0,$$

$$-i\left(\frac{\partial}{\partial z}\psi^{+}(z,\rho_2)\right)\sigma_3\psi(z,\rho_1)-i\partial_\alpha^2\psi^{+}(z,\rho_2)\sigma^\alpha\psi(z,\rho_1) + K_0n(z,\rho_2)\psi^{+}(z,\rho_2)\psi(z,\rho_1)=0;$$

subtracting the second equation from the first one and averaging, one obtains

$$i\frac{\partial}{\partial z}\Gamma_0(z;\rho_1,\rho_2)+i(\partial_\alpha^1+\partial_\alpha^2)\langle\psi^{+}(z,\rho_2)\sigma^\alpha\psi(z,\rho_1)\rangle + K_0\langle[\mu(z,\rho_1)-\mu(z,\rho_2)]Z_3(z;\rho_1,\rho_2)\rangle=0. \quad (26)$$

The last term on the left-hand side of (26) is computed as for Eq. (21); we just have to interchange the couples (Z_0,Γ_3) and (Z_3,Γ_0). Still assuming $A_n(\rho_1-\rho_2)=A_n(\rho_2-\rho_1)$, one has

$$\langle[\mu(z,\rho_1)-\mu(z,\rho_2)]Z_3(z;\rho_1,\rho_2)\rangle = iK_0[A_n(0)-A_n(\rho_1-\rho_2)]\Gamma_0(z;\rho_1,\rho_2). \quad (26')$$

Now according to the Pauli algebra, one has

$$\sigma_\alpha = i\epsilon_{\alpha\beta}\sigma^3\sigma^\beta,\quad \alpha,\beta=1,2,$$
$$\epsilon_{11}=\epsilon_{22}=0,\quad \epsilon_{12}=-\epsilon_{21}=1, \quad (27)$$

which implies

$$\langle\psi^{+}(z,\rho_2)\sigma^\alpha\psi(z,\rho_1)\rangle = i\epsilon^{\alpha\beta}\Gamma_\beta(z;\rho_1,\rho_2). \quad (26'')$$

Using (26') and (26''), Eq. (26) becomes

$$\left(\frac{\partial}{\partial z}+K_0^2[A_n(0)-A_n(\rho_1-\rho_2)]\right)\Gamma_0(z;\rho_1,\rho_2) + i\epsilon^{\alpha\beta}(\partial_\alpha^1+\partial_\alpha^2)\Gamma_\beta(z;\rho_1,\rho_2)=0. \quad (28)$$

This is the second equation for the mutual coherence 4-vector. To obtain the last two equations, we multiply (20a) on the left by $\psi^{+}(\bar{x}_2)\sigma_\alpha$, $\alpha=1,2$, and (20b) on the right by $\sigma_\alpha\psi(\bar{x}_1)$, $\alpha=1,2$. For $z=z_1=z_2$, we get

$$i\psi^{+}(z,\rho_2)\sigma_\alpha\sigma_3\left(\frac{\partial}{\partial z}\psi(z,\rho_1)\right)+i\psi^{+}(z,\rho_2)\sigma_\alpha\sigma^\beta\partial_\beta^1\psi(z,\rho_1) + K_0n(z,\rho_1)\psi^{+}(z,\rho_2)\sigma_\alpha\psi(z,\rho_1)=0,$$

$$-i\left(\frac{\partial}{\partial z}\psi^{+}(z,\rho_2)\right)\sigma_3\sigma_\alpha\psi(z,\rho_1)-i\partial_\beta^2\psi(z,\rho_2)\sigma^\beta\sigma_\alpha\psi(z,\rho_1) + K_0n(z,\rho_2)\psi^{+}(z,\rho_2)\sigma_\alpha\psi(z,\rho_1)=0,\quad \alpha=1,2;$$

summing these two equations, averaging, and using the following relations,

$$\sigma_\alpha\sigma_3+\sigma_3\sigma_\alpha=0,\quad \sigma_\alpha\sigma_\beta=\delta_{\alpha\beta}+i\epsilon_{\alpha\beta}\sigma_3,\quad \alpha,\beta=1,2, \quad (27')$$

one obtains

$$-i\frac{\partial}{\partial z}\Gamma_\alpha(z;\rho_1,\rho_2)+i(\partial_\alpha^1-\partial_\alpha^2)\Gamma_3(z;\rho_1,\rho_2) - \epsilon_\alpha^\beta(\partial_\beta^1+\partial_\beta^2)\Gamma_0(z;\rho_1,\rho_2)+2iK_0\epsilon_\alpha^\beta\Gamma_\beta(z;\rho_1,\rho_2) + iK_0\epsilon_\alpha^\beta\langle[\mu(z,\rho_1)+\mu(z,\rho_2)]Z_\beta(z;\rho_1,\rho_2)\rangle=0,\quad \alpha=1,2. \quad (29)$$

The last term on the left-hand side of (29) is computed as previously,

$$\langle\mu(z,\rho_1)Z_\beta(z;\rho_1,\rho_2)\rangle = \int d\rho_1'\, A_n(\rho_1-\rho_1')\left\langle\frac{\delta Z_\beta(z;\rho_1,\rho_2)}{\delta\mu(z,\rho_1')}\right\rangle,\quad \beta=1,2 \quad (30)$$

with

$$\frac{\delta Z_\beta(z;\rho_1,\rho_2)}{\delta\mu(z,\rho_1')} = \frac{\delta\psi^{+}(z,\rho_2)}{\delta\mu(z,\rho_1')}\sigma_3\sigma_\beta\psi(z,\rho_1) + \psi^{+}(z,\rho_2)\sigma_3\sigma_\beta\frac{\delta\psi(z,\rho_1)}{\delta\mu(z,\rho_1')}$$
$$= -\tfrac{1}{2}iK_0\delta(\rho_2-\rho_1')\psi^{+}(z,\rho_1')\sigma_\beta\psi(z,\rho_1) + \tfrac{1}{2}iK_0\delta(\rho_1-\rho_1')\psi^{+}(z,\rho_2)\sigma_3\sigma_\beta\sigma_3\psi(z,\rho_1')$$
$$= -\tfrac{1}{2}iK_0\delta(\rho_2-\rho_1')\psi^{+}(z,\rho_1')\sigma_\beta\psi(z,\rho_1) - \tfrac{1}{2}iK_0\delta(\rho_1-\rho_1')\psi^{+}(z,\rho_2)\sigma_\beta\psi(z,\rho_1'),\quad \beta=1,2.$$

from (11) and (18). Using (27) and averaging, we get

$$\left\langle\frac{\delta Z_\beta(z;\rho_1,\rho_2)}{\delta\mu(z,\rho_1')}\right\rangle = \tfrac{1}{2}K_0\epsilon_\beta^\gamma[\delta(\rho_2-\rho_1')\Gamma_\gamma(z;\rho_1,\rho_1')+\delta(\rho_1-\rho_1')\Gamma_\gamma(z;\rho_1',\rho_2)],\quad \beta=1,2. \quad (30')$$

Substituting (30') into (30) gives

$$\langle \mu(z,\rho_1)Z_\beta(z;\rho_1,\rho_2)\rangle = \tfrac{1}{2}K_0\epsilon_\beta^\gamma[A_n(0)+A_n(\rho_1-\rho_2)]\Gamma_\gamma(z;\rho_1,\rho_2), \quad \beta=1,2. \tag{31a}$$

By interchanging ρ_1 and ρ_2 in (30) and taking the Hermitian conjugate, we get

$$\langle \mu(z,\rho_2)Z_\beta^+(z;\rho_2,\rho_1)\rangle = \int d\rho_1'\, A_n(\rho_2-\rho_1')\left\langle\frac{\delta Z_\beta^+(z;\rho_2,\rho_1)}{\delta\mu(z,\mu_1')}\right\rangle, \quad \beta=1,2,$$

but according to (18), one has

$$Z_\beta^+(z;\rho_2,\rho_1) = -Z_\beta(z;\rho_1,\rho_2), \quad \beta=1,2.$$

Thus the last relation becomes

$$\langle \mu(z,\rho_2)Z_\beta(z;\rho_1,\rho_2)\rangle = \int d\rho_1'\, A_n(\rho_2-\rho_1')\left\langle\frac{\delta Z_\beta(z;\rho_1,\rho_2)}{\delta\mu(z,\mu_1')}\right\rangle, \quad \beta=1,2;$$

then using (30') together with the assumption $A_n(\rho_1-\rho_2) = A_n(\rho_2-\rho_1)$, we get

$$\langle \mu(z,\rho_2)Z_\beta(z;\rho_1,\rho_2)\rangle = \tfrac{1}{2}K_0\epsilon_\beta^\gamma[A_n(0)+A_n(\rho_1-\rho_2)]\Gamma_\gamma(z;\rho_1,\rho_2), \quad \beta=1,2\ . \tag{31b}$$

Substituting (31a) and (31b) into (29) and using the relation

$$\epsilon_\alpha^\beta\epsilon_\beta^\gamma = -\delta_\alpha^{\ \gamma}, \quad \alpha,\gamma=1,2 \tag{32}$$

one obtains

$$\left\{\frac{\partial}{\partial z}+K_0^2[A_n(0)+A_n(\rho_1-\rho_2)]\right\}\Gamma_\alpha(z;\rho_1,\rho_2) - (\partial_\alpha^1-\partial_\alpha^2)\Gamma_3(z;\rho_1,\rho_2) - i\epsilon_\alpha^\beta(\partial_\beta^1+\partial_\beta^2)\Gamma_0(z;\rho_1,\rho_2) - 2K_0\epsilon_\alpha^\beta\Gamma_\beta(z;\rho_1,\rho_2)=0, \quad \alpha=1,2. \tag{33}$$

This completes the system of equations for the mutual coherence 4-vector. If we set

$$F(\rho_1-\rho_2)=K_0^2[A_n(0)-A_n(\rho_1-\rho_2)],$$
$$G(\rho_1-\rho_2)=K_0^2[A_n(0)+A_n(\rho_1-\rho_2)], \tag{34}$$
$$D_f=\frac{\partial}{\partial z}+F(\rho_1-\rho_2), \quad D_g=\frac{\partial}{\partial z}+G(\rho_1-\rho_2).$$

Equations (25), (28), and (33) become

$$D_f\Gamma_3+(\partial_\alpha^1-\partial_\alpha^2)\Gamma^\alpha=0, \tag{35a}$$
$$D_f\Gamma_0+i\epsilon^{\alpha\beta}(\partial_\alpha^1+\partial_\alpha^2)\Gamma_\beta=0, \tag{35b}$$
$$D_g\Gamma_\alpha-(\partial_\alpha^1-\partial_\alpha^2)\Gamma_3-i\epsilon_\alpha^\beta(\partial_\beta^1+\partial_\beta^2)\Gamma_0-2K_0\epsilon_\alpha^\beta\Gamma_\beta=0, \quad \alpha=1,2 \tag{35c}$$

where to simplify we did not write the arguments z,ρ_1 and ρ_2. These equations can still be put in a more compact form. Indeed, using the definition (27) of $\epsilon_{\alpha\beta}$, we can explicitly write Eqs. (35) as

$$D_f\Gamma_3+(\partial_x^1-\partial_x^2)\Gamma_1+(\partial_y^1-\partial_y^2)\Gamma_2=0, \tag{36a}$$
$$D_f\Gamma_0+i(\partial_x^1+\partial_x^2)\Gamma_2-i(\partial_y^1+\partial_y^2)\Gamma_1=0, \tag{36b}$$
$$D_g\Gamma_1-2K_0\Gamma_2-(\partial_x^1-\partial_x^2)\Gamma_3-i(\partial_y^1+\partial_y^2)\Gamma_0=0, \tag{36c}$$
$$D_g\Gamma_2+2K_0\Gamma_1-(\partial_y^1-\partial_y^2)\Gamma_3+i(\partial_x^1+\partial_x^2)\Gamma_0=0. \tag{36d}$$

Let us now introduce the following functionals:

$$\Lambda(z;\rho_1,\rho_2)=\Gamma_3(z;\rho_1,\rho_2)+\Gamma_0(z;\rho_1,\rho_2) = 2\langle\tilde\psi_1(z,\rho_2)\psi_1(z,\rho_1)\rangle, \tag{37a}$$
$$\Xi(z;\rho_1,\rho_2)=\Gamma_3(z;\rho_1,\rho_2)-\Gamma_0(z;\rho_1,\rho_2) = 2\langle\tilde\psi_2(z,\rho_2)\psi_2(z,\rho_1)\rangle, \tag{37b}$$
$$\Omega(z;\rho_1,\rho_2)=\Gamma_1(z;\rho_1,\rho_2)+i\Gamma_2(z;\rho_1,\rho_2) = 2\langle\tilde\psi_1(z,\rho_2)\psi_2(z,\rho_1)\rangle, \tag{37c}$$
$$\Pi(z;\rho_1,\rho_2)=\Gamma_1(z;\rho_1,\rho_2)-i\Gamma_2(z;\rho_1,\rho_2) = -2\langle\tilde\psi_2(z,\rho_2)\psi_1(z,\rho_1)\rangle, \tag{37d}$$

which satisfy the obvious relations

$$\Lambda(z;\rho_1,\rho_2)=\tilde\Lambda(z;\rho_2,\rho_1), \quad \Xi(z;\rho_1,\rho_2)=\tilde\Xi(z;\rho_2,\rho_1),$$
$$\Omega(z;\rho_1,\rho_2)=-\tilde\Pi(z;\rho_2,\rho_1), \tag{38}$$

where the tilde denotes complex conjugation.

Then, summing and substracting successively (36a) and (36b), we get

$$D_f\Lambda+(\partial_x^1-i\partial_y^1)\Omega-(\partial_x^2+i\partial_y^2)\Pi=0,$$
$$D_f\Xi+(\partial_x^1+i\partial_y^1)\cdot\Pi-(\partial_x^2-i\partial_y^2)\Omega=0.$$

The same operations made with (36c) and (36d) give

$$(D_g-2iK_0)\Pi-(\partial_x^1-i\partial_y^1)\Xi+(\partial_x^2-i\partial_y^2)\Lambda=0,$$
$$(D_g+2iK_0)\Omega-(\partial_x^1+i\partial_y^1)\Lambda+(\partial_x^2+i\partial_y^2)\Xi=0.$$

Let us now introduce the spinors X and Y and the derivative operators ∇_α and $\hat\nabla_\alpha$, $\alpha=1,2$:

$$X=\begin{pmatrix}\Lambda\\ \Xi\end{pmatrix}, \quad Y=\begin{pmatrix}\Pi\\ \Omega\end{pmatrix},$$
$$\nabla_\alpha=\partial_\alpha^1-\partial_\alpha^2\sigma_1, \quad \hat\nabla_\alpha=(\partial_\alpha^1+\partial_\alpha^2\sigma_1), \quad \alpha=1,2; \tag{39}$$

then the last four equations take the simple form

$$D_fX+\nabla_\alpha\sigma^\alpha Y=0, \tag{40a}$$
$$(D_g-2iK_0\sigma_3)Y-\sigma^\alpha\nabla_\alpha X=0. \tag{40b}$$

Let us note that the operators ∇_α and σ_α, $\hat\nabla_\alpha$ and σ_α, and ∇_α and $\hat\nabla_\alpha$ do not commute.

IV. SOLUTIONS FOR THE MUTUAL COHERENCE 4-VECTOR

A. Approximation by a perturbation method

Multiplying (40a) by $2iK_0\sigma_3$, taking $2iK_0\sigma_3Y$ from (40b), and using the relation $\sigma_3\nabla_\alpha\sigma^\alpha=-\hat\nabla_\alpha\sigma^\alpha\sigma_3$, we get

$$2iK_0\sigma_3D_f+(\hat\nabla_\alpha\sigma^\alpha\sigma^\beta\nabla_\beta)X=\hat\nabla_\alpha\sigma^\alpha D_gY, \tag{41}$$

but from the relation $\sigma_\alpha\sigma_\beta=\delta_{\alpha\beta}+i\epsilon_{\alpha\beta}\sigma_3$, $\alpha,\beta=1,2$, we deduce

$$\hat\nabla_\alpha\sigma^\alpha\sigma^\beta\nabla_\beta=\delta^{\alpha\beta}\hat\nabla_\alpha\nabla_\beta+i\epsilon^{\alpha\beta}\hat\nabla_\alpha\sigma_3\nabla_\beta = \delta^{\alpha\beta}\hat\nabla_\alpha\nabla_\beta+i\epsilon^{\alpha\beta}\hat\nabla_\alpha\hat\nabla_\beta\sigma_3.$$

The second term on the right-hand side of the last relation is zero (antisymmetry of $\epsilon_{\alpha\beta}$) and from the definition (39) of the operators ∇_α and $\hat\nabla_\alpha$, one has

$$\hat\nabla_\alpha\sigma^\alpha\sigma^\beta\nabla_\beta=\partial_\alpha^{\,1}\partial^{\alpha,1}-\partial_\alpha^{\,2}\partial^{\alpha,2}=\Delta_1-\Delta_2,$$

where Δ denotes the transverse Laplacian operator. So Eq.

(41) becomes

$$(2iK_0\sigma_3 D_f + \Delta_1 - \Delta_2)X = \hat{\nabla}_\alpha \sigma^\alpha D_g Y. \tag{41'}$$

To obtain the solutions of Eqs. (40b) and (41'), we now introduce the dimensionless coordinates previously used by Lax *et al.*[4] Considering Gaussian-like beams, there exists some characteristic width ω_0 in the transverse dimension while the diffraction length l is characteristic of the longitudinal direction, and one has

$$l = K_0\omega_0^2, \quad \epsilon = \omega_0 l^{-1} \ll 1. \tag{42}$$

Let ξ, η, and ζ be the dimensionless coordinates

$$\xi = x\omega_0^{-1}, \quad \eta = y\omega_0^{-1}, \quad \zeta = zl^{-1}; \tag{43}$$

we denote by τ the couple (ξ,η) and by Δ^0, ∇_α^0, and $\hat{\nabla}_\alpha^0$, the dimensionless derivative operators while one has

$$F(\rho_1-\rho_2)\to F^0(\tau_1-\tau_2), \quad G(\rho_1-\rho_2)\to G^0(\tau_1-\tau_2), \tag{43'}$$
$$X(z;\rho_1,\rho_2)\to X^0(\zeta;\tau_1,\tau_2), \quad Y(z;\rho_1,\rho_2)\to Y^0(\zeta,\tau_1,\tau_2).$$

Then using (43) and (43'), Eqs. (41') and (40b) become

$$\left[2iK_0\sigma_3\left(\frac{1}{l}\frac{\partial}{\partial\zeta} + F^0\right) + \frac{1}{\omega_0^2}(\Delta_1^0 - \Delta_2^0)\right]X^0 = \frac{1}{\omega_0}\hat{\nabla}_\alpha^0\sigma^\alpha\left(\frac{1}{l}\frac{\partial}{\partial\zeta} + G^0\right)Y^0,$$

$$\left(\frac{1}{l}\frac{\partial}{\partial\zeta} + G^0 - 2iK_0\sigma_3\right)Y^0 = \frac{1}{\omega_0}\sigma^\alpha\nabla_\alpha^0 X^0,$$

where we used definition (34) of D_f and D_g. Taking (42) into account, we get

$$\left(2i\sigma_3\frac{\partial}{\partial\zeta} + 2il\sigma_3 F^0 + \Delta_1^0 - \Delta_2^0\right)X^0 = \epsilon\hat{\nabla}_\alpha^0\sigma^\alpha\left(\frac{\partial}{\partial\zeta} + lG^0\right)Y^0, \tag{44a}$$

$$\left(\epsilon\frac{\partial}{\partial\zeta} + \omega_0 T^0\right)Y^0 = \sigma^\alpha\nabla_\alpha^0 X^0, \quad T^0 = G^0 - 2iK_0\sigma_3. \tag{44b}$$

We first have to justify the way we write the right-hand side of (44a), which implies that $\epsilon|lG^0|$ is a quantity of order ϵ. Indeed, from definition (34) of the G function, one has

$$lG^0 = lK_0^2[A_n^0(0) - A_n^0(\tau_1-\tau_2)] \leqslant 2lK_0^2 A_n^2(0) = 4l\alpha_0,$$

where α_0 is the attenuation coefficient defined in Sec. II. But one has[6]

$$\alpha_0 \cong 0.392 C_n^2 K_0^2 L_0^{5/3},$$

where in strong turbulence the structure constant C_n^2 and the outer scale of turbulence L_0 are about 10^{-14}–10^{-15} $m^{-2/3}$ and 10 m, respectively. Taking $K_0 = 2\pi\times10^{-5}\ m^{-1}$, $\omega_0 = 10^{-2}$ m this leads to $\alpha_0 \simeq 10^{-2}$–$10^{-3}\ m^{-1}$ so that, for $l\simeq 60$ m, $\epsilon|lG^0|$ is actually of order ϵ.

Since $\epsilon \ll 1$, this suggests looking for the solution of Eqs. (44a) and (44b) by a perturbation technique, but we have to be cautious because, in (44b), ϵ is multiplying the highest derivative of z. Following a suggestion from Latta,[7] we set for Y^0

$$Y^0(\zeta;\tau_1,\tau_2) = U^0(\zeta;\tau_1,\tau_2) + \exp(-\omega_0 T^0\zeta/\epsilon)\times[Y^0(0;\tau_1,\tau_2) - U^0(0;\tau_1,\tau_2)] \tag{45a}$$

$$U^0(\zeta;\tau_1,\tau_2) = \sum_{n=0}^{\infty}\epsilon^n U_n^0(\zeta;\tau_1,\tau_2),$$

while, for X^0, one has

$$X^0(\zeta;\tau_1,\tau_2) = \sum_{n=0}^{\infty}\epsilon^n X_n^0(\zeta;\tau_1,\tau_2). \tag{45b}$$

Substituting (45) into (44) gives, to zeroth order,

$$\left(2i\sigma_3\frac{\partial}{\partial\zeta} + 2il\sigma_3 F^0 + \Delta_1^0 - \Delta_2^0\right)X_0^0 = 0, \tag{46a}$$

$$\omega_0 T^0 U_0^0 = \sigma^\alpha\nabla_\alpha^0 X_0^0 \tag{46b}$$

and, to any order $n>0$,

$$\left(2i\sigma_3\frac{\partial}{\partial\zeta} + 2il\sigma_3 F^0 + \Delta_1^0 - \Delta_2^0\right)X_n^0 = \hat{\nabla}_\alpha^0\sigma^\alpha\left(\frac{\partial}{\partial\zeta} + lG^0\right)Y_{n-1}^0, \tag{47a}$$

$$W_0 T^0 U_n^0 = \sigma^\alpha\nabla_\alpha^0 X_n^0 - \frac{\partial}{\partial\zeta}U_{n-1}^0. \tag{47b}$$

In this paper we do not discuss the convergence of the power series (45a), (45b), and we only consider the zeroth-order approximation.

B. Zeroth-order approximation to solution

Let us now come back to the Cartesian coordinates x, y, z: using definition (34) of the F function and definition (39) of the spinor X, we get from (46a)

$$\left(2iK_0\frac{\partial}{\partial z} + \Delta_1 - \Delta_2 + 2iK_0^3[A_n(0) - A_n(\rho_1-\rho_2)]\right)\times\Lambda_0(z;\rho_1,\rho_2) = 0, \tag{48a}$$

$$\left(2iK_0\frac{\partial}{\partial z} + \Delta_2 - \Delta_1 + 2iK_0^3[A_n(0) - A_n(\rho_1-\rho_2)]\right)\times\Xi_0(z;\rho_1,\rho_2) = 0. \tag{48b}$$

We see at once that (48a) is the exact equation supplied by the parabolic equation[6] for the scalar mutual coherence function Γ_{3p} [in agreement with (19), we denote this function by Γ_{3p} rather than Γ; the index p reminds us of its origin). As said in the Introduction, this result agrees with the fact proved by Lax *et al.*[4] that the parabolic equation is the zeroth order approximation to Maxwell's equations.

Remark: Eq. (48a) is often written with the correlation function $A_\epsilon(\rho_1-\rho_2)$, and, as already noted in Sec. II, one has $A_n(\rho) = 4A_\epsilon(\rho)$.

The solutions of (48a) are well known[6]; in terms of the coordinates

$$\rho = \tfrac{1}{2}(\rho_1+\rho_2), \quad r = \rho_1 - \rho_2, \tag{49}$$

which lead to

$$\frac{\partial}{\partial\rho_1} = \frac{1}{2}\frac{\partial}{\partial\rho} + \frac{\partial}{\partial r}, \quad \frac{\partial}{\partial\rho_2} = \frac{1}{2}\frac{\partial}{\partial\rho} - \frac{\partial}{\partial r}. \tag{49'}$$

One obtains for the functional $\Lambda_0(z;\rho,r)$:

$$\Lambda_0(z;\rho,r) = \int M_0(0,\chi,r-\chi z/K_0)\exp[i\chi\cdot\rho - H(z;\chi,r)]\,d\chi \tag{50a}$$

with

$$M_0(0;\chi,r) = \frac{1}{(2\pi)^2}\int \Lambda_0(0;\rho,r)e^{-i\chi\cdot\rho}\,d\rho \tag{50b}$$

and

$$H(z;\chi,r) = K_0^2\int_0^z [A_n(0) - A_n(r-\chi z'/K_0)]\,dz'. \tag{50c}$$

In these expressions, $\Lambda_0(0;\rho,r)$ is the boundary condition at $z=0$, χ is a two-dimensional vector dual to ρ (with respect to the Fourier transform), and $\chi\cdot\rho$ is the scalar product of both vectors χ and ρ.

Since one has $A_n(\rho_1-\rho_2) = A_n(\rho_2-\rho_1)$, one easily sees that we pass from (48a) to (48b) by interchanging the couple (ρ_1,ρ_2) into (ρ_2,ρ_1), that is (ρ,r) into $(\rho,-r)$; this gives, for $\Xi_0(z;\rho,r)$,

$$\Xi_0(z;\rho,r) = \int N_0(0;\chi,-r-Xz/K_0) \times\exp[i\chi\cdot\rho - H(z;\rho,-r)]\,d\chi, \tag{51a}$$

$$N_0(0;\chi,r) = \frac{1}{(2\pi)^2}\int \Xi_0(0;\rho,r)e^{-2i\chi\cdot\rho}\,d\rho. \tag{51b}$$

Using the variable $r' = r - \chi z/K_0$ makes it possible to obtain the expressions (50) and (51) in a more symmetric form.

With the coordinates z,ρ,r, Eq. (46b) gives

$$U_0(z;\rho,r) = T^{-1}\sigma^\alpha\nabla_\alpha X_0(z;\rho,r),$$

and from (45a) we get

$$Y_0(z;\rho,r) = T^{-1}\sigma^\alpha\nabla_\alpha X_0(z;\rho,r) + e^{-Tz}[Y(0;\rho,r) - U(0;\rho,r)]$$

which we write

$$Y_0(z;\rho,r) = T^{-1}\sigma^\alpha\nabla_\alpha X_0(z;\rho,r) + e^{-Tz}[Y_0(0;\rho,r) - T^{-1}\sigma^\alpha\nabla_\alpha X_0(0;\rho,r)]. \tag{52}$$

The functionals (50)–(52) are known as soon as the boundary conditions $\Lambda_0(0;\rho,r)$, $\Xi_0(0;\rho,r)$, and $Y_0(0;\rho,r)$ are given, which is discussed in the next section.

C. Boundary conditions

As for the coherent intensity in Sec. II, we must take into account that the propagation of the light beam is along the $0z$ axis. Using the coordinates z,ρ,r, the comparison of (2a) and (19) supplies the relations

$$\langle P_\alpha(z,\rho)\rangle = i\epsilon_\alpha^\beta\Gamma_\beta(z;\rho,0)\quad \alpha=1,2,$$
$$\langle P_3(z,\rho)\rangle = \Gamma_0(z;\rho,0),\quad \langle I(z,\rho)\rangle = \Gamma_3(z;\rho,0), \tag{53}$$

and the generalization of (14) to turbulent media leads to

$$\Gamma_\alpha(z;\rho,0) = 0,\quad \alpha=1,2, \tag{54a}$$

$$\Gamma_3(z;\rho,0) = |\Gamma_0(z;\rho,0)|, \tag{54b}$$

that is, in terms of the components ψ_1 and ψ_2 of the spinor ψ,

$$\langle\tilde\psi_1(z,\rho)\psi_2(z,\rho)\rangle = 0 = \langle\tilde\psi_2(z,\rho)\psi_1(z,\rho)\rangle,$$
$$\langle\tilde\psi_2(z,\rho)\psi_2(z,\rho)\rangle = 0 \tag{55a}$$

for $z=0$, these last relations are true with probability 1, which implies

$$\psi_2(0,\rho) = 0. \tag{55b}$$

Using (37), we get from (55a)

$$\Xi(z;\rho,0) = 0,\quad \Lambda(z;\rho,0) = 2\langle I(z,\rho)\rangle, \tag{56a}$$

$$\Omega(z;\rho,0) = 0,\quad \Pi(z;\rho,0) = 0, \tag{56b}$$

while (55b) gives

$$\Xi(0;\rho,r) = \Omega(0;\rho,r) = \Pi(0;\rho,r) = 0. \tag{57}$$

This supplies the boundary conditions for the three functionals Ξ, Ω, and Π. On the other hand, $\Lambda(0;\rho,r)$ is well known[6] from the physical characteristics of the light beam; in particular, $\Lambda(0,\rho,0) = 2I(0,\rho)$, where $I(0,\rho)$ is the intensity of the light source.

We now have to prove that if the conditions (56) are fulfilled for $z=0$ [in agreement with (57)], they are also fulfilled for any $z>0$. This follows from the fact that according to (40), one may write

$$X(z;\rho,r) = \int_0^z e^{F(z'-z)}\nabla_\alpha\sigma^\alpha Y(z';\rho,r)\,dz' + X(0;\rho,r),$$
$$Y(z;\rho,r) = \int_0^z e^{T(z'-z)}\sigma^\alpha\nabla_\alpha X(z';\rho,r)\,dz' + Y(0;\rho,r).$$

Let us now consider the boundary conditions for the zeroth-order approximation. In agreement with (56) and (57), we shall take

$$\Xi_0(0;\rho,r) = 0,\quad \Lambda_0(0;\rho,r) = \Gamma_{3p}(0;\rho,r), \tag{58a}$$

$$\Omega_0(0;\rho,r) = 0,\quad \Pi_0(0;\rho,r) = 0. \tag{58b}$$

These relations achieve the determination of the solutions (50)–(52). Let us then remark that according to (51) and (58a), one has

$$\Xi_0(z;\rho,r) = 0, \tag{59}$$

which implies from (37a) and (37b)

$$\Gamma_{3,0}(z;\rho,r) = \Gamma_{0,0}(z;\rho,r) = \Gamma_{3,p}(z;\rho,r); \tag{60}$$

that is, the mutual coherence function has the same expression to the zeroth-order approximation of the spinor formalism and to the paraxial approximation of the scalar formalism.

V. PLANE WAVE IN A STRONG TURBULENT MEDIUM

Since for a plane wave the parameter ϵ is useless, we have to come back to Eqs. (36), taking into account that now the 4-vector $\Gamma_\mu(z;\rho_1,\rho_2)$, $\mu=0,1,2,3$, is a function of z and $|r| = |\rho_1-\rho_2|$ only. Then using the coordinates (49) with $\rho=0$ together with the relation (49'), Eqs. (36) become

$$D_f\Gamma_3(z;0,r) + 2\partial_x\Gamma_1(z;0,r) + 2\partial_y\Gamma_2(z;0,r) = 0,\quad r=(x,y), \tag{61a}$$

$$D_f\Gamma_0(z;0,r) = 0, \tag{61b}$$

$$D_g\Gamma_1(z;0,r) - 2K_0\Gamma_2(z;0,r) - 2\partial_x\Gamma_3(z;0,r) = 0, \tag{61c}$$

$$D_g\Gamma_2(z;0,r) + 2K_0\Gamma_1(z;0,r) - 2\partial_y\Gamma_3(z;0,r) = 0. \tag{61d}$$

From (61b) and from definition (34) of the derivative operator D_f, we get

$$\Gamma_0(z;0,r) = \Gamma_0(0;0,r)\exp\{-K_0^2[A_n(0) - A_n(r)]z\}. \tag{62}$$

The three other equations (61) can be written:

$$D_f\Gamma_3 + (\partial_x - i\partial_y)(\Gamma_1 + i\Gamma_2) + (\partial_x + i\partial_y)(\Gamma_1 - i\Gamma_2) = 0, \tag{63a}$$

$$(D_g + 2iK_0)(\Gamma_1 + i\Gamma_2) - 2(\partial_x + i\partial_y)\Gamma_3 = 0, \tag{63b}$$

$$(D_g - 2iK_0)(\Gamma_1 - i\Gamma_2) - 2(\partial_x - i\partial_y)\Gamma_3 = 0. \tag{63c}$$

From (63b), (63c), using the fact that, as previously discussed, one has $\Gamma_1(0;0;r) = \Gamma_2(0;0,r) = 0$, we get

$$\Gamma_1(z;0,r) + i\Gamma_2(z;0,r) = 2\int_0^z e^{H(r)(z'-z)}(\partial_x + i\partial_y)\Gamma_3(z';0,r)\,dz', \tag{64a}$$

$$\Gamma_1(z;0,r) - i\Gamma_2(z;0,r) = 2\int_0^z e^{\tilde{H}(r)(z'-z)}(\partial_x - i\partial_y)\Gamma_3(z';0,r)\,dz' \tag{64b}$$

with $H(r) = G(r) + 2iK_0, \tilde{H}(r) = G(r) - 2iK_0$. Then a simple calculation gives

$$(\partial_x - i\partial_y)(\Gamma_1 + i\Gamma_2) = 2\int_0^z e^{H(r)(z'-z)}[\Delta_T\Gamma_3 + (z'-z)\nabla_T G\cdot\nabla_T\Gamma_3 + i(z'-z)\nabla_T G_\wedge \nabla_T\Gamma_3]\,dz', \tag{65a}$$

$$(\partial_x + i\partial_y)(\Gamma_1 - i\Gamma_2) = 2\int_0^z e^{\tilde{H}(r)(z'-z)}[\Delta_T\Gamma_3 + (z'-z)\nabla_T G\cdot\nabla_T\Gamma_3 - i(z'-z)\nabla_T G_\wedge \nabla_T\Gamma_3]\,dz' \tag{65b}$$

where Δ_T and ∇_T denote the Laplacian and the gradient transverse operators. Substituting (65a) and (65b) into (63a) gives

$$D_f\Gamma_3(z;0,r) + 4\int_0^z e^{G(r)(z'-z)}\{\cos[2K_0(z'-z)]\Delta_T\Gamma_3(z';0,r) + \cos[2K_0(z'-z)](z'-z)\nabla_T G(r)\cdot\nabla_T\Gamma_3(z';0,r) - \sin[2K_0(z'-z)](z'-z)\nabla_T G(r)\wedge\nabla_T\Gamma_3(z';0,r)\}\,dz' = 0. \tag{66}$$

This is the equation for the mutual coherence function of a plane wave in a strong turbulent medium.

Now in the Appendix we prove that in optics where K_0 is very large, one has for a bounded function $f(z)$ which varies smoothly on an interval π/K_0 approximately

$$\int_0^z \cos 2K_0(z-z')f(z')\,dz' \cong [f(0)/2K_0]\sin 2K_0 z, \tag{67a}$$

$$\int_0^z \sin 2K_0(z-z')f(z')\,dz' \cong [f(z)/2K_0] - [f(0)/2K_0]\cos 2K_0 z, \tag{67b}$$

$$\int_0^z (z-z')\cos 2K_0(z-z')f(z')\,dz' \cong [f(0)/2K_0^2](2K_0 z\sin 2K_0 z + \cos 2K_0 z \pm 1), \tag{67c}$$

$$\int_0^z (z-z')\sin 2K_0(z-z')f(z')\,dz' \cong [f(z)/2K_0^2] - [f(0)/2K_0^2] \times (2K_0 z\cos 2K_0 z - \sin 2K_0 z \pm 1). \tag{67d}$$

If we replace in these expressions $f(z')$ by $f(z-z')$, we obtain similar results; it is only sufficient to change $f(z)$ into $f(0)$ and vice versa.

Using these last results, the three terms to be integrated in Eq. (66) become

$$\int_0^z \cos 2K_0(z-z')e^{G(r)(z'-z)}\Delta_T\Gamma_3(z';0,r)\,dz' \cong (1/2K_0)e^{-G(r)z}\Delta_T\Gamma_3(0;0,r)\sin 2K_0 z,$$

$$\int_0^z (z'-z)\cos 2K_0(z-z')e^{G(r)(z'-z)}\nabla_T G(r)\cdot\nabla_T\Gamma_3(z';0,r)\,dz' \cong (1/2K_0^2)e^{-G(r)z}\nabla_T G(r)\cdot\nabla_T\Gamma_3(0;0,r) \times (2K_0 z\sin 2K_0 z + \cos 2K_0 z \pm 1),$$

$$\int_0^z (z'-z)\sin 2K_0(z-z')e^{G(r)(z'-z)}\nabla_T G(r)\wedge\nabla_T\Gamma_3(z';0,r)\,dz' \cong (1/2K_0^2)\nabla_T G(r)\wedge\nabla_T\Gamma_3(z;0,r) - (e^{-G(r)z}/2K_0^2)\nabla_T G(r)\wedge\nabla_T\Gamma_3(0;0,r) \times (2K_0 z\cos 2K_0 z - \sin 2K_0 z \pm 1).$$

For a wave perfectly coherent at $z = 0$, one has $\Gamma_3(0;0,r) = 1$, and these relations become

$$\int_0^z \cos 2K_0(z-z')e^{G(r)(z'-z)}\Delta_T\Gamma_3(z';0,r) \cong 0 \cong \int_0^z (z'-z)\cos 2K_0(z-z')e^{G(r)(z'-z)} \times \nabla_T G(r)\cdot\nabla_T\Gamma_3(z';0,r)\,dz', \tag{68a}$$

$$\int_0^z (z'-z)\sin 2K_0(z'-z)e^{G(r)(z'-z)}\nabla_T G(r)\wedge\nabla_T\Gamma_3(z';0,r)\,dz' \cong (1/2K_0^2)\nabla_T G(r)\wedge\nabla_T\Gamma_3(z;0,r). \tag{68b}$$

Substituting (68a) and (68b) into (66) and using definitions (34) of D_f, $F(r)$, and $G(r)$, we get

$$\left(\frac{\partial}{\partial z} + K_0^2[A_n(0) - A_n(r)] + 2\nabla_T A_n(r)\wedge\nabla_T\right)\Gamma_3(z;0,r) = 0. \tag{69}$$

If one has

$$|\nabla_T A_n(r)| \ll K_0|A_n(r)|, \tag{70a}$$

$$|\nabla_T\Gamma_3(z;0,r)| \ll K_0|\Gamma_3(z;0,r)|, \tag{70b}$$

the last term on the left-hand side of (69) can be neglected with respect to $K_0^2 A_n(r)\Gamma_3(z;0,r)$, and one obtains, in agreement with the solution supplied by the parabolic equation,[6]

$$\Gamma_3(z;0,r) = \exp\{-K_0^2[A_n(0) - A_n(r)]z\}. \tag{71}$$

To discuss (70a), let us think an instant of $A_n(r)$ as a Gaussian function $A_n(r) = (l_0/2\pi^{1/2})\exp(-r^2/l_0^2)$, where l_0 is the transverse correlation length, so that (70a) becomes

$$r \ll \tfrac{1}{2}K_0 l_0^2. \tag{72}$$

Although this particular form of (70a) is strictly valid only for a Gaussian function, it can be considered as a good approximation in many other cases provided that l_0 is a characteristic length of transverse turbulence. As a consequence, the expression (71) supplied by the parabolic equation is correct as long as (72) is fulfilled.

VI. CONCLUSIONS

From the previous results, we may state two conclusions:

(i) The spinor wave optics supplies more information about the statistical properties of a light beam propagating in a strong turbulent medium than the scalar wave optics since one may describe the fluctuations not only of intensity but also of the Poynting vector.

(ii) Since the spinor wave optics and the parabolic approximation to scalar optics lead to a first-order differential equation with respect to the leading variable z, they both correspond to a first-order Markov process, which differs for both theories but not in an essential way, especially when the refractive index satisfies Eq. (3b).

In fact we proved that when one is not too far from the light axis, that is, when $\epsilon \ll 1$ for a beam, and when (72) is fulfilled for a plane wave, both formalisms give similar results.

We shall see in the second paper on the intensity covariance function that these conclusions hold valid.

APPENDIX

We first remark that the following integral $\int_0^x \cos\theta\, d\theta$ (resp. $\int_0^x \sin\theta\, d\theta$) is zero on any subinterval $[m\pi,(m+1)\pi]$ (resp. $[(2m+1)\pi/2, (2m+3)\pi/2]$), $m = 0,1,2,\cdots$. Let $I_1 = \int_0^x \cos 2K_0(x-x')f(x')dx'$ and $I_2 = \int_0^x \sin 2K_0(x-x')f(x')\,dx'$.

In optics where K_0 is very large, one may consider that provided $f(x)$ does not oscillate with similar frequency, $f(x)$ is approximately constant on these intervals. So we may write for any bounded function $f(x)$, assuming $n\pi \leqslant 2K_0x < (n+1)\pi$,

$$\begin{aligned} I_1 &\cong (1/2K_0)\int_{n\pi}^{2K_0x} \cos\theta f(x-\theta/2K_0)\,d\theta \\ &\cong [f(0)/2K_0]\int_{n\pi}^{2K_0x}\cos\theta\,d\theta \\ &\cong [f(0)/2K_0]\sin 2K_0x \end{aligned} \tag{A1}$$

and, assuming $(2n_1+1)\pi/2 \leqslant 2K_0x < (2n_1+3)\pi/2$,

$$\begin{aligned} I_2 &\cong (1/2K_0)\int_0^{\pi/2}\sin\theta f(x-\theta/2K_0)\,d\theta \\ &\quad + (1/2K_0)\int_{(2n_1+1)\pi/2}^{2K_0x}\sin\theta f(x-\theta/2K_0)\,d\theta \\ &\cong [f(x)/2K_0]\int_0^{\pi/2}\sin\theta\,d\theta \\ &\quad + [f(0)/2K_0]\int_{(2n_1+1)\pi/2}^{2K_0x}\sin\theta\,d\theta \\ &\cong [f(x)/2K_0] - [f(0)/2K_0]\cos 2K_0x. \end{aligned} \tag{A2}$$

In the same way, one has

$$\begin{aligned} I_3 &= \int_0^x (x-x')\cos 2K_0(x-x')f(x')\,dx' \\ &\cong (1/2K_0^2)\int_{n\pi}^{2K_0x}\theta\cos\theta f(x-\theta/2K_0)\,d\theta \\ &\cong [f(0)/2K_0^2]\left[(\theta\sin\theta)_{n\pi}^{2K_0x} - \int_{n\pi}^{2K_0x}\sin\theta\,d\theta\right] \\ &\cong [f(0)/2K_0^2] \\ &\quad \times [2K_0x\sin 2K_0x + \cos 2K_0x + (-1)^{n+1}] \end{aligned} \tag{A3}$$

and

$$\begin{aligned} I_4 &= \int_0^x (x-x')\sin 2K_0(x-x')f(x')\,dx' \\ &\cong (1/2K_0^2)\int_0^{\pi/2}\theta\sin\theta f(x-\theta/2K_0)\,d\theta \\ &\quad + (1/2K_0^2)\int_{(2n_1+1)\pi/2}^{2K_0x}\theta\sin\theta f(x-\theta/2K_0)\,d\theta \\ &\cong [f(x)/2K_0^2]\int_0^{\pi/2}\theta\sin\theta\,d\theta + [f(0)/2K_0^2] \\ &\quad \times \int_{(2n_1+1)\pi/2}^{2K_0x}\theta\sin\theta\,d\theta \\ &\cong [f(x)/2K_0^2] - [f(0)/2K_0^2] \\ &\quad \times [2K_0x\cos 2K_0x - \sin 2K_0x + (-1)^{n_1}]. \end{aligned} \tag{A4}$$

Remark: If one considers instead of I_1 and I_2 the following integrals,

$$I_1' = \int_0^x \cos 2K_0(x-x')f(x-x')\,dx',$$

$$I_2' = \int_0^x \sin 2K_0(x-x')f(x-x')\,dx'$$

and similarly I_3' and I_4', then the previous results hold valid; it is only sufficient to change $f(x)$ into $f(0)$ and vice versa.

[1] P. Hillion, J. Math. Phys. **19**, 264 (1978).
[2] P. Hillion, J. Opt. (Paris) **10**, 21 (1979).
[3] J. Strohbehn, in *Laser Beam Propagation in the Atmosphere*, edited by J. W. Strohbehn (Springer–Verlag, Berlin, 1978).
[4] M. Lax, W. H. Louisell, and W. B. McKnight, Phys. Rev. A **11**, 1345 (1975).
[5] I. M. Gel'fand and S. V. Fomin, *Calculus of Variations* (Prentice-Hall, Englewood Cliff, N.J., 1963).
[6] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978), Vol. 2.
[7] M. Van Dyke, *Perturbation Methods in Fluid Mechanics* (Academic, New York, (1964); A. H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973).

Strong fluctuations in the spinor wave optics. Intensity covariance function

P. Hillion
Institut Henri Poincaré, 75231 Paris, France

S. Quinnez
Faculté des Sciences Mirande, 21000 Dijon, France

(Received 3 September 1981; accepted for publication 18 December 1981)

In this paper, which is a continuation of the preceding paper, we consider in the framework of the spinor wave optics the intensity covariance function of a light field propagating in a strong turbulent medium. In the preceding paper, for the second-order moments we had a 4-vector corresponding to intensity and to the Poynting vector; similarly, here we obtain a second-rank tensor with 16 components. But we limit ourselves to the zeroth-order approximation with respect to the parameter $\epsilon = \omega_0 l^{-1}$, where ω_0 is a transverse characteristic of the light beam and l the diffraction length. So only four components are needed, and provided that a particular assumption is made on the statistical behavior of the light beam, we get for the intensity covariance function the usual result obtained for a plane wave in the framework of the parabolic approximation. The interpretation of this result is discussed.

PACS numbers: 42.20. – y

I. INTRODUCTION

In Ref. 1 we considered the propagation of a light beam in a turbulent medium in the framework of the spinor wave optics, and we computed an approximation of the average intensity. Here we consider the intensity covariance function, and we first sum up the main notations introduced.

Let us consider a beam propagating along the $0z$ axis; then ρ denotes the transverse coordinates (x, y), $\psi(z, \rho)$ is a two-component spinor which describes the amplitude of the light field, $\psi^+(z, \rho)$ is the Hermitian conjugate. Then using the Pauli matrices $\sigma_j, j = 1, 2, 3$, and the 2×2 identity matrix σ_0, we introduced a mutual coherence 4-vector:

$$\Gamma_\mu(z;\rho_1,\rho_3) = \langle Z_\mu(z;\rho_1,\rho_3)\rangle = \langle \psi^+(z,\rho_3)\sigma_3\sigma_\mu\psi(z,\rho_1)\rangle, \quad \mu = 0,1,2,3, \tag{1}$$

where the angle brackets $\langle\ \rangle$ denote an ensemble average. Then the average intensity $\langle I(z,\rho_1)\rangle$ is

$$\langle I(z,\rho_1)\rangle = \lim_{\rho_3\to\rho_1} \Gamma_3(z;\rho_1,\rho_3) = \langle \psi^+(z,\rho_1)\psi(z,\rho_1)\rangle. \tag{2}$$

In a similar way we consider here a tensor of rank 2 with 16 components:

$$\begin{aligned}\Gamma_{\mu\nu}(z;\rho_1,\rho_3;\rho_2,\rho_4) &= \langle Z_{\mu\nu}(z;\rho_1,\rho_3;\rho_2,\rho_4)\rangle \\ &= \langle Z_\mu(z;\rho_1,\rho_3)Z_\nu(z;\rho_2,\rho_4)\rangle, \\ &\quad \mu,\nu = 0,1,2,3,\end{aligned} \tag{3}$$

with

$$Z_\mu(z;\rho_1,\rho_3) = \psi^+(z,\rho_3)\sigma_3\sigma_\mu\psi(z,\rho_1), \quad \mu = 0,1,2,3; \tag{3'}$$

then the fourth-order coherence function is

$$\Gamma_{33}(z;\rho_1,\rho_3;\rho_2,\rho_4) = \langle \psi^+(z,\rho_3)\psi(z,\rho_1)\psi^+(z,\rho_4)\psi(z,\rho_2)\rangle \tag{4}$$

so that the intensity correlation function is defined by the relation

$$\langle I(z,\rho_1)I(z,\rho_2)\rangle = \lim_{\substack{\rho_4\to\rho_2\\ \rho_3\to\rho_1}} \Gamma_{33}(z;\rho_1,\rho_3;\rho_2,\rho_4). \tag{4'}$$

From (2) and (4') we obtain the intensity covariance function

$$\begin{aligned}B_I(z;\rho_1,\rho_2) &= \langle I(z,\rho_1)I(z,\rho_2)\rangle - \langle I(z,\rho_1)\rangle\langle I(z,\rho_2)\rangle \\ &= \Gamma_{33}(z;\rho_1,\rho_1;\rho_2,\rho_2) - \Gamma_3(z;\rho_1,\rho_1)\Gamma_3(z;\rho_2,\rho_2).\end{aligned} \tag{5}$$

In this paper we limit ourselves in two aspects:

(i) First we do not consider all the components $\Gamma_{\mu\nu}$ but only those necessary to compute $B_I(z;\rho_1,\rho_2)$.

(ii) As in Ref. 1, we only calculate the zeroth-order approximation with respect to the parameter $\epsilon = \omega_0 l^{-1}$, where ω_0 is a transverse characteristic of the light beam and l the diffraction length.

Then we shall get the same formal intensity correlation function as the parabolic approximation for a plane wave, but we disagree with the interpretation plane wave of this result.

II. BASIC EQUATIONS

A. General

We do not write the arguments z, ρ_1, ρ_3 when no ambiguity is to be feared. The refractive index has the form $n = 1 + \mu$ and we denote by n_1, n_3, μ_1, and μ_3 the functions $n(z,\rho_1)$, $n(z,\rho_3)$, $\mu(z,\rho_1)$, and $\mu(z,\rho_3)$. The Greek indices α, β take the values 1, 2 corresponding to the transverse coordinates $\rho(x_1 = x, x_2 = y)$, ∂_α is the derivative $\partial/\partial x_\alpha$, and $\epsilon_{\alpha\beta}$ the antisymmetric tensor $\epsilon_{11} = \epsilon_{22} = 0, \epsilon_{12} = -\epsilon_{21} = 1$. In Ref. 1 we proved that the 4-vector (3') $Z_\mu, \mu = 0, 1, 2, 3$ is a solution of the following equations:

$$\frac{\partial}{\partial z}Z_3 + (\partial_\alpha^1 - \partial_\alpha^3)Z^\alpha - iK_0(\mu_1 - \mu_3)Z_0 = 0, \tag{6a}$$

$$\frac{\partial}{\partial z}Z_0 + i\epsilon^{\alpha\beta}(\partial_\alpha^1 + \partial_\alpha^3)Z_\beta - iK_0(\mu_1 - \mu_3)Z_3 = 0, \tag{6b}$$

$$\frac{\partial}{\partial z}Z_\alpha - (\partial_\alpha^1 - \partial_\alpha^3)Z_3 - i\epsilon_\alpha^\beta(\partial_\beta^1 + \partial_\beta^3)Z_0 - K_0(n_1 + n_3)\epsilon_\alpha^\beta Z_\beta = 0, \quad \alpha = 1, 2; \tag{6c}$$

 0022-2488/83/030721-07$02.50

in these expressions we use the summation convention $\partial_\alpha Z^\alpha = \partial_1 Z_1 + \partial_2 Z_2$. The explicit form of Eqs. (6a) and (6b) is

$$\frac{\partial}{\partial z} Z_3 + (\partial_x^1 - \partial_x^3)Z_1 + (\partial_y^1 - \partial_y^3)Z_2 - iK_0(\mu_1 - \mu_3)Z_0 = 0,$$

$$\frac{\partial}{\partial z} Z_0 + i(\partial_x^1 + \partial_x^3)Z_2 - i(\partial_y^1 + \partial_y^3)Z_1 - iK_0(\mu_1 - \mu_3)Z_3 = 0.$$

Summing and subtracting these two equations, we get

$$\left(\frac{\partial}{\partial z} - iK_0(\mu_1 - \mu_3)\right)(Z_3 + Z_0) + (\partial_x^1 - i\partial_y^1)(Z_1 + iZ_2) - (\partial_x^3 + i\partial_y^3)(Z_1 - iZ_2) = 0, \tag{7a}$$

$$\left(\frac{\partial}{\partial z} + iK_0(\mu_1 - \mu_3)\right)(Z_3 - Z_0) - (\partial_x^3 - i\partial_y^3)(Z_1 + iZ_2) + (\partial_x^1 + i\partial_y^1)(Z_1 - iZ_2) = 0.$$

In the same way the explicit form of Eqs. (6c) is

$$\frac{\partial}{\partial z} Z_1 - (\partial_x^1 - \partial_x^3)Z_3 - i(\partial_y^1 + \partial_y^3)Z_0 - K_0(n_1 + n_3)Z_2 = 0,$$

$$\frac{\partial}{\partial z} Z_2 - (\partial_y^1 - \partial_y^3)Z_3 + i(\partial_x^1 + \partial_x^3)Z_0 + K_0(n_1 + n_3)Z_1 = 0.$$

Summing and subtracting these two equations, the second of which being multiplied by i, we get

$$\left(\frac{\partial}{\partial z} + iK_0(n_1 + n_3)\right)(Z_1 + iZ_2) - (\partial_x^1 + i\partial_y^1)(Z_3 + Z_0) + (\partial_x^3 + i\partial_y^3)(Z_3 - Z_0) = 0, \tag{7b}$$

$$\left(\frac{\partial}{\partial z} - iK_0(n_1 + n_3)\right)(Z_1 - iZ_2) + (\partial_x^3 - i\partial_y^3)(Z_3 + Z_0) - (\partial_x^1 - i\partial_y^1)(Z_3 - Z_0) = 0.$$

Let us now introduce the following matrices:

$$\nabla_\alpha \sigma^\alpha = (\partial^1_\alpha - \partial^3_\alpha \sigma_1)\sigma^\alpha, \tag{8a}$$

$$\sigma^\alpha \nabla_\alpha = \sigma^\alpha(\partial^1_\alpha - \partial^3_\alpha \sigma_1), \tag{8b}$$

and the two component spinors

$$\mathcal{U} = \begin{pmatrix} Z_3 + Z_0 \\ Z_3 - Z_0 \end{pmatrix}, \quad \mathcal{V} = \begin{pmatrix} Z_1 - iZ_2 \\ Z_1 + iZ_2 \end{pmatrix}. \tag{9}$$

Then Eqs. (7a) and (7b) become

$$\left(\frac{\partial}{\partial z} - iK_0(\mu_1 - \mu_3)\sigma_3\right)\mathcal{U} + \nabla_\alpha \sigma^\alpha \mathcal{V} = 0,$$

$$\left(\frac{\partial}{\partial z} - iK_0(n_1 + n_3)\sigma_3\right)\mathcal{V} - \sigma^\alpha \nabla_\alpha \mathcal{U} = 0. \tag{10}$$

These basic equations will be used to compute the intensity covariance function.

B. Relations between $\Gamma_{\mu\nu}$ and both spinors $\mathcal{U}$ and $\mathcal{V}$

We now have to define $\Gamma_{\mu\nu}$ in terms of both spinors $\mathcal{U}$ and $\mathcal{V}$. Let us consider the transformation

$$Z_\mu(z;\rho_1,\rho_3) \to Z^*_\mu(z;\rho_1,\rho_3) = (-1)^{\delta_{\mu;1,2}} Z^+_\mu(z;\rho_3,\rho_1), \quad \mu = 0,1,2,3, \tag{11}$$

obtained by associating the Hermitian conjugation with the interchanging of the transverse coordinates ρ_1 and ρ_3; $\delta_{\mu;1,2}$ is the Kronecker symbol, zero for $\mu \neq 1,2$ and equal to unity for $\mu = 1,2$. Then from definition (3') of $Z_\mu(z;\rho_1,\rho_3)$ and from the properties of the Pauli matrices, we get

$$Z^*_\mu(z;\rho_1,\rho_3) = Z_\mu(z;\rho_1,\rho_3), \quad \mu = 0,1,2,3, \tag{11'}$$

and from definition (3) of $Z_{\mu\nu}$ it becomes

$$Z_{\mu\nu}(z;\rho_1,\rho_3,\rho_2,\rho_4) = Z_\mu(z;\rho_1,\rho_3)Z^*_\nu(z;\rho_2,\rho_4), \quad \mu,\nu = 0,1,2,3. \tag{12a}$$

If we now introduce the two spinors $\chi^* = (Z_1^*, Z_2^*, Z_3^*, Z_0^*)$, $\chi^T = (Z_1, Z_2, Z_3, Z_0)$, where T denotes the transposition, relation (12a) can be written:

$$Z_{\mu\nu} = \chi^* \gamma_{\mu\nu} \chi, \quad \mu,\nu = 0,1,2,3, \tag{12b}$$

where $\gamma_{\mu\nu}$ is a 4×4 matrix with only one entry not zero (and equal to 1). But one has

$$\chi = S\psi, \quad \chi^* = \psi^* S^+, \quad \psi^T = (\mathcal{U}^T, \mathcal{V}^T),$$

$$\psi^* = (\mathcal{U}^*, \mathcal{V}^*), \quad S = \tfrac{1}{2}\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & i & -i \\ 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix}, \tag{13}$$

so that substituting (13) into (12b) with the remark that S is a unitary matrix, we get

$$Z_{\mu\nu} = \psi^* S^{-1} \gamma_{\mu\nu} S \psi, \quad \mu,\nu = 0,1,2,3. \tag{12c}$$

Using four 2×2 matrixes $A^{\mu\nu}_{\alpha\beta}$, $\alpha,\beta = 1,2$, we may write

$$S^{-1}\gamma_{\mu\nu}S = \begin{bmatrix} A^{\mu\nu}_{11} & A^{\mu\nu}_{12} \\ A^{\mu\nu}_{21} & A^{\mu\nu}_{22} \end{bmatrix}, \quad \mu,\nu = 0,1,2,3,$$

and, since the Pauli matrices σ_μ, $\mu = 0,1,2,3$, are a basis of the 2×2 matrix algebra, one has

$$A^{\mu\nu}_{\alpha\beta} = \sum_{\lambda=0}^{4} C_\lambda (A^{\mu\nu}_{\alpha\beta})\sigma^\lambda,$$

where the constants C_λ depend on the matrices $A^{\mu\nu}_{\alpha\beta}$. As a consequence, there exists a linear relationship between the components of the tensor $Z_{\mu\nu}$ and the components of the four following vectors:

$$\mathcal{U}^*\sigma_\mu \mathcal{U}, \quad \mathcal{U}^*\sigma_\mu \mathcal{V}, \quad \mathcal{V}^*\sigma_\mu \mathcal{U}, \quad \mathcal{V}^*\sigma_\mu \mathcal{V}, \quad \mu = 0,1,2,3; \tag{14}$$

for instance, one has

$$Z_{33} = \tfrac{1}{4}\mathcal{U}^*(\sigma_0 + \sigma_1)\mathcal{U}. \tag{14'}$$

So, to obtain $\Gamma_{\mu\nu}$, we just have to look for the average of the four vectors (14) taking into account Eqs. (10).

This last result is interesting for computing all the components $\Gamma_{\mu\nu}$, but, as stated in the Introduction, we only consider the components entering into the formula of the intensity covariance function. Moreover, as we shall now see, to zeroth-order approximation with respect to ϵ, one has further simplifications so that we shall no longer use (14).

C. Approximate theory

The problem of computing $\Gamma_{\mu\nu}$ without any approximation would lead to a formidable task, so we make three simplifications which considerably reduce the work.

1. First, as in Ref. 1, we introduce the dimensionless coordinates

$$\xi = x\omega_0^{-1}, \quad \eta = y\omega_0^{-1}, \quad \zeta = z\,l^{-1}, \tag{15}$$

where ω_0 is some characteristic width in the transverse dimension while l is a diffraction length in the longitudinal direction with

$$l = K_0\omega_0^2, \quad \epsilon = \omega_0\, l^{-1} \ll 1. \tag{16}$$

We use the following notations:

$$\mu(\rho)\to\mu^0(\tau), \quad n(\rho)\to n^0(\tau), \quad \tau = (\xi, \eta), \tag{17}$$

$$\mathscr{U}(z;\rho_1,\rho_3)\to\mathscr{U}^0(\zeta;\tau_1,\tau_3), \quad \mathscr{V}(z;\rho_1,\rho_3)\to\mathscr{V}^0(\zeta;\tau_1,\tau_3).$$

Here also we do not write the arguments ζ, τ_1, τ_3 when there is no ambiguity. Let us now come back to Eqs. (10) which we transform in the following way. Multiplying the first one by $2iK_0\sigma_3$ and using the relation

$$\sigma_3\nabla_\alpha\sigma^\alpha = -\hat{\nabla}_\alpha\sigma^\alpha\sigma_3, \quad \hat{\nabla}_\alpha = \partial_\alpha^1 + \partial_\alpha^3\sigma_1, \quad \alpha = 1, 2, \tag{18a}$$

we get

$$\left(2iK_0\sigma_3\frac{\partial}{\partial z} + 2K_0^2(\mu_1 - \mu_3)\right)\mathscr{U} = \hat{\nabla}_\alpha\sigma^\alpha(2iK_0\sigma_3\mathscr{V}), \tag{19a}$$

and the second Eq. (10) is

$$\left(\frac{\partial}{\partial z} - iK_0(\mu_1 + \mu_3)\sigma_3 - 2iK_0\sigma_3\right)\mathscr{V} - \sigma^\alpha\nabla_\alpha\mathscr{U} = 0. \tag{19b}$$

Taking $2iK_0\sigma_3\mathscr{V}$ out of (19b) and substituting it into (19a), we get

$$\left(2iK_0\sigma_3\frac{\partial}{\partial z} + 2K_0^2(\mu_1 - \mu_3)\right)\mathscr{U} = \hat{\nabla}_\alpha\sigma^\alpha\left[\left(\frac{\partial}{\partial z} - iK_0(\mu_1 + \mu_3)\sigma_3\right)\mathscr{V} - \sigma^\beta\nabla_\beta\mathscr{U}\right].$$

Now using the following relation proved in Ref. 1,

$$\hat{\nabla}_\alpha\sigma^\alpha\sigma^\beta\nabla_\beta = \Delta_1 - \Delta_3, \tag{18b}$$

where Δ denotes the transverse Laplacian operator, we get

$$\left(2iK_0\sigma_3\frac{\partial}{\partial z} + 2K_0^2(\mu_1 - \mu_3) + \Delta_1 - \Delta_3\right)\mathscr{U} = \hat{\nabla}_\alpha\sigma^\alpha\left(\frac{\partial}{\partial z} - iK_0(\mu_1 + \mu_3)\sigma_3\right)\mathscr{V}; \tag{20a}$$

in addition to this equation, we must consider the second Eq. (10)

$$\left(\frac{\partial}{\partial z} - iK_0(n_1 + n_3)\sigma_3\right)\mathscr{V} - \sigma^\alpha\nabla_\alpha\mathscr{U} = 0. \tag{20b}$$

We now use the dimensionless coordinates (15) and denote by Δ^0, ∇_α^0, $\hat{\nabla}_\alpha^0$ the derivative operators defined with these last variables. Then Eqs. (20) become

$$\left(2i\sigma_3\frac{\partial}{\partial\zeta} + 2K_0^2\,\omega_0^2(\mu_1^0 - \mu_3^0) + \Delta_1^0 - \Delta_3^0\right)\mathscr{U}^0 = \epsilon\hat{\nabla}_\alpha^0\sigma^\alpha\left(\frac{\partial}{\partial\zeta} - iK_0 l(\mu_1^0 + \mu_3^0) + \sigma_3\right)\mathscr{V}^0, \tag{21a}$$

$$\left(\epsilon\frac{\partial}{\partial\zeta} - iK_0\omega_0(n_1^0 + n_3^0)\sigma_3\right)\mathscr{V}^0 - \sigma^\alpha\nabla_\alpha^0\mathscr{U}^0 = 0. \tag{21b}$$

As in Ref. 1 we look for the solution of these equations by a perturbation technique with

$$\mathscr{U}^0(\zeta;\tau_1,\tau_3) = \sum_{n=0}^{\infty}\epsilon^n\mathscr{U}_n^0(\zeta;\tau_1,\tau_3), \tag{22a}$$

$$\mathscr{V}^0(\zeta;\tau_1,\tau_3) = \mathscr{W}^0(\zeta;\tau_1,\tau_3) + \exp[(iK_0\omega_0/\epsilon)(n_1^0 + n_3^0)\sigma_3\zeta]\times[\mathscr{V}^0(0;\tau_1,\tau_3) - \mathscr{W}^0(0;\tau_1,\tau_3)], \tag{22b}$$

$$\mathscr{W}^0(\zeta;\tau_1,\tau_3) = \sum_{n=0}^{\infty}\epsilon^n\mathscr{W}_n^0(\zeta;\tau_1,\tau_3); \tag{22c}$$

the choice of the particular form (22b) is discussed in Ref. 1.

As stated in the Introduction, we only consider here the zeroth-order approximation $\mathscr{U}_0^0$, $\mathscr{V}_0^0$, and, substituting (22) into (21), we get

$$\left(2i\sigma_3\frac{\partial}{\partial\zeta} + 2K_0^2\omega_0^2(\mu_1^0 - \mu_3^0) + \Delta_1^0 - \Delta_3^0\right)\mathscr{U}_0^0 = 0, \tag{23a}$$

$$iK_0\omega_0(n_1^0 + n_3^0)\sigma_3\mathscr{W}_0^0 + \sigma^\alpha\nabla_\alpha^0\mathscr{U}_0^0 = 0. \tag{23b}$$

We shall no longer write the lower index since it must be understood that we always speak about the zeroth-order approximation.

2. The second simplification comes from the following remark: Since we only intend to compute the intensity covariant function (5), we just need to know according to (4) the solution $\mathscr{U}^0$ to Eq. (23a). Now using definition (9) of $\mathscr{U}$ and coming back to the coordinates x, y, z, we can explicitly write Eq. (23a) as

$$\left(2iK_0\frac{\partial}{\partial z} + 2K_0^2(\mu_1 - \mu_3) + \Delta_1 - \Delta_3\right)\times(Z_3(z;\rho_1,\rho_3) + Z_0(z;\rho_1,\rho_3)) = 0, \tag{24a}$$

$$\left(-2iK_0\frac{\partial}{\partial z} + 2K_0^2(\mu_1 - \mu_3) + \Delta_1 - \Delta_3\right)\times(Z_3(z;\rho_1,\rho_3) - Z_0(z;\rho_1,\rho_3)) = 0. \tag{24b}$$

These equations are the basic equations we shall use later.

Of course, we would like to deal only with Γ_{33}:

$$\Gamma_{33}(z;\rho_1,\rho_3;\rho_2,\rho_4) = \langle Z_{33}(z;\rho_1,\rho_3;\rho_2,\rho_4)\rangle = \langle Z_3(z;\rho_1,\rho_3)Z_3(z;\rho_2,\rho_4)\rangle, \tag{25}$$

but it is clear from Eqs. (24) that we also have to take into account Γ_{00}, Γ_{30}, and Γ_{03}, these functionals having a definition similar to (25).

3. Finally let $R = \frac{1}{4}(\rho_1 + \rho_2 + \rho_3 + \rho_4)$ be the center of mass on the four points ρ_1, ρ_2, ρ_3, and ρ_4; we assume that turbulence is stationary in the transverse direction and that one has

$$\nabla_R\Gamma_{ij} = 0, \quad i = 0, 3, \quad j = 0, 3, \tag{26}$$

where ∇_R denotes the gradient with respect to the variable R. We shall later discuss this condition.

III. THE INTENSITY COVARIANCE FUNCTION TO ZEROTH-ORDER APPROXIMATION

A. Equations for the quadruplet Γ_{33}, Γ_{00}, Γ_{03}, Γ_{30}

To simplify calculations, we introduce the operator

$$L = 2K_0^2\mu + \Delta \tag{27a}$$

and the following notation:

$$Z(1,3)=Z(z;\rho_1,\rho_3)\quad Z(2,4)=Z(z;\rho_2,\rho_4)$$
$$X=Z_3+Z_0,\quad Y=Z_3-Z_0. \tag{27b}$$

Then Eqs. (24) give

$$\left(2iK_0\frac{\partial}{\partial z}+L_1-L_3\right)X(1,3)=0, \tag{28a}$$
$$\left(2iK_0\frac{\partial}{\partial z}+L_2-L_4\right)X(2,4)=0, \tag{28b}$$
$$\left(-2iK_0\frac{\partial}{\partial z}+L_1-L_3\right)Y(1,3)=0, \tag{29a}$$
$$\left(-2iK_0\frac{\partial}{\partial z}+L_2-L_4\right)Y(2,4)=0, \tag{29b}$$

and, by summing two by two these equations after multiplying them on the left by the ad hoc function, we get

$$\begin{aligned}
&2iK_0\frac{\partial}{\partial z}(X(1,3)X(2,4))+X(2,4)(L_1-L_3)X(1,3)\\
&\quad+X(1,3)(L_2-L_4)X(2,4)=0,\\
&2iK_0\frac{\partial}{\partial z}(Y(1,3)Y(2,4))-Y(2,4)(L_1-L_3)Y(1,3)\\
&\quad-Y(1,3)(L_2-L_4)Y(2,4)=0,\\
&2iK_0\frac{\partial}{\partial z}(X(1,3)Y(2,4))+Y(2,4)(L_1-L_3)X(1,3)\\
&\quad-X(1,3)(L_2-L_4)Y(2,4)=0,\\
&2iK_0\frac{\partial}{\partial z}(Y(1,3)X(2,4))-X(2,4)(L_1-L_3)Y(1,3)\\
&\quad+Y(1,3)(L_2-L_4)X(2,4)=0.
\end{aligned} \tag{30}$$

Using some obvious linear transformations and coming back to Z_0, Z_3, we get

$$\begin{aligned}
&2iK_0\frac{\partial}{\partial z}(Z_3(1,3)Z_3(2,4))+Z_3(2,4)(L_1-L_3)Z_0(1,3)\\
&\quad+Z_3(1,3)(L_2-L_4)Z_0(2,4)=0,\\
&2iK_0\frac{\partial}{\partial z}(Z_0(1,3)Z_3(2,4))+Z_3(2,4)(L_1-L_3)Z_3(1,3)\\
&\quad+Z_0(1,3)(L_2-L_4)Z_0(2,4)=0,\\
&2iK_0\frac{\partial}{\partial z}(Z_3(1,3)Z_0(2,4))+Z_0(2,4)(L_1-L_3)Z_0(1,3)\\
&\quad+Z_3(1,3)(L_2-L_4)Z_3(2,4)=0,\\
&2iK_0\frac{\partial}{\partial z}(Z_0(1,3)Z_0(2,4))+Z_0(2,4)(L_1-L_3)Z_3(1,3)\\
&\quad+Z_0(1,3)(L_2-L_4)Z_3(2,4)=0.
\end{aligned}$$

Using (27) and taking the ensemble average, these equations become

$$\begin{aligned}
&2iK_0\frac{\partial}{\partial z}\Gamma_{33}+(\Delta_1-\Delta_3)\Gamma_{03}+(\Delta_2-\Delta_4)\Gamma_{30}\\
&\quad+2K_0^2\langle(\mu_1-\mu_3)Z_3(2,4)Z_0(1,3)\\
&\quad+(\mu_2-\mu_4)Z_3(1,3)Z_0(2,4)\rangle=0,
\end{aligned} \tag{31a}$$

$$\begin{aligned}
&2iK_0\frac{\partial}{\partial z}\Gamma_{03}+(\Delta_1-\Delta_3)\Gamma_{33}+(\Delta_2-\Delta_4)\Gamma_{00}\\
&\quad+2K_0^2\langle(\mu_1-\mu_3)Z_3(2,4)Z_3(1,3)\\
&\quad+(\mu_2-\mu_4)Z_0(1,3)Z_0(2,4)\rangle=0,
\end{aligned} \tag{31b}$$

$$\begin{aligned}
&2iK_0\frac{\partial}{\partial z}\Gamma_{30}+(\Delta_1-\Delta_3)\Gamma_{00}+(\Delta_2-\Delta_4)\Gamma_{33}\\
&\quad+2K_0^2\langle(\mu_1-\mu_3)Z_0(2,4)Z_0(1,3)\\
&\quad+(\mu_2-\mu_4)Z_3(1,3)Z_3(2,4)\rangle=0,
\end{aligned} \tag{31c}$$

$$\begin{aligned}
&2iK_0\frac{\partial}{\partial z}\Gamma_{00}+(\Delta_1-\Delta_3)\Gamma_{30}+(\Delta_2-\Delta_4)\Gamma_{03}\\
&\quad+2K_0^2\langle(\mu_1-\mu_3)Z_0(2,4)Z_3(1,3)\\
&\quad+(\mu_2-\mu_4)Z_0(1,3)Z_3(2,4)\rangle=0.
\end{aligned} \tag{31d}$$

We now have to compute the last term on the left-hand side of each equation (31), and this requires some results previously proved in Ref. 1. First we assume that $\mu(\bar{x})$ is a random Gaussian field with the correlation function

$$B_n(\bar{x},\bar{x}')=\langle\mu(\bar{x})\mu(\bar{x}')\rangle=A_n(\rho-\rho')\delta(z-z'),\quad \bar{x}=(\rho,z), \tag{32}$$

where $\delta(z-z')$ is the Dirac distribution while $A_n(\rho-\rho')$ satisfies

$$A_n(\rho-\rho')=A_n(\rho'-\rho). \tag{32'}$$

Then one may apply the Novikov–Furutsu formula and $F(z,\rho)$ being a differentiable function, one has

$$\langle\mu(z,\rho)F(z,\rho)\rangle=\int d\rho'\,A_n(\rho-\rho')\left\langle\frac{\delta F(z,\rho)}{\delta\mu(z,\rho')}\right\rangle, \tag{33}$$

where $\delta F/\delta\mu$ is a variational derivative. This leads to the following results[1]

$$\begin{aligned}
\frac{\delta Z_0(z;\rho_1,\rho_3)}{\delta\mu(z,\rho_1')}=&-\tfrac{1}{2}iK_0\,\delta(\rho_3-\rho_1')Z_3(z;\rho_1,\rho_1')\\
&+\tfrac{1}{2}iK_0\,\delta(\rho_1-\rho_1')Z_3(z;\rho_1',\rho_3),
\end{aligned} \tag{34a}$$

$$\begin{aligned}
\frac{\delta Z_3(z,\rho_1,\rho_3)}{\delta\mu(z,\rho_1')}=&-\tfrac{1}{2}iK_0\,\delta(\rho_3-\rho_1')Z_0(z;\rho_1,\rho_1')\\
&+\tfrac{1}{2}iK_0\,\delta(\rho_1-\rho_1')Z_0(z;\rho_1',\rho_3).
\end{aligned} \tag{34b}$$

Let us then consider the term $\langle\mu_1,Z_3(2,4)Z_0(1,3)\rangle$ on the left-hand side of (31a); one has

$$\langle\mu_1,Z_3(2,4)Z_0(1,3)\rangle=\int d\rho_1'\,A_n(\rho_1-\rho_1')\left\langle Z_3(z;\rho_2,\rho_4)\frac{\delta Z_0(z;\rho_1,\rho_3)}{\delta\mu(z,\rho_1')}+\frac{\delta Z_3(z;\rho_2,\rho_4)}{\delta\mu(z,\rho_1')}Z_0(z;\rho_1,\rho_3)\right\rangle. \tag{35}$$

Substituting (34a) and (34b) into (35), we get

$$\langle\mu_1,Z_3(2,4)Z_0(1,3)\rangle=\tfrac{1}{2}iK_0[A_n(0)-A_n(\rho_1-\rho_3)]\Gamma_{33}+\tfrac{1}{2}iK_0[A_n(\rho_1-\rho_2)-A_n(\rho_1-\rho_4)]\Gamma_{00}. \tag{36a}$$

Let us take the Hermitian conjugate of (35) and then interchange ρ_1 and ρ_3, ρ_2 and ρ_4; because of the equalities

$Z(1,3) = Z^+(3,1)$ and $Z(2,4) = Z^+(4,2)$, we get

$$\langle \mu_3, Z_3(2,4) Z_0(1,3)\rangle = \int d\rho_1' \, A_n(\rho_3 - \rho_1') \left(Z_3(z;\rho_2,\rho_4) \frac{\delta Z_0(z;\rho_1,\rho_3)}{\delta \mu(z,\rho_1')} + \frac{\delta Z_3(z;\rho_2,\rho_4)}{\delta \mu(z,\rho_1')} Z_0(z;\rho_1,\rho_3) \right).$$

Using (34a), (34b), and taking (32') into account, this last relation becomes

$$\langle \mu_3, Z_3(2,4) Z_0(1,3)\rangle = -\tfrac{1}{2} i K_0 [A_n(0) - A_n(\rho_1 - \rho_3)] \Gamma_{33} - \tfrac{1}{2} i K_0 [A_n(\rho_3 - \rho_4) - A_n(\rho_3 - \rho_2)] \Gamma_{00}. \tag{36b}$$

We introduce the following notation:

$$H(1,3) = K_0^2 [A_n(0) - A_n(\rho_1 - \rho_3)], \quad H(2,4) = K_0^2 [A_n(0) - A_n(\rho_2 - \rho_4)],$$
$$Q = \tfrac{1}{2} K_0^2 [A_n(\rho_1 - \rho_2) + A_n(\rho_3 - \rho_4) - A_n(\rho_1 - \rho_4) - A_n(\rho_3 - \rho_2)]; \tag{37}$$

then Eqs. (36a) and (36b) lead to

$$K_0^2 \langle \mu_1 - \mu_3, Z_3(2,4) Z_0(1,3)\rangle = i K_0 [H(1,3)\Gamma_{33} + Q\Gamma_{00}]. \tag{38a}$$

Still using (32'), a similar computation gives

$$K_0^2 \langle \mu_2 - \mu_4, Z_3(2,4) Z_0(1,3)\rangle = i K_0 [H(2,4)\Gamma_{00} + Q\Gamma_{33}]. \tag{38b}$$

Substituting (38a) and (38b) into (31a) gives

$$2iK_0 \left(\frac{\partial}{\partial z} + H(1,3) + Q \right) \Gamma_{33} + 2iK_0 [H(2,4) + Q] \Gamma_{00} + (\Delta_1 - \Delta_3)\Gamma_{03} + (\Delta_2 - \Delta_4)\Gamma_{30} = 0. \tag{39a}$$

One obtains in the same way

$$2iK_0 \left(\frac{\partial}{\partial z} + H(1,3) + Q \right) \Gamma_{03} + 2iK_0 [H(2,4) + Q] \Gamma_{30} + (\Delta_1 - \Delta_3)\Gamma_{33} + (\Delta_2 - \Delta_4)\Gamma_{00} = 0, \tag{39b}$$

$$2iK_0 \left(\frac{\partial}{\partial z} + H(1,3) + Q \right) \Gamma_{30} + 2iK_0 [H(2,4) + Q] \Gamma_{03} + (\Delta_1 - \Delta_3)\Gamma_{00} + (\Delta_2 - \Delta_4)\Gamma_{33} = 0, \tag{39c}$$

$$2iK_0 \left(\frac{\partial}{\partial z} + H(1,3) + Q \right) \Gamma_{00} + 2iK_0 [H(2,4) + Q] \Gamma_{33} + (\Delta_1 - \Delta_3)\Gamma_{30} + (\Delta_2 - \Delta_4)\Gamma_{03} = 0. \tag{39d}$$

This completes the set of equations for the quadruplet $\{\Gamma_{33}, \Gamma_{00}, \Gamma_{30}, \Gamma_{03}\}$.

B. Use of a special coordinate frame

We now consider the coordinate transformation[2,3]

$$R = \tfrac{1}{4}(\rho_1 + \rho_2 + \rho_3 + \rho_4), \quad \rho = \rho_1 - \rho_3 + \rho_2 - \rho_4,$$
$$\rho_1 = R + \tfrac{1}{2}(r_1 + r_2) + \tfrac{1}{4}\rho, \quad \rho_2 = R - \tfrac{1}{2}(r_1 + r_2) + \tfrac{1}{4}\rho,$$
$$r_1 = \tfrac{1}{2}(\rho_1 + \rho_3 - \rho_2 - \rho_4), \quad r_2 = \tfrac{1}{2}(\rho_1 - \rho_3 - \rho_2 + \rho_4), \tag{40}$$
$$\rho_3 = R + \tfrac{1}{2}(r_1 - r_2) - \tfrac{1}{4}\rho, \quad \rho_4 = R - \tfrac{1}{2}(r_1 - r_2) - \tfrac{1}{4}\rho.$$

Let $\nabla_R, \nabla_\rho, \nabla_{r_1}, \nabla_{r_2}$ denote the gradients with respect to the variables R, ρ, r_1, r_2; a simple calculation gives

$$\Delta_1 - \Delta_3 + \Delta_2 - \Delta_4 = 2(\nabla_R \cdot \nabla_\rho + \nabla_{r_1} \cdot \nabla_{r_2}). \tag{40'}$$

Summing and subtracting (39a) and (39d), then (39b) and (39c), we get

$$2iK_0 \left(\frac{\partial}{\partial z} + H(1,3) + H(2,4) + 2Q \right)(\Gamma_{33} + \Gamma_{00}) + (\Delta_1 - \Delta_3 + \Delta_2 - \Delta_4)(\Gamma_{30} + \Gamma_{03}) = 0, \tag{41a}$$

$$2iK_0 \left(\frac{\partial}{\partial z} + H(1,3) - H(2,4) \right)(\Gamma_{33} - \Gamma_{00}) + (\Delta_1 - \Delta_3 - \Delta_2 + \Delta_4)(\Gamma_{30} - \Gamma_{03}) = 0, \tag{41b}$$

$$2iK_0 \left(\frac{\partial}{\partial z} + H(1,3) + H(2,4) + 2Q \right)(\Gamma_{03} + \Gamma_{30}) + (\Delta_1 - \Delta_3 + \Delta_2 - \Delta_4)(\Gamma_{33} + \Gamma_{00}) = 0, \tag{41c}$$

$$2iK_0 \left(\frac{\partial}{\partial z} + H(1,3) - H(2,4) \right)(\Gamma_{03} - \Gamma_{30}) + (\Delta_1 - \Delta_3 - \Delta_2 + \Delta_4)(\Gamma_{33} - \Gamma_{00}) = 0. \tag{41d}$$

Now the quadruplet $\{\Gamma_{ij}\}$, $i,j = 0, 3$, is invariant under the transformation

$$(\rho_1, \rho_3) \Leftrightarrow (\rho_2, \rho_4), \quad \Gamma_{33} \Leftrightarrow \Gamma_{33}, \quad \Gamma_{00} \Leftrightarrow \Gamma_{00}, \quad \Gamma_{30} \Leftrightarrow \Gamma_{03} \tag{42}$$

so that if we impose the invariance of Eqs. (41) under this transformation, it follows from (41b) and (41d) that one has

$$\Gamma_{33} = \Gamma_{00}, \quad \Gamma_{30} = \Gamma_{03}, \tag{42'}$$

and Eqs. (41a) and (41c) reduce to

$$iK_0 \left(\frac{\partial}{\partial z} + H(1,3) + H(2,4) + 2Q \right) \Gamma_{33} + (\Delta_1 - \Delta_3 + \Delta_2 - \Delta_4)\Gamma_{30} = 0, \tag{43a}$$

$$iK_0 \left(\frac{\partial}{\partial z} + H(1,3) + H(2,4) + 2Q \right) \Gamma_{30} + (\Delta_1 - \Delta_3 + \Delta_2 - \Delta_4)\Gamma_{33} = 0. \tag{43b}$$

Let us now use the coordinates (40) and take into account (26) which imposes $\nabla_R \Gamma_{ij} = 0$, $i,j = 0, 3$, so that we may set $R = 0$; but one sees at once using (40') that ∇_ρ is eliminated from Eqs. (43a) and (43b). Thus ρ is a parameter that, for convenience, we may set equal to zero; thus one has

$$R = \rho = 0, \quad r_1 = \rho_1 - \rho_4 = \rho_3 - \rho_2,$$
$$r_2 = \rho_1 - \rho_3 = \rho_4 - \rho_2, \tag{44a}$$

which gives

$$\rho_1 = \tfrac{1}{2}(r_1 + r_2), \quad \rho_3 = \tfrac{1}{2}(r_1 - r_2), \quad \rho_2 = -\tfrac{1}{2}(r_1 + r_2),$$
$$\rho_4 = -\tfrac{1}{2}(r_1 - r_2). \tag{44b}$$

We set

$$\tilde{\Gamma}_{ij}(z; r_1, r_2) = \Gamma_{ij}(z; \rho_1, \rho_3; \rho_2, \rho_4)_{R=\rho=0}, \quad i,j = 0, 3. \tag{44c}$$

Using (44b), the relations (37) become

$$H(1,3) = H(2,4) = H = K_0^2 [A_n(0) - A_n(r_2)],$$
$$Q = \tfrac{1}{2} K_0^2 [A_n(r_1 + r_2) + A_n(r_1 - r_2) - 2A_n(r_1)]. \tag{44d}$$

If we substitute these last results into Eqs. (43), we get

$$iK_0\left(\frac{\partial}{\partial z}+2H+2Q\right)\tilde{\Gamma}_{33}+\nabla_{r_1}\cdot\nabla_{r_2}\tilde{\Gamma}_{30}=0,$$

$$iK_0\left(\frac{\partial}{\partial z}+2H+2Q\right)\tilde{\Gamma}_{30}+\nabla_{r_1}\cdot\nabla_{r_2}\tilde{\Gamma}_{33}=0,$$

and by summing and subtracting successively,

$$iK_0\left(\frac{\partial}{\partial z}+2H+2Q\right)(\tilde{\Gamma}_{33}+\tilde{\Gamma}_{30})+\nabla_{r_1}\cdot\nabla_{r_2}(\tilde{\Gamma}_{33}+\tilde{\Gamma}_{30})=0, \tag{45a}$$

$$iK_0\left(\frac{\partial}{\partial z}+2H+2Q\right)(\tilde{\Gamma}_{33}-\tilde{\Gamma}_{30})-\nabla_{r_1}\cdot\nabla_{r_2}(\tilde{\Gamma}_{33}-\tilde{\Gamma}_{30})=0. \tag{45b}$$

We shall now discuss the solutions of these equations.

IV. SOLUTIONS FOR THE FOURTH-ORDER COHERENCE FUNCTIONS

A. Boundary conditions

We still have to give the boundary conditions, and we repeat here some of the remarks mentioned in Ref. 1. In the spinor wave optics and for a nonturbulent medium, one has

$$I(z,\rho)=\psi^+(z,\rho)\psi(z,\rho),$$

$$P_j(z,\rho)=\psi^+(z,\rho)\sigma_j\psi(z,\rho),\ j=1,2,3,$$

$$I^2(z,\rho)=P^j(z,\rho)P_j(z,\rho), \tag{46}$$

and, since we consider a light beam propagating along the $0z$ axis, then both transverse components of $P_j(z,\rho)$ are zero, which gives

$$P_\alpha(z,\rho)=0,\quad \alpha=1,2,\quad I(z,\rho)=|P_3(z,\rho)|. \tag{47a}$$

This leads to

$$P_i(z,\rho_1)P_j(z,\rho_2)=\delta_{i3}\delta_{j3}I(z,\rho_1)I(z,\rho_2). \tag{47b}$$

From definition (5′) of the currents $Z_\mu(z;\rho_1,\rho_2),\mu=0,1,2,3$, we get

$$P_\alpha(z,\rho_1)=i\epsilon_\alpha^\beta Z_\beta(z;\rho_1,\rho_2),\quad \alpha=1,2,$$

$$P_3(z,\rho_1)=Z_0(z;\rho_1,\rho_1),\quad I(z,\rho_1)=Z_3(z;\rho_1,\rho_1). \tag{48}$$

The generalization of (47b) to a turbulent medium is

$$\Gamma_{\alpha\mu}(z;\rho_1,\rho_1;\rho_2,\rho_2)=0,\quad \alpha=1,2,\quad \mu=0,1,2,3,$$

$$\Gamma_{3j}(z;\rho_1,\rho_1;\rho_2,\rho_2)=\Gamma_{0i}(z,\rho_1,\rho_1;\rho_2,\rho_2),\quad i,j=0,3. \tag{49}$$

Using the coordinates r_1, r_2, the conditions on Γ_{3j} and Γ_{0i} become

$$\tilde{\Gamma}_{3j}(z;r_1,0)=\tilde{\Gamma}_{0i}(z;r_1,0),\quad i,j=0,3,$$

and, taking (42′) into account, we only have

$$\tilde{\Gamma}_{33}(z;r_1,0)=\tilde{\Gamma}_{30}(z;r_1,0), \tag{50}$$

a relation to be satisfied by the boundary conditions at $z=0$, which is in agreement with (47b):

$$\tilde{\Gamma}_{33}(0;r_1,r_2)=\tilde{\Gamma}_{30}(0;r_1,r_2)=I(0,(r_1+r_2)/2)I(0,-(r_1+r_2)/2). \tag{51}$$

This completes the data required to obtain the solutions of Eqs. (45).

B. Discussion of the solutions for Eqs. (45)

According to (51), Eq. (45b) has the trivial solutions

$$\tilde{\Gamma}_{33}(z;r_1,r_2)=\tilde{\Gamma}_{30}(z;r_1,r_2) \tag{52}$$

so that Eq. (45a) becomes

$$iK_0\left(\frac{\partial}{\partial z}+2H+2Q\right)\tilde{\Gamma}_{33}(z;r_1,r_2)+\nabla_{r_1}\cdot\nabla_{r_2}\tilde{\Gamma}_{33}(z;r_1,r_2)=0. \tag{53}$$

This is also the equation obtained for the fourth-order coherence function of a plane wave propagating in a strong turbulent medium starting with the parabolic equation (paraxial approximation of the scalar Helmholtz equation),[2,3] and the solution can be written in the form[2]

$$\tilde{\Gamma}_{33}(z;r_1,r_2)=\int\int_{-\infty}^{\infty}M(z;r_1,\chi)e^{i\chi\cdot r_2}d\chi$$

with

$$M(z;r,\chi)=M(0;r-\chi z/K_0,\chi)\times\exp\left[-\tfrac{1}{2}\pi K_0^2\int_0^z H(\zeta,r-\chi(z-\zeta)/K_0))\,d\zeta\right]+\int_0^z\exp\left[-\tfrac{1}{2}\pi K_0^2\int_{z'}^z H(\zeta,\chi(z-\zeta)/K_0)\,d\zeta\right]\times G(z',\chi(zz')/K_0,\chi)\,dz' \tag{54}$$

(for the definition of the functions G and H, see Ref. 2). This result needs the following comments:

1. As in Ref. 1 and in agreement with Lax *et al.*,[4] we see that the parabolic equation corresponds to a zeroth-order approximation of the light field equations in terms of the parameter $\epsilon=\omega_0 l^{-1}$ and that it supplies results only valid to this order of approximation.

2. As a consequence, the solution is not valid for a plane wave so that the condition (26), which is natural for a plane wave, appears now as an ad hoc tool for obtaining analytic solution.

Whatever it may be, Eq. (53) has been extensively studied, but it is a formidable task to look for a solution no matter whether an analytical or a numerical technique is used. Although a great deal of effort has been expended, only approximate asymptotic solutions have been obtained (see, for instance, Ref. 3).

Some recent surveys[5,6] discuss the experimental and theoretical situation on scintillations in strong turbulence, and we note that some results were obtained on a heuristic technique based on the extended Huyghens–Fresnel principle.

As a final comment, let us remark that with the choice of the r_1, r_2 coordinates and taking into account (44a), one cannot compute the covariance function for any couple of points (ρ_1,ρ_2); one only has

$$B_r(z;r_1,-r_1)=\langle I(z,r_1)I(z,-r_1)\rangle-\langle I(z,r_1)\rangle\langle I(z,-r_1)\rangle, \tag{55}$$

which is rather limited.

V. CONCLUSIONS

The present work suggests the following remarks:

1. The interpretation of the condition (26) $\nabla_R\Gamma_{ij}=0$

should be clarified.

2. Assuming $\Gamma_{33} = \Gamma_{30}$, it would be interesting, without using (26), to obtain a solution of the system (43a), (43b); that is, to solve the following equation:

$$iK_0\left(\frac{\partial}{\partial z} + H(1,3) + H(2,4) + 2Q\right)\Gamma_{33} + (\Delta_1 - \Delta_3 + \Delta_2 - \Delta_4)\Gamma_{33} = 0. \quad (56)$$

Even with numerical methods, it seems to be a difficult task.

3. We could also consider a plane wave with the coordinates (40) since in this case the condition (26) (more generally $\nabla_R \Gamma_{\mu\nu} = 0, \mu, \nu = 0,1,2,3$) is fulfilled, but we should have to start directly with Eqs. (20a), (20b). In Ref. 1, we proved, for the second-order moment, that the plane wave case is rather intricate; this suggests it could be here analytically unmanageable.

4. We conjecture that as long as one is not too far from the light axis the condition (26) for a finite beam and the zeroth-order approximation with repsect to ϵ are approximately valid.

[1] P. Hillion and S. Quinnez, "Strong fluctuations in the spinor wave optics. Mutual coherence function," J. Math. Phys. **24**, 712 (1983).

[2] J. W. Strohbehn, in *Laser Beam Propagation in the Atmosphere*, edited by J. W. Strohbehn (Springer-Verlag, Berlin, 1978).

[3] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978), Vol. 2.

[4] M. Lax, W. H. Louisell, and W. B. McKnight, Phys. Rev. A **11**, 1345 (1975).

[5] R L. Fante, Proc. IEEE **63**, 1669 (1975); **68**, 1424 (1980).

[6] A. M. Prokhorov, F. V. Bunkin, K. S. Gochelashvily, and V. I. Shishov, Proc. IEEE **63**, 790 (1975).