The unitary multiplicity-free representations of $\overline{SL_4(\mathbb{R})}$ [a)]

John L. Friedman
Physics Department, University of Wisconsin, Milwaukee, Wisconsin
Rafael D. Sorkin
Enrico Fermi Institute, University of Chicago, Chicago, Illinois

(Received 7 August 1979; accepted for publication 8 February 1980)

Those unitary representations of $\overline{SL_4(\mathbb{R})}$ are classified which contain each irreducible $\overline{SO(4)}$ representation at most once.

1. INTRODUCTION

Unitary representations of the group $\overline{SL_4(\mathbb{R})}$ [the universal covering group of $SL_4(\mathbb{R})$] have received some attention recently.[1-3] In this paper we classify all irreducible unitary representations ("unirreps") that are multiplicity-free with respect to an $\overline{SO_4}$ subgroup. Our method is based on a technique developed for SL_3 ($\mathbb{R}$) by Ogievetskii and Sokachev.[4] Our results disagree substantially with those given previously by Kihlberg[5] and by Ne'eman and Sijacki; in particular, we find that several of their claimed representations do not exist.

2. FORMALISM

The group SL_4 ($\mathbb{R}$) has SO_4 ($\mathbb{R}$) as its maximal compact subgroup. By choosing a metric $g_{\alpha\beta}$ on $\mathbb{R}^4$ we single out a particular SO_4 ($\mathbb{R}$) subgroup. The corresponding SO_4 Lie algebra $\mathscr{O}$ is spanned by the trace-free tensors $L^{\alpha}{}_{\beta}$ which are antisymmetric when their first index is lowered by $g_{\alpha\beta}$:

$$L_{\alpha\beta} = -L_{\beta\alpha}. \tag{1}$$

This subalgebra $\mathscr{O}$ acts by commutation (the adjoint representation) on the complementary subspace of trace-free tensors $Q^{\alpha}{}_{\beta}$, symmetric when lowered by $g_{\alpha\beta}$:

$$Q_{\alpha\beta} = Q_{\beta\alpha},$$

and under this action the Q's transform as a tensor representation.

The subalgebra $\mathscr{O}$ can be further decomposed into its even (self-dual) and odd (anti-self-dual) subspaces with respect to the operation $*$, defined on antisymmetric tensors $L_{\alpha\beta}$ by

$$*L_{\alpha\beta} := \tfrac{1}{2}\epsilon_{\alpha\beta}{}^{\gamma\delta}L_{\gamma\delta},$$

where $\epsilon_{\alpha\beta\gamma\delta}$ is the totally antisymmetric tensor normalized with respect to the metric $g_{\alpha\beta}$ (the sign of $\epsilon_{\alpha\beta\gamma\delta}$ can be chosen arbitrarily). Each of these subspaces is an ideal of $\mathscr{O}$, isomorphic as a Lie algebra to so_3 ($\mathbb{R}$), and the decomposition thereby realizes the relation

$$so_4 \simeq so_3 \oplus so_3. \tag{2}$$

Denote by iJ and iK the inclusion maps of the first and second so_3 subalgebras in the direct sum (2). Because our so_4, $\mathscr{O}$, is a space of tensors, we can regard iJ (iK) as itself a tensor in $\mathbb{E}^3 \otimes \mathbb{E}^4 \otimes \mathbb{E}^4$. The lower case Latin indices a, b and c will refer to vectors in $\mathbb{E}^3$ that correspond to self-dual tensors in $\mathbb{E}^4 \otimes \mathbb{E}^4$, and the lower case Latin indices p, q and r will refer to vectors in (strictly speaking a distinct) $\mathbb{E}^3$ that correspond to anti-self-dual tensors in $\mathbb{E}^4 \otimes \mathbb{E}^4$. Thus J and K will be written in the form $J_{a\alpha\beta}$, $K_{p\alpha\beta}$. In terms of J and K, the commutation relations of so_4 ($\mathbb{R}$) have the form

$$J_a{}^{\alpha}{}_{\gamma} J_b{}^{\gamma}{}_{\beta} - J_b{}^{\alpha}{}_{\gamma} J_a{}^{\gamma}{}_{\beta} = i\epsilon_{ab}{}^{c} J_c{}^{\alpha}{}_{\beta}, \tag{3}$$

$$K_p{}^{\alpha}{}_{\gamma} K_q{}^{\gamma}{}_{\beta} - K_q{}^{\alpha}{}_{\gamma} K_p{}^{\gamma}{}_{\beta} = i\epsilon_{pq}{}^{r} K_r{}^{\alpha}{}_{\beta}, \tag{4}$$

$$J_a{}^{\alpha}{}_{\gamma} K_p{}^{\gamma}{}_{\beta} - K_p{}^{\alpha}{}_{\gamma} J_a{}^{\gamma}{}_{\beta} = 0, \tag{5}$$

or, suppressing the $\mathbb{E}^4$ indices,

$$[J_a, J_b] = i\epsilon_{ab}{}^{c} J_c, \tag{6}$$

$$[K_p, K_q] = i\epsilon_{pq}{}^{r} K_r, \tag{7}$$

$$[J_a, K_p] = 0. \tag{8}$$

Here the tensor ϵ_{abc} is an antisymmetric tensor on $\mathbb{E}^3$ normalized with respect to the metric g_{ab} of $\mathbb{E}^3$.

The space of symmetric tracefree tensors $Q_{\alpha\beta}$ can be identified with the space of tensors iQ_{ap} having two $\mathbb{E}^3$ indices; in this case, the isomorphism is a tensor in $\mathbb{E}^3 \otimes \mathbb{E}^3 \otimes \mathbb{E}^4 \otimes \mathbb{E}^4$, namely

$$iQ_{ap}{}^{\alpha\beta} = 2J_a{}^{\alpha}{}_{\gamma} K_p{}^{\beta\gamma} = 2J_a{}^{\beta}{}_{\gamma} K_p{}^{\alpha\gamma}. \tag{9}$$

If, as in Eqs. (6)–(8), we suppress the $\mathbb{E}^4$ indices, the action of the so_4 subalgebra $\mathscr{O}$ on the symmetric tensors is expressed by the commutation relations

$$[J_a, Q_{bp}] = i\epsilon_{ab}{}^{c} Q_{cp}, \tag{10}$$

$$[K_p, Q_{aq}] = i\epsilon_{pq}{}^{r} Q_{ar}. \tag{11}$$

Equations (6)–(8), (10) and (11) exhibit the commutation relations of the antisymmetric ($\mathbb{E}^4$) tensors with each other and with the symmetric tensors. The remaining commutation relations of sl_4 ($\mathbb{R}$), those of the symmetric tensors with themselves are given by

$$[Q_{ap}, Q_{bq}] + i\epsilon_{ab}{}^{c} J_c g_{pq} + i\epsilon_{pq}{}^{r} K_r g_{ab} = 0. \tag{12}$$

Any unirrep of sl_4 ($\mathbb{R}$), when restricted to so_4, can be decomposed into a direct sum of irreducible finite dimensional so_4 representations. We will consider only multiplicity-free representations, those for which no representation of so_4 occurs more than once in the decomposition. A representation of so_4 is characterized by the eigenvalues $j(j+1)$ and $k(k+1)$ of the operators $\tilde{J}^2$ and $\tilde{K}^2$ that correspond to J^2 and

[a)] Work supported in part by the National Science Foundation under grants PHY 77-0711 at the University of Wisconsin-Milwaukee and PHY 78-24275 at the University of Chicago.

K^2. We will label the representation by the integers $m = 2j$ and $n = 2k$, eigenvalues of the number operators M and N, defined by

$$J^2 = \tfrac{1}{4}M(M+2), \quad K^2 = \tfrac{1}{4}N(N+2). \tag{13}$$

If we denote by $\mathscr{H}_{mn}$ the $(m+1)(n+1)$-dimensional substratum of a representation (m,n), then the substratum of an arbitrary multiplicity-free representation of $sl_4(\mathbb{R})$ can be regarded as a subspace of the Hilbert space

$$\mathscr{H} := \bigoplus_{m,n=0}^{\infty} \mathscr{H}_{mn}. \tag{14}$$

The associated lattice of points (m, n) will be denoted by H, so that $\mathscr{H} = \oplus_{(m,n)\in H}\mathscr{H}_{mn}$. The representation will be unitary if and only if Q_{ap} is represented by a self-adjoint operator $\widetilde{Q}_{ap}$.

We now follow an approach analogous to that used by Ogievetskii and Sokachev[4] in dealing with representations of $SL_3(\mathbb{R})$. We introduce ladder operators (operator valued $\mathbb{E}^3$ vectors) u_a and v_p which raise and lower by 2 the values of m and n, respectively:

$$[M,u_a] = -2u_a, \tag{15}$$

$$[N,v_a] = -2v_a; \tag{16}$$

u_a and v_p commute with each other,

$$[u_a,v_p] = 0, \tag{17}$$

and with each other's adjoints,

$$[u_a,v_p^*] = 0; \tag{18}$$

and, in their commutation relations with the operators $\tilde{J}_a$ and $\widetilde{K}_a$ that correspond to J_a and K_a, they transform as SO_4 tensors of types (1,0) and (0,1) respectively:

$$[\tilde{J}_a,u_b] = i\epsilon_{ab}{}^c u_c, \quad [\widetilde{K}_p,u_b] = 0, \tag{19}$$

$$[\tilde{J}_a,v_p] = 0, \quad [\widetilde{K}_p,v_q] = i\epsilon_{pq}{}^r v_r. \tag{20}$$

They satisfy an additional set of relations given in the Appendix (and in Ref. 4).

A tensor operator on $\mathscr{H}$ such as $\widetilde{Q}_{ap}$ whose commutation relations with J and K provide a (1,1) representation of so_4 can be written in terms of u_a, v_p, J_a, and K_p in the manner,

$$\widetilde{Q}_{ap} = u_a v_p A + \widehat{A}u_a^* v_p^* + v_p B u_a^* + u_a \widehat{B} v_p^* + J_a v_p C + \widehat{C} J_a v_p^* + u_a K_p D + \widehat{D} u_a^* K_p + J_a K_p E, \tag{21}$$

where $A = A(M,N)$, $\widehat{A} = \widehat{A}(M,N)$,...,$E = E(M,N)$ are SO_4-scalar operators constructed from M and N only. The substratum of an irreducible representation will not be $\mathscr{H}$, but rather a subspace

$$\mathscr{S} = \bigoplus_{(m,n)\in S} \mathscr{H}_{mn}, \tag{22}$$

where S is a sublattice of H. We may nonetheless regard $\widetilde{Q}$ as an operator acting on $\mathscr{H}$ by setting $\widetilde{Q}$ to zero on $\mathscr{H} - \mathscr{S}$ and setting $A(M,N)$,...,$E(M,N)$ to zero whenever the associated term in (21) would otherwise map $\mathscr{S}$ to $\mathscr{H} - \mathscr{S}$. Thus, for example, if $(m, n)\in S$ and $(m-2, n-2)\notin S$, the eigenvalue A_{mn} of A will vanish. In this way, the operator $\widetilde{Q}$ of any representation on $\mathscr{S} \subset \mathscr{H}$ can be expressed in the form (21).

With $\tilde{J}_a$, and $\widetilde{K}_p$ and $\widetilde{Q}_{ap}$ in this form, the commutation relations (6)–(8), (10) and (11), which express the SO_4 character of the operators, are automatically satisfied. The remaining commutation relation (12) then provides a set of algebraic relations among the coefficients A,...,E. These we give explicitly in the Appendix. Here, however, we will restrict ourselves to unitary representations. For these, it is implicit in Kihlberg's work[6] and is explicitly noted by Ne'eman and Sijacki[2] that $\widetilde{Q}_{ap} = \widetilde{Q}^*_{ap}$ can only connect a given subspace $\mathscr{H}_{mn}$ to the diagonally related subspaces $\mathscr{H}_{m+2,n+2}$, $\mathscr{H}_{m+2,n-2}$, $\mathscr{H}_{m-2,n+2}$, $\mathscr{H}_{m-2,n-2}$. For unitary, multiplicity-free representations, $\widetilde{Q}_{am}$ therefore has the form

$$\widetilde{Q}_{ap} = u_a v_p A + A^* u_a^* v_p^* + v_p B u_a^* + u_a B^* v_p^*. \tag{23}$$

The commutator $[\widetilde{Q}_{ap},\widetilde{Q}_{bq}]$ is then [see Appendix, Eq. (A6)] given by

$$[\widetilde{Q}_{ap},\widetilde{Q}_{bq}] + i\epsilon_{ab}{}^c\tilde{J}_c g_{pq} + i\epsilon_{pq}{}^r\widetilde{K}_r g_{ab} = i\epsilon_{ab}{}^c\tilde{J}_c[R_{J\delta}g_{pq} + R_{JKK}\widetilde{K}_p\widetilde{K}_q + R_{Jvv}v_p v_q + v_p^* v_q^* R_{Jv^*v^*}] + i\epsilon_{pq}{}^r\widetilde{K}_r[R_{K\delta}g_{ab} + R_{KJJ}\tilde{J}_a\tilde{J}_b + R_{Kuu}u_a u_b + u_a^* u_b^* R_{Ku^*u^*}], \tag{24}$$

where, denoting $A(M,N)$ and $B(M,N)$ by A_{MN} and B_{MN},

$$R_{J\delta} = -4(M+3)(N+2)^2|A_{M+2\ N+2}|^2 + 4(M-1)N^2|A_{MN}|^2 - 4(M+3)N^2|B_{M+2\ N}|^2 + 4(M-1)(N+2)^2|B_{MN+2}|^2 + 1, \tag{25}$$

$$R_{JKK} = 16(M+3)|A_{M+2\ N+2}|^2 - 16(M-1)|A_{MN}|^2 + 16(M+3)|B_{M+2\ N}|^2 - 16(M-1)|B_{MN+2}|^2, \tag{26}$$

$$R_{Jvv} = -4(M+3)A_{M+2\ N+2}B_{M+2\ N+4} + 4(M-1)A_{M\ N+4}B_{M\ N+2}, \tag{27}$$

$$R_{Jv^*v^*} = R_{Jvv}{}^*, \tag{28}$$

$$R_{K\delta} = -4(N+3)(M+2)^2|A_{M+2\ N+2}|^2 + 4(N-1)M^2|A_{MN}|^2 - 4(N+3)M^2|B_{M\ N+2}|^2 + 4(N-1)(M+2)^2|B_{M+2N}|^2 + 1, \tag{29}$$

$$R_{KJJ} = 16(N+3)|A_{M+2\ N+2}|^2 - 16(N-1)|A_{MN}|^2 + 16(N+3)|B_{M\ N+2}|^2 - 16(N-1)|B_{M+2N}|^2, \tag{30}$$

$$R_{Kuu} = -4(N+3)A_{M+2\ N+2}B^*_{M+4\ N+2} + 4(N-1)A_{M+4\ N}B^*_{M+2N}, \tag{31}$$

$$R_{Ku^*u^*} = R_{Kuu}{}^*, \tag{32}$$

The operator $\widetilde{Q}_{ap}$, and hence the commutator (24), is invariant under the simultaneous interchange

$$M \longleftrightarrow N, \quad u_a \longleftrightarrow v_p, \quad A \longleftrightarrow A, \quad B \longleftrightarrow B^*, \quad \tilde{J}_a \longleftrightarrow \widetilde{K}_p, \tag{33}$$

which exchange the roles of the two so_3's in so_4. Equations (25)–(32) for the coefficients $R_{J\delta},\ldots,R_{Kv^*v^*}$ are therefore also invariant under the interchange (33).

Let $\mathscr{S} \subset \mathscr{H}$ be the substratum of our representation. The operators A and B must then satisfy the relation

$$P_{mn}\{[\widetilde{Q}_{ap},\widetilde{Q}_{bq}] + i\epsilon_{ab}{}^c\tilde{J}_c g_{pq} + i\epsilon_{pq}{}^r\widetilde{K}_p g_{ab}\} = 0, \tag{34}$$

where P_{mn} is the projection operator onto any subspace $\mathscr{H}_{mn} \subset \mathscr{S}$. Equation (34) implies, for $(m, n) \in S$ that

$$R_{J\delta} = 0, \quad m \neq 0, n \neq 1; \quad R_{K\delta} = 0, \quad n \neq 0, m \neq 1; \tag{35}$$

$$R_{JKK} = 0, \quad m \neq 0, n \neq 0,1; \quad R_{KJJ} = 0, \quad m \neq 0,1, n \neq 0; \tag{36}$$

$$R_{Jvv} = 0, \quad m \neq 0; \quad R_{Kuu} = 0, \quad n \neq 0. \tag{37}$$

When $n = 1$, $\widetilde{K}_p \widetilde{K}_q = \frac{1}{3}g_{pq}\widetilde{K}^2 = \frac{1}{4}g_{pq}$, because $K^2 = \frac{1}{4}N(N+2)$, and the $R_{J\delta}$ and R_{JKK} terms provide only one independent relation, namely

$$R_{J\delta} + \tfrac{1}{4}R_{JKK} = 0, \quad n = 1, m \neq 0. \tag{38}$$

Similarly, at $m = 1$ we have

$$R_{K\delta} + \tfrac{1}{4}R_{KJJ} = 0, \quad m = 1, n \neq 0. \tag{39}$$

Equations (35)–(39) are equivalent to the commutation relation (34). Consequently, given any operators A and B and subset $S \subset H$, such that A and B leave invariant $\mathscr{S} = \oplus_{(m,n)\in S}\mathscr{H}_{mn}$ and satisfy equations (35)–(39) for all $(m, n) \in S$, we get a representation of $sl_4(\mathbb{R})$ with substratum $\mathscr{S}$.

Equations (35) and (36) have a relatively simple explicit form when written as the linear combinations

$$R_{J\delta} + \tfrac{1}{4}N^2 R_{JKK} = 0, \quad R_{K\delta} + \tfrac{1}{4}M^2 R_{KJJ} = 0, \tag{40}$$

$$R_{J\delta} + \tfrac{1}{4}(N+2)^2 R_{JKK} = 0, \quad R_{K\delta} + \tfrac{1}{4}(M+2)^2 R_{KJJ} = 0. \tag{41}$$

Equation (40) reduces when $n = 1$ ($m = 1$) to Eqs. (38) and (39); Eqs. (40) and (41) are therefore equivalent to Eqs. (35), (38) and (39). Explicitly, the full set of independent equations is [in the order (40), (41), (37)]

$$(M+3)(N+1)|A_{M+2\ N+2}|^2 - (M-1)(N+1) \times |B_{M\ N+2}|^2 = \tfrac{1}{16}, \quad m \neq 0, \tag{42}$$

$$(M+1)(N+3)|A_{M+2\ N+2}|^2 - (M+1)(N-1) \times |B_{M+2\ N}|^2 = \tfrac{1}{16}, \quad n \neq 0, \tag{42'}$$

$$(M-1)(N+1)|A_{MN}|^2 - (M+3)(N+1)|B_{M+2\ N}|^2 = \tfrac{1}{16}, \quad m \neq 0, n \neq 0,1, \tag{43}$$

$$(M+1)(N-1)|A_{MN}|^2 - (M+1)(N+3)|B_{M\ N+2}|^2 = \tfrac{1}{16}, \quad m \neq 0,1, n \neq 0, \tag{43'}$$

$$(M+3)A_{M+2\ N+2}B_{M+2\ N+4} - (M-1)A_{M\ N+4}B_{M\ N+2} = 0, \quad m \neq 0, \tag{44}$$

$$(N+3)A_{M+2\ N+2}B^*_{M+4\ N+2} - (N-1)A_{M+4\ N}B^*_{M+2\ N} = 0, \quad n \neq 0. \tag{44'}$$

The equation hold, with the exclusions given, when $(m, n) \in S$. The exclusions mean geometrically that an equation is not valid at (m, n) unless every operator A or B that appears in the equation with nonzero coefficient connects (m, n) to a point (m', n') in H. In other words, precisely those equations are excluded in which meaningless eigenvalues would otherwise occur (with nonzero coefficients).

Ignoring for a moment the fact that the nature of the substratum may affect the solution to these equations, it is easy to check that they admit the following generic family of positive real solutions, parameterized by a real number k:

$$A_{MN}(k) = \frac{1}{8}\left[\frac{k + (M+N)^2}{(M^2-1)(N^2-1)}\right]^{1/2}, \tag{45a}$$

$$B_{MN}(k) = \frac{1}{8}\left[\frac{k + (M-N)^2}{(M^2-1)(N^2-1)}\right]^{1/2}, \tag{45b}$$

We will find that, to within unitary equivalence, this family encompasses all multiplicity-free unirreps of $sl_4(\mathbb{R})$.

Note that these solutions blow up at $m = 1$ and $n = 1$. However, because the operator A occurs in $\widetilde{Q}$ only in the combination uvA, and $uvA(\mathscr{H}_{mn}) = 0$, for m or $n \leqslant 1$, these eigenvalues A_{mn} for m or $n \leqslant 1$ have no meaning for a representation in any event. Similarly, B_{mn} with m or $n \leqslant 1$ has no meaning for any representation. Nonetheless, the fact that the form (45) blows up for m or $n = 1$ has an important consequence: In valid equations where A_{mn} or B_{mn}, with m or $n = 1$ formally occur, their coefficients vanish—for example, in equation (43), the coefficient of A_{MN} is $(M-1)(N+1)$ which vanishes at $m = 1$. Thus an expression 0/0 occurs, and form (45) in general fails to satisfy the recursion relations (42), (43), and (44) at $m = 1$ and (42′), (43′), and (44′) at $n = 1$. Additional checks are therefore necessary to verify the existence of a representation when points with $m = 1$ or $n = 1$ occur in S.

The parameter k may be expressed in terms of the quadratic Casimir operator $\widetilde{Q}_{ap}\widetilde{Q}^{ap} - \tilde{J}^2 - \widetilde{K}^2$ by the equation

$$k = 4(\widetilde{Q}_{ap}\widetilde{Q}^{ap} - \tilde{J}^2 - \widetilde{K}^2) - 16. \tag{46}$$

3. CLASSIFICATION OF THE MULTIPLICITY-FREE UNIRREPS

We begin by stating as a theorem the central result of this section and then devote the remainder of our space to its proof. The substratum $\mathscr{S}$ of an irreducible representation can contain at most all subspaces $\mathscr{H}_{m'n'}$ connected to a single $\mathscr{H}_{mn}$ by repeated action of the operators uvA, vBu^*, and their adjoints; that is, S must be a subset of

$$I_{mn} = \{(m',n')| \bmod 4, m'-m \equiv n'-n \equiv 0 \text{ or } 2\}, \tag{47}$$

for some integers $m = 0, 1, 2, 3$, $n = 0, 1$. It turns out that the only sublattices which arise are the I_{00}, I_{20}, the lines

$$L_{00} = \{m',m')| m' \equiv 0 \mod 4\}, \tag{48}$$

$$L_{11} = \{(m',m')|m'\equiv 1 \mod 4\}, \tag{49}$$

and the triangular lattices (for $m, n > 0$)

$$T_{m0} = \{(m',n')\in I_{m0}|m' - m \geqslant n'\}, \tag{50}$$

$$T_{0n} = \{(m',n')\in I_{0n}|n' - n \geqslant m'\}. \tag{51}$$

Theorem: Every multiplicity-free unitary irreducible representation of the Lie algebra $sl_4(\mathbb{R})$ is unitarily equivalent to a representation of the form (45) with sublattice and associated value of the real parameter k given as follows:

$S = I_{00}$, $k > -4$,

$S = I_{20}$, $k > 0$,

$S = L_{00}$, $k = -4$,

$S = L_{11}$, $k = -4$,

$S = T_{m0}$, $k = -(m-2)^2$ for $m > 0$,

$S = T_{0n}$, $k = -(n-2)^2$ for $n > 0$.

Note that only the last two cases correspond to spinor representations of $SL_4(\mathbb{R})$ [representations of $SL_4(\mathbb{R})$ that are not also representations of $SL_4(\mathbb{R})$], and for these the parameter k can only take a discrete set of values.

In order to establish this result, several steps are needed. Considering first the absolute values $|A|$ and $|B|$ of the operators A and B of a representation with sublattice S, we show that $|A|$ and $|B|$ provide a representation given by the generic formula (45) for some k, and also having sublattice S. We then show that when S is convex, the replacement of A and B by their absolute values is a unitary equivalence. Finally, we classify the (irreducible) generic representations, finding in particular that their associated sublattices *are* convex. It thus follows that every unirrep is equivalent to one of the generic representations just classified.

In what follows, the operator A will be said to connect points $(m-2, n-2)$ and (m, n) in H if $uv\,A\,(\mathscr{H}_{mn}) \neq 0$. Similarly, B connects $(m-2, n)$ and $(m, n-2)$ if $vBu^*(\mathscr{H}_{m-2\,n}) \neq 0$. And a subset of H will be called "connected" (with respect to a particular A, B) if it does not fall into disjoint subsets which are not connected to each other by A or B. We will denote by A_{mn} (B_{mn}) the eigenvalue of A (B) corresponding to eigenvalues m and n of the operators M and N; that is, $Aw = A_{mn}w$ for $w \in \mathscr{H}_{mn}$.

A. Two lemmas

Lemma 1: If A and B furnish a unirrep on $\mathscr{S}$, then $|A|$ and $|B|$ also furnish a unirrep on $\mathscr{S}$. Moreover, $|A|$ and $|B|$ are given on S by the generic formula (45) for some fixed value of the parameter k; that is, there is a number k such that if $(m-2, n-2)$, $(m, n)\in S$, $|A_{mn}|$ is given by the generic formula (45a),

$$|A_{mn}|^2 = \frac{1}{64}\,\frac{k + (m+n)^2}{(m^2-1)(n^2-1)},$$

and if $(m-2, n)$, $(m, n-2)\in S$, $|B_{mn}|$ is given by (45b),

$$|B_{mn}|^2 = \frac{1}{64}\,\frac{k + (m-n)^2}{(m^2-1)(n^2-1)}.$$

Proof: It follows by inspection that $|A|$ and $|B|$ satisfy the recursion relations (42)–(44′) whenever A and B do so. Furthermore, $|A|$ (respectively $|B|$) vanishes if and only if A (respectively B) vanishes, so that $\mathscr{S}$ remains an irreducible invariant subspace. This establishes the first part of the Lemma.

Now let $\widetilde{S}$ be a connected subset of S containing at least two points (m_1, n_1) and (m_2, n_2), and let $\widetilde{S}' \subset S$ be the (connected) subset that results from adjoining to $\widetilde{S}$ all (m', n') connected by A or B to some $(m, n)\in\widetilde{S}$. Suppose that $|A|$ or $|B|$ are generic on $\widetilde{S}$; we will first show that $|A|$ and $|B|$ are generic on $\widetilde{S}'$.

A point in $\widetilde{S}$ can connected to S-$\widetilde{S}$ four ways:

1. A connects $(m-2, n-2)\in\widetilde{S}$ to $(m, n)\notin\widetilde{S}$,
2. A connects $(m-2, n-2)\notin\widetilde{S}$ to $(m, n)\in\widetilde{S}$,
3. B connects $(m-2, n)\in\widetilde{S}$ to $(m, n-2)\notin\widetilde{S}$,
4. B connects $(m-2, n)\notin\widetilde{S}$ to $(m, n-2)\in\widetilde{S}$,

Case 1: We must show that $|A_{mn}|$ is generic, $|A_{mn}| = A_{mn}(k)$. Since $\widetilde{S}$ is connected, $(m-2, n-2)$ is connected either (a) to $(m-4, n-4)\in\widetilde{S}$ by A, (b) to $(m-4, n)\in\widetilde{S}$ by B, or (c) to $(m, n-4)\in\widetilde{S}$ by B.

(1a). Here $|A_{m-2\,n-2}|$ is generic and $(m-4, n-4)\in\widetilde{S}\Rightarrow m\geqslant 4$, $n\geqslant 4\Rightarrow$Eq. (43′) holds at $(m-2, n-2)$. This equation gives $|B_{m-2\,n}|$:

$$(m-1)(n-3)|A_{m-2\,n-2}|^2 - (m-1)(n+1)|B_{m-2\,n}|^2 = \tfrac{1}{16}$$

$$\Rightarrow |B_{m-2\,n}|^2 = -\frac{1}{16(m-1)(n+1)} + \frac{n-3}{n+1}\cdot\frac{1}{64}\,\frac{k+(m+n-4)^2}{[(m-2)^2-1][(n-2)^2-1]}$$
$$= \frac{k+(m-2-n)^2}{[(m-2)^2-1][(n^2-1)]} \tag{52}$$

That is, $|B_{m-2\,n}|$ is generic. Then $m, n\geqslant 4$ implies that Eq. (42) is valid at $(m-2, n-2)$, and when $|B_{m-2\,n}|$ is generic, Eq. (42) implies $|A_{mn}|$ is generic.

(1b). Here $|B_{m-2\,n}|$ is generic. Equation (42) is valid at $(m-2, n-2)$ and implies that $|A_{mn}|$ is generic.

(1c). $|B_{m\,n-2}|$ is generic and Eq. (42′) implies that $|A_{mn}|$ is generic.

Case 2: Again we must show that $|A_{mn}|$ is generic. Here $(m, n)\in\widetilde{S}$ is connected either (a) to $(m+2, n+2)\in\widetilde{S}$ by A, (b) to $(m-2, n+2)\in\widetilde{S}$ by B, or (c) to $(m+2, n-2)\in\widetilde{S}$ by B.

(2a). $|A_{m+2\,n+2}|$ is generic and Eq. (42) at (m, n) implies $|B_{m\,n+2}|$ is generic. Equation (43′) holds at (m, n) and implies $|A_{mn}|$ is generic.

(2b). $|B_{m\,n+2}|$ is generic, and Eq. (43′), valid at (m, n), implies $|A_{mn}|$ is generic.

(2c). $|B_{m+2\,n}|$ is generic and Eq. (43), valid at (m, n), implies $|A_{mn}|$ is generic.

Case 3: We must show that B_{mn} is generic. $(m-2, n)\in\widetilde{S}$ is connected either (a) to $(m, n+2)$ by A, (b) to $(m-4, n-2)$ by A, or (c) to $(m-4, n+2)$ by B.

(3a). $|A_{m\,n+2}|$ is generic; and Eq. (42′), valid at $(m-2, n)$ because $(m, n-2)\in S\Rightarrow n\neq 0$, implies $|B_{mn}|$ is generic.

(3b). $|A_{m-2\,n}|$ is generic, and Eq. (43), valid at $(m-2, n)$, implies $|B_{mn}|$ is generic.

(3c). $|B_{m-2\,n+2}|$ is generic, and Eq. (42), valid at $(m-2, n)$, implies $|A_{m\,n+2}|$ is generic. Equation (42′) holds at $(m-2, n)$ and implies $|B_{mn}|$ is generic.

Case 4: This is identical to case 3 with m and n exchanged. Thus if $|A|$ and $|B|$ are generic on $\widetilde{S}$, they are generic on $\widetilde{S}'$. Let $\widetilde{S}$ be a maximal connected subset of S such that $|A|$ and $|B|$ are generic on $\widetilde{S}$. $\widetilde{S}$ contains at least two points (m_1, n_1) and (m_2, n_2), because we can choose any two points connected by, for example, $uvA:\mathscr{H}_{mn}\rightarrow\mathscr{H}_{m-2\,n-2}$ and define k by

$$k = 64(m^2-1)(n^2-1)|A_{mn}|^2 - (m+n)^2.$$

If $\widetilde{S}\neq S$, A or B connects $\widetilde{S}$ to S-$\widetilde{S}$ because S is irreducible. We have just shown that $|A|$ and $|B|$ would then be generic on the extended subspace $\widetilde{S}'$ and $\widetilde{S}$ would not be maximal. Thus $\widetilde{S}=S$ and $|A|$ and $|B|$ are generic on S, as asserted. □

Lemma 2: Let A and B, given on S by the generic form (45) with parameter k, provide a unirrep with substratum $\mathscr{S}$ given by (22). Then:

(a) if $(m, n)\in S$, $A_{mn}(k)=0$ if and only if $(m-2, n-2)\in H$-S;

(b) if $(m-2, n-2)\in S$, $A_{mn}(k)=0$ if and only if $(m, n)\in H$-S;

(c) similarly, if $(m-2, n)\in S$, $B_{mn}(k)=0$ if and only if $(m, n-2)\in H$-S;

(d) if $(m, n-2)\in S$, $B_{mn}(k)=0$ if and only if $(m-2, n)\in H$-S.

In other words, if (m_1, n_1) is in S, then any generic coefficient $A_{mn}(k)$ or $B_{mn}(k)$ that would connect (m_1, n_1) to a second point $(m_2, n_2)\in H$ vanishes if and only if $(m_2, n_2)\notin S$.

Proof: If $(m, n)\in S$ and $(m-2, n-2)\in H$-S, or vice versa, then A_{mn} must vanish in order that Q not map $\mathscr{S}$ to $\mathscr{H}-\mathscr{S}$. But, as was seen in cases 1 and 2 of the proof of Lemma 1, A_{mn} appears in a valid equation with $B_{m'n'}$ whose value is generic, and the equation implies that $|A_{mn}|$ is itself generic, that $|A_{mn}| = A_{mn}(k)$. Thus $A_{mn}(k)=0$.

Analogously, if $(m-2, n)\in S$ and $(m, n-2)\in H$-S, or vice versa, then $B_{mn}=0$. From cases 3 and 4 of the proof of Lemma 1, we find that B_{mn} appears in a valid equation with $A_{m'n'}$ whose value is generic, whence $|B_{mn}|$ is generic and $B_{mn}(k)=0$.

Conversely, if $(m, n)\in S$ and $A_{mn}(k)=0$, then by the generic formula, $(m+n)^2=-k$, whence $A_{m'n'}(k)=0$ for $(m'+n')^2=-k$. Consequently, the representation connects no point (m', n') of S with $(m'+n')^2\geqslant -k$ to a point (m'', n'') with $(m''+n'')^2<-k$. Since S contains the point (m, n), it can have no point (m', n') with $(m'+n')^2<-k$, for if it had, then S would fall into two disconnected pieces, contradicting the irreducibility of the representation. Thus $(m-2, n-2)\notin S$.

The argument for B is again analogous. □

B. Unitary equivalence

We will consider unitary transformations $U:\mathscr{H}\rightarrow\mathscr{H}$ which commute with the operators $\tilde{J}$ and $\widetilde{K}$ and will observe that they lead to a symmetry of the recursion relations (42)–(44′). Any unitary operator U on $\mathscr{H}$ which commutes with the so_4 subalgebra $\mathscr{O}$ has the property that

$$Uw = e^{i\theta_{mn}}w \quad \text{for} \quad w\in\mathscr{H}_{mn},$$

which follows from Schur's Lemma and the fact that $\tilde{J}$ and $\widetilde{K}$ generate an irreducible representation of $\overline{SO_4}$ on each subspace $\mathscr{H}_{mn}$. The operator U can then be written in the form

$$U = e^{i\theta_{MN}}, \tag{53}$$

with θ a real function of M and N.

Given a representation of sl_4 in which $\widetilde{Q}_{ap}$ has the form (23), for some operators A and B, the transformation U furnishes a unitarily equivalent representation of the same form, but with operators $\widehat{A}$ and $\widehat{B}$ defined by

$$U\,uvA\,U^{-1} = uv\widehat{A}, \tag{54a}$$

$$U\,vBu^*U^{-1} = v\widehat{B}u^*, \tag{54b}$$

Using Eqs. (15), (16), and (53), we have

$$\widehat{A} = e^{i(\theta_{M-2\,N-2}-\theta_{MN})}A, \tag{55a}$$

$$\widehat{B} = e^{i(\theta_{M\,N-2}-\theta_{M-2\,N})}B. \tag{55b}$$

Because U commutes with $\tilde{J}$ and $\widetilde{K}$, and $\widetilde{Q}_{ap}$ satisfies the commutation relation (34), $U\widetilde{Q}_{ap}U^{-1}$ also satisfies (34), and therefore $\widehat{A}$ and $\widehat{B}$ satisfy the recursion relations (42)–(44′) for a lattice S whenever A and B do so.

This latter fact is easily verified directly as follows. Equations (42)–(43′) involve only $|A|$ and $|B|$ and are automatically satisfied, since $|\widehat{A}| = |A|$ and $|\widehat{B}| = |B|$. Equation (44),

$$(M+3)A_{M+2\,N+2}B_{M+2\,N+4} = (M-1)A_{M\,N+4}B_{M\,N+2},$$

implies, using (55),

$$\begin{aligned}(M+3)&\widehat{A}_{M+2\,N+2}e^{-i(\theta_{MN}-\theta_{M+2\,N+2})}\widehat{B}_{M+2\,N+4}\\ &\times e^{-i(\theta_{M+2\,N+2}-\theta_{MN+4})}\\ &=(M-1)\widehat{A}_{MN+4}e^{-i(\theta_{M-2\,N+2}-\theta_{MN+4})}\widehat{B}_{MN+2}\\ &\times e^{-i(\theta_{MN}-\theta_{M-2\,N+2})}\\ &\Rightarrow(M+3)\widehat{A}_{M+2\,N+2}\widehat{B}_{M+2\,N+4}\\ &=(M-1)\widehat{A}_{MN+4}\widehat{B}_{MN+2}.\end{aligned}$$

(The operator $\theta_{M-2\,N+2}$ that appears is undefined when $m=0$ or 1, but points having $m=0$ are excluded from the range of validity of (44), and when $m=1$, the expression containing $\theta_{M-2\,N+2}$ vanishes.) Equation (44′) is similarly satisfied by $\widehat{A}$ and $\widehat{B}$ when satisfied by A and B.

We can now prove

Lemma 3: Any unitary irreducible representation for which $\widetilde{Q}_{ap}$ is of the form (23) and whose substratum is associated with a convex lattice S is unitarily equivalent to a representation of the form (23) with A and B generic.

Proof: Pick as a base point any $(m_0, n_0)\in S$. Since S cannot be a single point, (m_0, n_0) will be connected by A or B to some other point of S, for example to (m_0-2, n_0-2). (The other three cases are analogous.) Then, by Lemma 2, $A_{m_0n_0}\neq 0$, and there is, according to (55a) a unique choice of $\theta_{m_0-2\,n_0-2}$ which renders $A_{m_0n_0}$ positive, if we set $\theta_{m_0n_0}=0$. Then, by Lemma 1, $A_{m_0n_0}$ will be generic. Proceeding in this way from (m_0-2, n_0-2) we can reach any point (m', n') of S and assign a unique $\theta_{m'n'}$ such that each A_{mn} or B_{mn} corresponding to a "link" of the path has its generic value.

Moreover, we claim the resulting θ is independent of path. For because S is convex, any path from (m_0, n_0) to $(m',$

n') can be deformed into any other such path by a series of deformations, each of which simply changes the order of a pair of A and B steps. (Geometrically, this is changing the way the path connects one corner of a diamond to its opposite corner.) There are four different cases to consider, but because the argument is the same for each, we give only one explicitly. Suppose that a path $\cdots (m, n), (m-2, n-2), (m, n-4), \cdots$ is to be deformed into $\cdots (m, n), (m+2, n-2), (m, n-4), \cdots$. We claim that for a given θ_{mn}, both paths produce the same $\theta_{m\,n-4}$. According to (55), the reality of $\hat{A}_{mn}$ is equivalent (calling the phases of A and B "α" and "β", respectively) to

$$0 = \hat{\alpha}_{mn} = \theta_{m-2\;n-2} - \theta_{mn} + \alpha_{mn} \quad \text{(all phases mod } 2\pi\text{)},$$
$$\Rightarrow \theta_{m-2\;n-2} = \theta_{mn} - \alpha_{mn}.$$

Then $\hat{\beta}_{m\;n-2} = 0 \Rightarrow$

$$\theta_{m\,n-4} = \theta_{m-2\;n-2} - \beta_{m\;n-2} = \theta_{mn} - \alpha_{mn} - \beta_{m\,n-2}. \tag{56}$$

Similarly, $\hat{\beta}_{m+2\,n} = 0 = \hat{\alpha}_{m+2\;n-2} \Rightarrow$

$$\theta_{m+2\,n-2} = \theta_{mn} - \beta_{m+2\,n},$$
$$\theta_{m\,n-4} = \theta_{m+2\,n-2} - \alpha_{m+2\;n-2} = \theta_{mn} - \beta_{m+2\,n} - \alpha_{m+2\,n-2}. \tag{57}$$

The two paths produce the same θ_{mn-4} if Eqs. (56) and (57) agree, namely if

$$\alpha_{mn} + \beta_{m\,n-2} = \beta_{m+2\,n} + \alpha_{m+2\;n-2}. \tag{58}$$

But this last equation, (58), follows immediately from Eq. (44) at $(m, n-4)$. Notice in this connection first that since $(m, n-4)$ is one of the points of one of our paths it lies in S; second, that (44) is not excluded there since the occurrence of $(m-2, n-2)$ in a path implies $m \geqslant 2$, hence $m \neq 0$; and third, that $m+3$ and $m-1$ in (44) are necessarily positive at $(m, n-4)$, since $m \geqslant 2$. □

C. Classification of the generic representations

Let us begin with

Case 0: Unirreps with min $(m+n) = 0$ have either
(a) sublattice L_{00} and $k = -4$, or
(b) sublattice I_{00} and $k > -4$.

Proof: min $(m+n) = 0 \Rightarrow (0,0) \in S \Rightarrow (2,2) \in S$ because $(2, 2)$ is the only point of H that can be connected to $(0, 0)$ by A or B. Suppose $(4, 0) \notin S$ [this will correspond to (a) above]. Then by Lemma 2, $B_{42}(k) = 0 \Rightarrow k = -4$, by inspection of $(45b) \Rightarrow B_{m\,m+2} = 0$ and $B_{m+2\;m} = 0$ for all integers m, and thus L_{00} is not connected to H-L_{00}. Moreover, $A_{mn} \neq 0$ for $m \geqslant 2$, and therefore $(0, 0)$ is connected to all of L_{00} by successive applications of $A^*u^*v^*$ (and hence of $\tilde{Q}$) to $\mathscr{H}_{00}$. Consequently, $S = L_{00}$ and so $(4, 0) \notin S \Rightarrow$(a) above.

If on the other hand, $(4, 0) \in S$, $B_{42}(k) \neq 0$ [otherwise we would have $k = -4$ and the representation would be (a)]. Since if $k < -4$, the generic representation does not furnish a real number $B_{42}(k)$, we must have $k > -4$. Because $A_{mn}(k)$ and $B_{mn}(k)$ are then nonvanishing on I_{00} [in the sense that any $A_{mn}(k)$ or $B_{mn}(k)$ which can connect two points of S is nonzero], Lemma 2 implies that S cannot be a proper subset of I_{00}. Therefore $S = I_{00}$. Because I_{00} and L_{00} contain no points with m or $n = 1$, the generic form satisfies all recursion relations and therefore provides a representation of sl_4, on L_{00} when $k = -4$ and on I_{00} for every $k > -4$.

Case 1: Unirreps with min $(m+n) = 1$ have sublattice T_{10} or T_{01} and $k = -1$.

Proof: If min $(m+n) = 1$, either $(1, 0)$ or $(0, 1)$ are in S. If, say, $(1, 0) \in S$, Eq. (42) holds at $(1, 0)$ and gives

$$4|A_{32}|^2 = \tfrac{1}{16} \Rightarrow |A_{32}| = \tfrac{1}{8} \Rightarrow k = -1.$$

Then $k = -1 \Rightarrow B_{s\,s+1} = 0$ for all integers $s \neq 0$, 1 and $B_{mn}, A_{mn} \neq 0$ on T_{10}. Therefore, by Lemma 2, all points in T_{10} must be in S, and, since $B_{s\,s+1} = 0$, T_{10} is not connected to H-T_{10}, whence $T_{10} = S$. The only point in T_{10} with m or $n = 1$ is $(1, 0)$ and the only equations valid at $(1, 0)$ are (42) and (44). But (42) has already been verified and (44) holds because $B_{34} = 0$. At all other points of T_{10}, the generic form with $k = -1$ automatically satisfies the recursion relations (42)–(44') and thereby furnishes a representation. The case where $(0, 1) \in S$ ($S = T_{01}$) is symmetric.

Case 2. Unirreps with $\min(m+n) = 2$ have either
(a) sublattice L_{11} and $k = -4$,
(b) sublattice T_{20} or T_{02} and $k = 0$, or
(c) sublattice I_{20} and $k > 0$.

Proof. If $\min(m+n) = 2$, either $(1, 1)$, $(0, 2)$, or $(2, 0)$ are in S. If $(1, 1) \in S$, (42) holds at $(1, 1)$ and gives

$$8|A_{33}|^2 = \tfrac{1}{16} \Rightarrow k = -4.$$

Then $k = -4 \Rightarrow B_{s\,s+2} = B_{s+2\,s} = 0$ for all integers $s \neq 1$ and $A_{ss} \neq 0$ for any integer $s \geqslant 3$. Thus $S = L_{11}$. Now $(1, 1)$ is the only point of L_{11} with m or $n = 1$, and so we need only check those equations valid at $(1, 1)$, namely (42), (42'), (44), and (44'). But (42) was verified above and (42') is identical; (44) and (44') are satisfied when B_{35} and B_{53} vanish, which is the case here. Therefore the generic form with $k = -4$ furnishes a representation on L_{11}.

If $(2, 0) \in S$, either B_{22} or A_{42} must be nonzero in order that $(2, 0)$ be connected to the rest of I_{20}. If $B_{22} = 0$, $k = 0$, which implies that $B_{ss} = 0$, $s \neq 1$, and hence that no point in T_{20} is connected to $I_{20} - T_{20}$. Moreover, A and B are nonzero on the remainder of T_{20}, so $S = T_{20}$. Since neither $m = 1$ nor $n = 1$ occur on T_{20}, the generic form with $k = 0$ furnishes a representation on T_{20} (or on T_{02} by the $m \longleftrightarrow n$ symmetry).

If $(2, 0) \in S$ but $B_{22} > 0$, then $k > 0$, in which case A and B are nonzero on the whole of I_{20}, and so $S = I_{20}$. Again, because $m = 1$ and $n = 1$ do not occur, the generic form furnishes a representation on I_{20} for all $k > 0$.

Case 3. Unirreps with $\min(m+n) = s \geqslant 3$ have sublattice T_{s0} or T_{0s} and $k = -(s-2)^2$.

Proof. We show first that the only possible points $(m, n) \in S$ for which $m + n = s$ are $(s, 0)$ and $(0, s)$—that is, that $(s-t, t) \notin S$, for $0 < t < s$. For $t = 1$, if $(s-1, 1) \in S$, then (44') holds at $(s-1, 1) \Rightarrow A_{s+1\,3} B^*_{s+3\,3} = 0 \Rightarrow A_{s+1\,3} = 0$ or $B_{s+3\,3} = 0$. If $A_{s+1\,3} = 0$, $k = -(s+4)^2 \Rightarrow |B_{s-1\,3}|^2 < 0$, an impossibility. If $B_{s+3\,3} = 0$, $k = -s^2 \Rightarrow |B_{s-1\,3}|^2 < 0$ again. Thus $(s-1, 1) \notin S$. Similarly, $(1, s-1) \notin S$. For the remaining cases, $2 \leqslant t \leqslant s-2$, the fact that $\min(m+n) = s$ requires $A_{s-t\,t} = 0 \Rightarrow k = -s^2 \Rightarrow |B_{s-t+2\,t}|^2 < 0$, an impossibility. Thus $(s-t, t) \notin S$ for $0 < t < s$ as asserted.

Suppose then that $(s, 0) \in S$. $(s-2, 2) \notin S \Rightarrow B_{s2} = 0 \Rightarrow k = -(s-2)^2 B_{s+t\,t+2} = 0$ for all t and A and B nonzero on $T_{s0} \Rightarrow S = T_{s0}$. Finally, because no point (m, n) with m or $n = 1$ occurs in T_{s0}, the generic form with $k = -(s-2)^2$ furnishes a representation thereon. By the $m \longleftrightarrow n$ symmetry, the generic form with $k = -(s-2)^2$ provides a representation on T_{0s} as well.

Case 0–3 exhaust the possible values of $\min(m+n)$.

The only sublattices that arise in the above classification are convex. Hence, by Lemma 1 and 3, any mutiplicity-free unirrep is equivalent to one of the generic representations whose classification we have just concluded. This establishes the Theorem. □

ACKNOWLEDGMENTS

We would like to thank Paul Sally and Birgit Speh for helpful discussions and the relativity group at Chicago for their hospitality.

APPENDIX

We first list the relations satisfied by the operators u and v, given initially by Ogievetskii and Sokachev.[4] We then consider a general multiplicity-free representation of sl_4 for which, as was noted in the text, the symmetric matrices are represented by an operator $\widetilde{Q}_{ap}$ of the form (21); we present the set of equations among the coefficients $A, \widehat{A}, \ldots, E$ that imply the commutation relation (12). It may be of interest that because we are not assuming unitarity here, this set of equations can be used to write the finite dimensional (tensor) representations of sl_4 in the form (21).

In addition to Eqs. (15)–(20) of the text, the operator u satisfies the following relations, where parentheses about indices denote symmetrization and square brackets denote antisymmetrization.

$$u^*_{[a} u_{b]} = 2i(M-1)\epsilon_{ab}{}^c \tilde{J}_c \quad u_{[a} u^*_{b]} = -2i(M+3)\epsilon_{ab}{}^c \tilde{J}_c, \tag{A1}$$

$$u^*_{(a} u_{b)} = M^2 g_{ab} - 4\tilde{J}_{(a}\tilde{J}_{b)},$$
$$u_{(a} u^*_{b)} = (M+2)^2 g_{ab} - 4\tilde{J}_{(a}\tilde{J}_{b)}, \tag{A2}$$

$$\tilde{J}_{[a} u_{b]} = -\frac{i}{4} M \epsilon_{ab}{}^c u_c \quad u_{[a}\tilde{J}_{b]} = \frac{i}{4}(M+4)\epsilon_{ab}{}^c u_c. \tag{A3}$$

The operator v satisfies the parallel set of equations obtained from (A1)–(A3) by the replacements (33); for example (A1) becomes

$$v^*_{[p} v_{q]} = 2i(N-1)\epsilon_{pq}{}^r \widetilde{K}_r \quad v_{[p} v^*_{q]} = -2i(N+3)\epsilon_{pq}{}^r \widetilde{K}_r. \tag{A4}$$

To write the commutation relation (12) in terms of an operator $\widetilde{Q}_{ap}$ of the form (21), it is helpful to use the identity

$$S_a T_b U_p V_q - S_b T_a U_q V_p = 2S_{(a}T_{b)}U_{[p}V_{q]} + 2S_{[a}T_{b]}U_{(p}V_{q)}. \tag{A5}$$

By means of equations (15)–(20) of the text, together with (A1)–(A3) and their counterparts for v, we find

$$\begin{aligned}
&[\widetilde{Q}_{ap}, \widetilde{Q}_{bq}] + i\epsilon_{ab}{}^c \tilde{J}_c g_{pq} + i\epsilon_{pq}{}^r \widetilde{K}_r g_{ab} \\
&\quad = (R_{J\delta} g_{pq} + R_{JKK}\widetilde{K}_{(p}\widetilde{K}_{q)} + R_{Jvv} v_p v_q + v^*_p v^*_q \widehat{R}_{Jvv} + R_{JvK} v_{(p}\widetilde{K}_{q)} + \widetilde{K}_{(p} v^*_{q)} \widehat{R}_{JvK}) i\epsilon_{ab}{}^c \tilde{J}_c \\
&\qquad + (R_{u\delta} g_{pq} + R_{uKK}\widetilde{K}_{(p}\widetilde{K}_{q)} + R_{uvv} v_p v_p + R_{uv^*v^*} v^*_p v^*_q + R_{uvK} v_{(p}\widetilde{K}_{q)} + R_{uv^*K} v^*_{(p}\widetilde{K}_{q)}) i\epsilon_{ab}{}^c u_c \\
&\qquad + i\epsilon_{ab}{}^c u^*_c (g_{pq}\widehat{R}_{u\delta} + \widetilde{K}_{(p}\widetilde{K}_{q)}\widehat{R}_{uKK} + v^*_p v^*_q \widehat{R}_{uvv} + v_p v_q \widehat{R}_{uv^*v^*} + \widetilde{K}_{(p} v^*_{q)} \widehat{R}_{uvK} + \widetilde{K}_{(p} v_{q)} \widehat{R}_{uv^*K}) \\
&\qquad + (\text{above terms with the replacements } a \longleftrightarrow p \quad b \longleftrightarrow q \quad c \longleftrightarrow r \quad J \longleftrightarrow K \quad u \longleftrightarrow v).
\end{aligned} \tag{A6}$$

Here

$$\begin{aligned}
R_{J\delta} = {} & -4(M+3)(N+2)^2 A_{M+2\;N+2}\widehat{A}_{M+2\;N+2} + 4(M-1)N^2 A_{MN}\widehat{A}_{MN} - 4(M+3)N^2 B_{M+2\;N}\widehat{B}_{M+2\;N} \\
& + 4(M-1)(N+2)^2 B_{MN+2}\widehat{B}_{MN+2} + (N+2)^2 C_{MN}\widehat{C}_{MN} + N^2 C_{MN-2}\widehat{C}_{MN-2} + 1,
\end{aligned} \tag{A7}$$

$$\begin{aligned}
R_{JKK} = {} & 16(M+3)A_{M+2\;N+2}\widehat{A}_{M+2\;N+2} - 16(M-1)A_{MN}\widehat{A}_{MN} + 16(M+3)B_{M+2N}\widehat{B}_{M+2N} \\
& - 16(M-1)B_{MN+2}\widehat{B}_{MN+2} - 4C_{MN}\widehat{C}_{MN} - 4C_{MN-2}\widehat{C}_{MN-2} \\
& - 4(M+3)D_{MN}\widehat{D}_{MN} + 4(M-1)D_{M-2N}\widehat{D}_{M-2N} + E_{MN}\widehat{E}_{MN},
\end{aligned} \tag{A8}$$

$$R_{Jvv} = -4(M+3)A_{M+2\;N+2}B_{M+2\;N+4} + 4(M-1)A_{MN+4}B_{MN+2} + C_{MN}C_{MN+2}, \tag{A9}$$

$$\begin{aligned}
R_{JvK} = {} & -4(M+3)A_{M+2\;N+2}\widehat{D}_{MN+2} + 4(M-1)A_{MN+2}\widehat{D}_{M-2N} - 4(M+3)\widehat{B}_{M+2\;N+2}D_{MN} \\
& + 4(M-1)\widehat{B}_{MN+2}D_{M-2\;N+2} + D_{MN}E_{MN} + C_{MN}E_{MN+2},
\end{aligned} \tag{A10}$$

$$\begin{aligned}
R_{u\delta} = {} & \tfrac{1}{2}(M+4)(N+2)^2 A_{M+2\;N+2}\widehat{C}_{M+2N} - \tfrac{1}{2}MN^2 A_{M+2N}\widehat{C}_{MN-2} \\
& - \tfrac{1}{2}M(N+2)^2 \widehat{B}_{M+2\,N+2} C_{MN} + \tfrac{1}{2}(M+4)N^2 \widehat{B}_{M+2N} C_{M+2\;N-2},
\end{aligned} \tag{A11}$$

$$\begin{aligned}
R_{uKK} = {} & -2(M+4)A_{M+2\;N+2}\widehat{C}_{M+2N} + 2MA_{M+2N}\widehat{C}_{MN-2} + 2M\widehat{B}_{M+2\,N+2}C_{MN} \\
& - 2(M+4)\widehat{B}_{M+2N}C_{M+2\;N-2} + \tfrac{1}{2}(M+4)D_{MN}E_{M+2\;N} - \tfrac{1}{2}MD_{MN}E_{MN},
\end{aligned} \tag{A12}$$

$$R_{uvv} = \tfrac{1}{2}(M+4)A_{M+2\;N+2}C_{M+2\;N+2} - \tfrac{1}{2}MA_{M+2\;N+4}C_{MN}, \tag{A13}$$

$$R_{uv^*v^*} = -\tfrac{1}{2}M\widehat{B}_{M+2\,N-2}\widehat{C}_{MN-2} + \tfrac{1}{2}(M+4)\widehat{B}_{M+2N}\widehat{C}_{M+2\,N-4}, \tag{A14}$$

$$R_{uvK} = \tfrac{1}{2}(M+4)A_{M+2\;N+2}E_{M+2\;N+2} - \tfrac{1}{2}MA_{M+2\;N+2}E_{MN} + \tfrac{1}{2}(M+4)C_{M+2N}D_{MN} - \tfrac{1}{2}MC_{MN}D_{MN+2}, \tag{A15}$$

$$R_{uv^*K} = \tfrac{1}{2}(M+4)\widehat{B}_{M+2N}E_{M+2\;N-2} - \tfrac{1}{2}M\widehat{B}_{M+2N}E_{MN} + \tfrac{1}{2}(M+4)\widehat{C}_{M+2\,N-2}D_{MN} - \tfrac{1}{2}M\widehat{C}_{MN-2}D_{MN-2}. \tag{A16}$$

The hatted coefficients, $\widehat{R}_{Jvv}, \widehat{R}_{JuK},...,\widehat{R}_{uv^*K}$, are obtained for the corresponding unhatted expressions by the replacements

$$A \longleftrightarrow \widehat{A}, \quad B \longleftrightarrow \widehat{B}, \quad C \longleftrightarrow \widehat{C}, \quad D \longleftrightarrow \widehat{D}.$$

The operators $A, \widehat{A},..., E$ of a representation with sublattice S satisfy the equations

$$P_{mn}\{[\widetilde{Q}_{ap}\widetilde{Q}_{bq}] + i\epsilon_{ab}{}^c\widetilde{J}_c g_{pq} + i\epsilon_{pq}{}^r\widetilde{K}_r g_{ab}\} = 0, \quad \text{(A17)}$$

where P_{mn} is the projection operator onto any subspace $\mathscr{H}_{mn}$ with $(m, n) \in S$. Equation (A6) is equivalent to requiring that $A, \widehat{A},..., E$ satisfy the following set of equations at each $(m, n) \in S$, with the exclusions listed:

$$R_{J\delta} = 0, \quad m \neq 0, n \neq 1; \quad R_{JKK} = 0, \quad m = 0, n \neq 0,1;$$
$$R_{Jvv} = \widehat{R}_{Jvv} = 0, \quad m \neq 0; \quad \widehat{R}_{JvK} = R_{JvK} = 0, \quad m \neq 0, n \neq 0;$$
$$R_{u\delta} = \widehat{R}_{u\delta} = 0, \quad n \neq 1; \quad R_{uKK} = \widehat{R}_{uKK} = 0, \quad n \neq 0,1;$$
$$R_{uvv} = \widehat{R}_{uvv} = 0; \quad R_{uv^*v^*} = \widehat{R}_{uv^*v^*} = 0, \quad n \neq 0,1,2,3;$$
$$R_{uvK} = \widehat{R}_{uvK} = 0; \quad n \neq 0, \quad R_{uv^*K} = \widehat{R}_{uv^*K} = 0, \quad n \neq 0,1. \quad \text{(A18)}$$

[1]Y. Ne'eman, Proc. Nat. Acad. Sci. USA **74**, 4157 (1977).
[2]Y. Ne'eman and Dj. Sijacki, TAUP 699-78 preprint.
[3]Y. Ne'eman and Dj. Sijacki, Ann. Phys. **120**, 292 (1979).
[4]V.I. Ogievetskii and E. Sokachev, Teor. Mat. Fiz. **23**, 214 (1975).
[5]A. Kihlberg, Ark. Fys. **32**, 241 (1966).
[6]Equation (A4) of Kihlberg[5] implies that for multiplicity-free representations (Kihlberg's "degenerate series") no operator connects $\mathscr{H}_{mn}$ to subspaces other than $\mathscr{H}_{m+2,n+2}$ or $\mathscr{H}_{m+2,n\mp 2}$.

The representations of spl(2,1)—an example of representations of basic superalgebras

M. Marcu
Physikalisches Institut der Universitat Bonn, Bonn, West Germany

(Received 27 September 1979; accepted for publication 21 January 1980)

The example of spl(2,1) illustrates some important properties of the representations of basic superalgebras. Typical representations are most similar to the representations of semisimple Lie algebras. We argue that nontypical representations are a degenerate type of representation. The structure of not fully reducible nontypical representations is discussed in detail. As opposed to semisimple Lie algebras, spl(2,1) also admits representations with nondiagonal Cartan subalgebra. There is a one-to-one correspondence between the representations of spl(2,1) and those of the generalized superalgebra osp(2,0,1,1). The osp(2,0,1,1) counterparts of nontypical irreducible representations of spl(2,1) are unfaithful.

1. INTRODUCTION

A. Representations of basic superalgebras

The structure of finite-dimensional superalgebras is by now well understood.[1-5] Many superalgebras are the semidirect sum of a solvable superalgebra and a semisimple one. Although the algorithm according to which semisimple superalgebras are built out of simple ones is not as simple as in the Lie algebra case, it is known. Simple superalgebras have been classified into classical and Cartan-type superalgebras. Classical superalgebras that admit a nondegenerate metric (supertrace) tensor are called *basic superalgebras*[6] (for the other simple superalgebras, including those of the Cartan type, the metric tensor is zero in every representation). Basic superalgebras resemble semisimple Lie algebras most closely. They admit a canonical form of the commutation relations as follows: α_i, $i = 1,...,r$, is a system of simple positive roots; H_i are the elements of the (r-dimensional) Cartan subalgebra H; X_i are simple positive root vectors ("step-up operators"), Y_i are simple negative root vectors ("step-down operators"), and

$$[H_i,H_j] = 0, \quad [H_i,X_j] = (\alpha_j)_i X_j,$$
$$[H_i,Y_j] = -(\alpha_j)_i Y_j, \quad \langle X_i,Y_j\rangle = H_i\delta_{ij}. \tag{1.1}$$

α_i are vectors in the dual space of H: $\alpha_i \in H^*$. H^* is called root space. All other root vectors are found by repeatedly forming (anti) commutators of the various X_i among temselves (the same with Y_i) and all remaining structure constants are uniquely determined by the Jacobi identity once we have fixed the normalization of the simple positive roots. In particular, each positive root is a linear combination of the simple positive roots, all coefficients being positive integers. Each root vector has a well-defined grade. The corresponding roots are accordingly even or odd. $\langle\ ,\ \rangle$ means commutation or anticommutation, according to the grades of X_i and Y_j. The Cartan subalgebra is entirely even. The matrix $A_{ij} = (\alpha_j)_i$ is called the Cartan matrix. It uniquely determines a basic subalgebra. It has a very useful graphical representation in the form of Dynkin diagrams.

The root space is endowed with a nondegenerate bilinear form

$$g_{ij} = \mathrm{str}(H_i\, H_j), \tag{1.2}$$

which is the restriction of the metric tensor to the Cartan subalgebra. We denote the scalar product determined by g_{ij} by (,). Except for the osp(1,2n) superalgebras, this bilinear form on H^* is indefinite. The roots α obeying $(\alpha,\alpha) = 0$ ("on the light cone") are exactly those odd roots that are not colinear to any even root. The existence of such roots will bring about the existence of a class of representations of basic superalgebras (called nontypical) that have no counterpart for semisimple Lie algebras.

The representations of osp(1,2n) have all the nice properties of representations of semisimple Lie algebras: every irreducible representation has a unique highest weight, all reducible representations are fully reducible (and therefore the Wigner–Eckhart theorem holds), the characters of a representation have almost the same analytic form and play the same role as in the case of a semisimple Lie algebra, etc.

Is there a class of representations of the other basic superalgebras which have similar properties? Kac answered this question.[6] All irreducible representations of basic superalgebras are obtainable from a highest weight. The Schur lemma holds under the usual form. A *typical representation* is an irreducible representation that can be encountered in a reducible representation with diagonal Cartan subalgebra only as a direct summand (it always "splits"). A necessary and sufficient condition for an irreducible representation to be typical is

$$(\Lambda + \rho,\alpha) \neq 0 \quad \text{for all odd roots } \alpha \text{ obeying } (\alpha,\alpha) = 0. \tag{1.3}$$

(Λ is the highest weight and ρ is half the sum of all positive even roots minus half the sum of all positive odd roots.) If a representation is typical, the representation space $V = V_{\bar 0} \oplus V_{\bar 1}$ ($V_{\bar 0}$ is the even subspace and $V_{\bar 1}$ the odd one) obeys

$$\dim V_{\bar 0} = \dim V_{\bar 1}. \tag{1.4}$$

The heighest weight, the dimension, the characters, and the supercharacters are known for each typical representation.[6]

Representations with diagonal Cartan subalgebra that do not obey (1.3) are called *nontypical*. Nontypical represen-

 0022-2488/80/061277-07$1.00

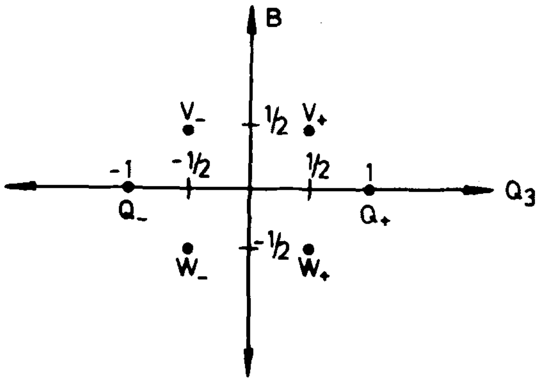

FIG. 1. Root diagram of spl (2, 1).

tations are in many respects degenerate. There is, however, no theory to tell us exactly what this degeneracy means or what are all the nontypical irreducible representations. Reducible nontypical representations can occur as semidirect sums (read: "not fully reducible", "reducible and indecomposable") of several nontypical irreducible representations. Note that as long as the Cartan subalgebra is diagonalizable, typical representations can only combine into direct sums. Of course, direct sums of nontypical representations are also possible.

The even part of a basic superalgebra is either a semisimple or a semisimple $\oplus$ u(1) Lie algebra. In the latter case the u(1) is contained in the Cartan subalgebra. In an arbitrary representation this u(1) may be nondiagonalizable (i.e., it can at best be put under Jordan canonical form). It turns out that in general the supplementary constraints [imposed by the (anti) commutation relations, involving the odd generators] are not strong enough to eliminate the possibility of *not fully reducible representations with nondiagonal Cartan subalgebra*.[7] By Schur's lemma the irreducible constituents will be either identical copies of one typical representation (and thus we exhaust all possible combinations of typical representations up to direct sums) or nontypical representations. Since the nontypical representations are not always fully reducible, all nontypical constitutents need not be identical.

The aim of this paper is to see in an example [spl(2,1)$\approx$osp(2,2)] all these concepts at work. The next paper[8] calculates the Clebsch–Gordan series and the (unnormalized) Clebsch–Gordan coefficients for the tensor product of two irreducible representations of spl(2,1) (both typical and nontypical), thereby getting a better understanding of the statement that nontypical representations are degenerate.

In Sec. 2 we review all irreducible representations of spl(2,1).[9] In Sec. 3 we derive nontypical representations that are semidirect sums of respectively, 2, 3, and 4 irreducible nontypical representations. The knowledge of these particular cases is important for the next paper.[8] We also explain the general mechanism for the computation of all nontypical representations. Throughout Secs. 2 and 3 we take the Cartan subalgebra diagonal.

Section 4 shows (by explicit construction) that there are indeed representations of spl(2,1) with nondiagonal Cartan subalgebra. We find all ways in which irreducible typical representations combine to not fully reducible representations with nondiagonal Cartan subalgebra. Two irreducible nontypical representations cannot enter such a combination. However, this becomes possible if we start with more complicated nontypical representations (themselves not fully reducible).

B. Representations of generalized Lie algebras and superalgebras

In 1978 Rittenberg and Wyler generalized Lie algebras and superalgebras.[10,11] These generalized structures are Γ-graded[12] vector spaces (i.e., the set of grades is an Abelian group Γ) endowed with a graded multiplication law (i.e., a multiplication law consistent with the addition of grades) called generalized commutator and obeying the generalized Jacobi identity. Are all representations also Γ-graded? The answer is negative but all Γ-graded representations are known[12]: for each generalized Lie algebra (superalgebra) there is exactly one Lie algebra (superalgebra) so that there is a one-to-one correspondence between their Γ-graded representations. The problem of finding the non-Γ-graded representations has not been solved.

Along these lines we shall explain the connection between the osp(2,0,1,1) generalized superalgebra[11] and the spl(2,1) superalgebra. This example will also throw some light on the properties of irreducible tensors of superalgebras.

We get all representations of osp(2,0,1,1). We prove that there is a one-to-one correspondence between *all* representations of osp(2,0,1,1) and those of spl(2,1), not only between the Γ-graded ones ($\Gamma = Z_2 \oplus Z_2$ in our case). From this point of view the representations of osp(2,0,1,1) are not a good example for studying the general properties of representations of generalized superalgebras. However, we find a surprising new feature for the irreducible nontypical representations of spl(2,1): their osp(2,0,1,1) counterparts are not faithful. We regard this fact as another aspect of the "degenerate" properties of nontypical representations of spl(2,1).

2. TYPICAL REPRESENTATIONS OF spl(2,1)

The even part of spl(2,1) is a su(2) $\oplus$ u(1) Lie algebra. $Q_\pm$ and Q_3 generate the su(2) ("isospin") and B generates the u(1) ("hypercharge"). The odd part of spl(2,1) contains two isospin $\frac{1}{2}$ tensors of su(2): $V_\pm$ with hypecharge $\frac{1}{2}$ and $W_\pm$ with hypercharge $-\frac{1}{2}$. The commutation relations are:

$$[Q_3,Q_\pm] = \pm Q_\pm, \quad [Q_+,Q_-] = 2Q_3, \quad [B,Q_\pm] = [B,Q_3] = 0, \tag{2.1}$$

$$[Q_3,V_\pm] = \pm\tfrac{1}{2}V_\pm, \quad [Q_3,W_\pm] = \pm\tfrac{1}{2}W_\pm, \quad [Q_\pm,V_\mp] = V_\pm, \quad [Q_\pm,W_\mp] = W_\pm, \tag{2.2}$$

$$[Q_\pm,V_\pm] = 0, \quad [Q_\pm,W_\pm] = 0, \quad [B,V_\pm] = \tfrac{1}{2}V_\pm, \quad [B,W_\pm] = -\tfrac{1}{2}W_\pm,$$

$$\{V_\pm, V_\pm\} = \{V_\pm, V_\mp\} = \{W_\pm, W_\pm\} = \{W_\pm, W_\mp\} = 0, \tag{2.3}$$

$$\{V_\pm, W_\pm\} = \pm Q_\pm, \quad \{V_\pm, W_\mp\} = -Q_3 \pm B.$$

An automorphism of spl (2.1) is:

$$Q_\pm \to Q_\pm, \quad Q_3 \to Q_3, \quad B \to -B, \quad V_\pm \to W_\pm, \quad W_\pm \to V_\pm. \tag{2.4}$$

The commutation relations are already written in canonical form. The root diagram is shown in Fig. 1.

The highest weight ϕ_0 of an irreducible representation is defined by

$$Q_+ \phi_0 = V_+ \phi_0 = W_+ \phi_0 = 0. \tag{2.5}$$

The root diagram and (2.3) tell us that an irreducible representation has at most four su (2) $\oplus$ u (1) multiplets. The actual form of all irreducible representations is known.[9] An irreducible representation is labelled by the eigenvalues of B and Q_3 on ϕ_0 (call them b and q). The basis vectors of the representation space are uniquely specified by the su (2) $\oplus$ u(1) multiplet they belong to and by the eigenvalue of Q_3 (an su (2) $\oplus$ u(1) multiplet is specified by its hypercharge and its isospin), in this order. Then

$$\phi_0 = |b, q, q\rangle \tag{2.6}$$

and the whole basis is given by ($q_3 = q_{3,\max}, q_{3,\max} - 1, \ldots, q_{3,\min}$):

$$|b, q, q_3\rangle, \quad |b + \tfrac{1}{2}, q - \tfrac{1}{2}, q_3\rangle, \quad |b - \tfrac{1}{2}, q - \tfrac{1}{2}, q_3\rangle, \quad |b, q - 1, q_3\rangle. \tag{2.7}$$

The matrix elements of Q_3 and B are obvious; for $Q_\pm$

$$Q_\pm |b, q, q_3\rangle = \sqrt{(q \mp q_3)(q \pm q_3 + 1)}\,|b, q, q_3 \pm 1\rangle. \tag{2.8}$$

It is convenient to write $V_\pm$ and $W_\pm$ in block matrix form. The blocks will be labelled by the su (2) $\oplus$ u (1) multiplets in parentheses. For a nonzero block all matrix elements in a column are zero except for the one in the line $q_3 + \frac{1}{2}$ (for V_+ and W_+) or $q_3 - \frac{1}{2}$ (for V_- and W_-). The expression for these nonzero matrix elements is written explicitly in our formulas. This convention will be used throughout this work unless explicitly stated otherwise. The result is:

$$\begin{array}{cccc} (b,q) & (b+\frac{1}{2},q-\frac{1}{2}) & (b-\frac{1}{2},q-\frac{1}{2}) & (b,q-1) \end{array}$$

$$V_\pm = \begin{pmatrix} 0 & 0 & \epsilon\sqrt{q \pm q_3 + \frac{1}{2}} & 0 \\ \pm\alpha\sqrt{q \mp q_3} & 0 & 0 & \tau\sqrt{q \pm q_3} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \pm\zeta\sqrt{q \mp q_3 - \frac{1}{2}} & 0 \end{pmatrix},$$

$$W_\pm = \begin{pmatrix} 0 & \gamma\sqrt{q \pm q_3 + \frac{1}{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \pm\beta\sqrt{q \mp q_3} & 0 & 0 & \omega\sqrt{q \pm q_3} \\ 0 & \pm\delta\sqrt{q \mp q_3 - \frac{1}{2}} & 0 & 0 \end{pmatrix}. \tag{2.9}$$

This form was obtained using the Wigner-Eckart theorem for su (2) $\oplus$ u (1); $\alpha,\ldots,\omega$ are the reduced matrix elements. Using (2.3) we get ($q \geqslant 1$)

$$\begin{pmatrix} \alpha & \zeta \\ \omega & \gamma \end{pmatrix}\begin{pmatrix} \epsilon & \delta \\ \tau & \beta \end{pmatrix} = 0, \quad \alpha\gamma = \zeta\omega = \frac{q+b}{2q},$$

$$\beta\epsilon = \delta\tau = \frac{q-b}{2q}. \tag{2.10}$$

Note that b can be any complex number. From (1.3) a representation is typical if and only if $b \neq \pm q$. We denote it by $[b, q]$. Equation (2.10) allows for three free constants. They give the relative normalization of the four su (2) $\oplus$ u (1) multiplets. This means that $[b, q]$ is unique, up to equivalence. If $q = \frac{1}{2}$, $b \neq \pm\frac{1}{2}$, the representation obtained is still typical. There are only three su (2) $\oplus$ u (1) multiplets [$(b, q - 1)$ is missing]. In this case Eq. (2.10) reduces to

$$\zeta = \delta = \omega = \tau = 0, \quad \alpha\gamma = \tfrac{1}{2} + b, \quad \beta\epsilon = \tfrac{1}{2} - b. \tag{2.11}$$

We do not make a canonical choice for the reduced matrix elements because no such choice is known for arbitrary b. Introducing a Hermitian scalar product in the representation space, every typical representation is equivalent to a symmetric one:

$$V'_+ = \pm W_-, \quad V'_- = \mp W_+, \quad W'_- = \pm V_+, \quad W'_+ = \mp V_-. \tag{2.12}$$

This is, however, of no great help. Hermitian representations are what we would like to have. On top of Eq. (2.12), $V_\pm$ and $W_\pm$ have to be real in a Hermitian representation. This is possible only for real b obeying $|b| > q$.[9,13]

For a representation uniquely characterized by the Casimir operators we can easily prove that it is a direct summand in every reducible representation,[1] as long as the Cartan subalgebra is diagonal. Typical representations of spl (2,1) are indeed uniquely characterized by their Casimirs.[9] Thus we have explicitly seen that the typical representations of spl (2,1) "split".

3. NONTYPICAL REPRESENTATIONS (IRREDUCIBLE AND NOT FULLY REDUCIBLE ONES) OF spl (2,1)

A. Irreducible nontypical representations

1. $b=q$

Equation (2.10) admits the solution

$$\beta=\delta=\epsilon=\zeta=\tau=\omega=0,\quad \alpha\gamma=1, \tag{3.1}$$

the su (2) $\oplus$ u (1) multiplets $(b-\frac{1}{2}, q-\frac{1}{2})$ and $(b, q-1)$ being absent. We shall denote this irreducible nontypical representation by $[q]_+$.

2. $b=-q$

Equation (2.10) admits the solution

$$\alpha=\gamma=\delta=\zeta=\tau=\omega=0,\quad \beta\epsilon=1. \tag{3.2}$$

$(b+\frac{1}{2}, q-\frac{1}{2})$ and $(b, q-1)$ are absent. Notation: $[q]_-$.

B. Not fully reducible nontypical representations containing two irreducible nontypical representations

This type of representation is obtained from Eq. (2.10), with $b=\pm q$ [all four su (2) $\oplus$ u (1) multiplets are retained].

1. $[q]_+ \,\oplus\!\!\!\!\!\supset\, [q-\frac{1}{2}]_+$

" $\oplus\!\!\!\!\supset$ " means semidirect sum (the representation space of the first summand is an invariant subspace of the whole representation). This representation is characterized by

$$b=q,\quad \beta=\delta=0,\quad \alpha\gamma=\zeta\omega=1,\quad \alpha\epsilon+\zeta\tau=0. \tag{3.3}$$

We denote it by $[q,q-\frac{1}{2}]_+$.

2. $[q-\frac{1}{2}]_+ \,\oplus\!\!\!\!\!\supset\, [q]_+$

This representation is characterized by

$$b=q,\quad \epsilon=\tau=0,\quad \alpha\gamma=\zeta\omega=1,\quad \alpha\delta+\beta\zeta=0. \tag{3.4}$$

Notation: $[q-\frac{1}{2},q]_+$.

3. $[q]_- \,\oplus\!\!\!\!\!\supset\, [q-\frac{1}{2}]_-$

This representation is characterized by

$$b=-q,\quad \alpha=\zeta=0,\quad \beta\epsilon=\delta\tau=1,\quad \beta\gamma+\delta\omega=0. \tag{3.5}$$

Notation: $[q,q-\frac{1}{2}]_-$.

4. $[q-\frac{1}{2}]_- \,\oplus\!\!\!\!\!\supset\, [q]_-$

This representation is characterized by

$$b=-q,\quad \gamma=\omega=0,\quad \beta\epsilon=\delta\tau=1,\quad \beta\zeta+\delta\alpha=0. \tag{3.6}$$

Notation: $[q-\frac{1}{2},q]_-$.

$[q,q-\frac{1}{2}]_+$ and $[q-\frac{1}{2},q]_+$ are representations of different types. The automorphism (2.4) takes $[q,q-\frac{1}{2}]_+$ into $[q,q-\frac{1}{2}]_-$ and $[q-\frac{1}{2},q]_+$ into $[q-\frac{1}{2},q]_-$. If $q=\frac{1}{2}$, we have to use (2.11) instead of (2.10).

There are no other not fully reducible nontypical representations with two irreducible components. We can convince ourselves that this is indeed the case by explicitly trying to construct representations of the form $[q_1] \,\oplus\!\!\!\!\!\supset\, [q_2]$, where $|q_1-q_2|\neq\frac{1}{2}$.

The Casimir operators become identically zero for all the representations discussed in Secs. A. and B., thus no longer characterizing a particular representation. We get the feeling that nontypical representations are a kind of "degeneration" of the typical ones.

Note that for $[q,q-\frac{1}{2}]_\pm$ the highest weight does not correspond to a cyclic vector.

In Secs. C. and D. we are going to use these results as building blocks for more complicated nontypical representations.

C. Not fully reducible nontypical representations containing three irreducible nontypical representations

Equations (3.3)–(3.6) severely limit the nontypical representations that can combine into a new type of representation (i.e., a representation that is not equivalent to a direct sum of representations already described in Secs. A. and B.).

1. $[q-\frac{1}{2}]_\pm \,\oplus\!\!\!\!\!\supset\, [q]_\pm \,\subset\!\!\!\!\!\oplus\, [q+\frac{1}{2}]_\pm$

The representation spaces of $[q+\frac{1}{2}]_\pm$ and $[q-\frac{1}{2}]_\pm$ are invariant. In matrix form:

$$\begin{pmatrix} [q-\frac{1}{2}]_\pm & 0 & M \\ 0 & [q+\frac{1}{2}]_\pm & N \\ 0 & 0 & [q]_\pm \end{pmatrix}. \tag{3.7}$$

The notation should be self evident: The diagonal blocks are the three irreducible nontypical representations. M and N are known from Eqs. (3.3)–(3.6). Notation: $[q-\frac{1}{2},q+\frac{1}{2},q]_+$ The plus and minus cases are related by Eq. (2.4).

2. $[q-\frac{1}{2}]_\pm \,\subset\!\!\!\!\!\oplus\, [q]_\pm \,\oplus\!\!\!\!\!\supset\, [q+\frac{1}{2}]_\pm$

The representation space of $[q]_+$ is invariant. In matrix form:

$$\begin{pmatrix} [q]_\pm & R & S \\ 0 & [q-\frac{1}{2}]_\pm & 0 \\ 0 & 0 & [q+\frac{1}{2}]_\pm \end{pmatrix}. \tag{3.8}$$

R and S are known from Eqs. (3.3)–(3.6). Notation: $[q,q-\frac{1}{2},q+\frac{1}{2}]_\pm$. There is no other type of representation than those described. We might think that we have also to consider cases like $[q]_\pm \,\oplus\!\!\!\!\!\supset\, [q-\frac{1}{2}]_\pm \,\subset\!\!\!\!\!\oplus\, [q]_\pm$. There is, however, an obvious basis transformation that leads to $[q]_\pm \oplus [q,q-\frac{1}{2}]_\pm$.

D. Not fully reducible nontypical representations containing four irreducible nontypical representations

Equations (3.3)–(3.8) severely limit the possible combinations of four irreducible nontypical representations.

1. $[q]_+ \,\oplus\!\!\!\!\!\supset\, [q\pm\frac{1}{2}]_+ \,\subset\!\!\!\!\!\oplus\, [q\pm 1]_+ \,\oplus\!\!\!\!\!\supset\, [q\pm\frac{3}{2}]_+$

The representation space of $[q]_+$ and that of $[q \pm 1]_+$ are invariant. In block matrix form:

$$\begin{pmatrix} [q]_+ & 0 & A & 0 \\ 0 & [q\pm 1]_+ & B & C \\ 0 & 0 & [q \pm \frac{1}{2}]_+ & 0 \\ 0 & 0 & 0 & [q \pm \frac{3}{2}]_+ \end{pmatrix}. \tag{3.9}$$

The blocks $A \neq 0$, $B \neq 0$, and $C \neq 0$ are known from (3.7), (3.8), and (2.3). Notation: $[q, q \pm 1, q \pm \frac{1}{2}, q \pm \frac{3}{2}]_+$. By (2.4) we get the representations $[q, q \pm 1, q \pm \frac{1}{2}, q \pm \frac{3}{2}]_-$.

$$2.\ [q]_+ \begin{array}{c} \nearrow \otimes [q-\frac{1}{2}]_+ \otimes \searrow \\ \searrow \otimes [q+\frac{1}{2}]_+ \otimes \nearrow \end{array} [q]_+$$

This is an entirely new type of representation. Up to now the requirement that the whole nontypical representation should not be equivalent to a direct sum forced all irreducible nontypical components to be different from one another.

By (3.7), (3.8), and by the Wigner–Eckart theorem for su (2) ⊕ u (1), we know all nondiagonal blocks up to some multiplicative constants. These constants are determined using (2.3). Here we give the explicit form of $V_\pm$ and $W_\pm$ with the conventions from Sec. 2:

$$V_\pm = \begin{pmatrix}
0 & 0 & \epsilon\sqrt{q \pm q_3 + \frac{1}{2}} & 0 & 0 & 0 & 0 & 0 \\
\pm\alpha\sqrt{q \mp q_3} & 0 & 0 & \tau\sqrt{q \pm q_3} & 0 & 0 & \pm\rho\sqrt{q \mp q_3} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \pm\zeta\sqrt{q \mp q_3 - \frac{1}{2}} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \eta\sqrt{q \pm q_3 + 1} & 0 \\
0 & 0 & 0 & 0 & \pm\mu\sqrt{q \mp q_3 + \frac{1}{2}} & 0 & 0 & \lambda\sqrt{q \pm q_3 + \frac{1}{2}} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \pm\pi\sqrt{q \mp q_3} & 0
\end{pmatrix},$$

(column blocks: $[q]_+$, $[q-\frac{1}{2}]_+$, $[q+\frac{1}{2}]_+$, $[q]_+$)

$$W_\pm = \begin{pmatrix}
0 & \gamma\sqrt{q \pm q_3 + \frac{1}{2}} & 0 & 0 & \pm\beta\sqrt{q \mp q_3 + \frac{1}{2}} & 0 & 0 & \theta\sqrt{q \pm q_3 + \frac{1}{2}} \\
0 & 0 & 0 & 0 & 0 & \pm\delta\sqrt{q \mp q_3} & 0 & 0 \\
0 & 0 & 0 & \omega\sqrt{q \pm q_3} & 0 & 0 & \pm\kappa\sqrt{q \mp q_3} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \pm\upsilon\sqrt{q \mp q_3 - \frac{1}{2}} \\
0 & 0 & 0 & 0 & 0 & \nu\sqrt{q \pm q_3 + 1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma\sqrt{q \pm q_3 + \frac{1}{2}} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}. \tag{3.10}$$

The constants $\alpha,\ldots,\omega$ obey

$$\begin{aligned}
&\alpha\gamma = \zeta\omega = \mu\nu = \pi\sigma = 1, \\
&\alpha\epsilon + \zeta\tau = 0, \quad \gamma\delta + \nu\beta = 0, \quad \mu\eta + \pi\lambda = 0, \\
&\omega\upsilon + \sigma\kappa = 0, \\
&\pi^2 = \alpha^2, \\
&2q(\gamma\rho + \pi\theta) = \beta\eta, \\
&2q\epsilon\kappa + (2q+1)\beta\eta = 0.
\end{aligned} \tag{3.11}$$

Notation: $[q, q - \frac{1}{2}, q + \frac{1}{2}, q]_+$. By (2.4) we get $[q, q - \frac{1}{2}, q + \frac{1}{2}, q]_-$. Note that $[0, -\frac{1}{2}, \frac{1}{2}, 0]_+ = [0, -\frac{1}{2}, \frac{1}{2}, 0]_-$. Notation: $[0, -\frac{1}{2}, \frac{1}{2}, 0]$.

We had a good reason for explicitly writing down the matrices: This type of representation will play an important role in the study of the decomposition of the tensor product of two irreducible representations of spl (2,1).[8]

The highest weight is a cyclic vector for no not fully reducible nontypical representations except $[q - \frac{1}{2}, q]_\pm$. $[q, q - \frac{1}{2}, q + \frac{1}{2}]_\pm$ and $[q, q \pm 1, q \pm \frac{1}{2}, q \pm \frac{3}{2}]_\pm$ do not even have a cyclic vector.

The task of finding all nontypical representations (with diagonal Cartan subalgebra) is now extremely simple in principle: all difficulties that may appear are known from our examples. Given the diagonal matrix blocks (the irreducible nontypical constituents) we can easily decide whether we can find nondiagonal ones so that the whole representation is not equivalent to a direct sum. On the other hand, the setting up of a list with all nontypical representations is a long and tedious affair.

4. REPRESENTATIONS OF SPL (2,1) WITH NONDIAGONAL CARTAN SUBALGEBRA

Each representation of su (2) ⊕ u (1) is a tensor product of a representation of su (2) and one of u (1). Any matrix is a representation of u (1). It can be brought to Jordan canonical form $B = \Sigma \oplus B_i$, where B_i are "Jordan boxes":

$$B_i = \begin{pmatrix} b_i & 1 & & & \\ 0 & b_i & 1 & & \\ & & \ddots & \ddots & \\ & & & b_i & 1 \\ & & & 0 & b_i \end{pmatrix} \tag{4.1}$$

Consider a typical representation with highest weight ϕ_0 and consider an additional vector ψ_0 so that

$$B\phi_0 = b\phi_0,$$
$$B\psi_0 = b\psi_0 + \phi_0 \qquad (4.2)$$

$$V_+\psi_0 = W_+\psi_0 = Q_+\psi_0 = 0.$$

Acting repeatedly with V_-, W_-, and Q_- on ϕ_0 and ψ_0, we obtain the representation $\phi_0 \in [b,q] \oplus [b,q] \ni \psi_0$. The explicit form of the generators is ($Q_\pm$, Q_3 are obvious):

$$[b,q]\ni\phi_0 \qquad\qquad [b,q]\ni\psi_0$$

$$B = \begin{array}{c} \\ \end{array}\begin{array}{cccccccc} (b,q) & (b+\tfrac{1}{2},q-\tfrac{1}{2}) & (b-\tfrac{1}{2},q-\tfrac{1}{2}) & (b,q-1) & (b,q) & (b+\tfrac{1}{2},q-\tfrac{1}{2}) & (b-\tfrac{1}{2},q-\tfrac{1}{2}) & (b,q-1) \end{array}$$

$$B = \begin{pmatrix} b & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & b+\tfrac{1}{2} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & b-\tfrac{1}{2} & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & b & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & b & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & b+\tfrac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & b-\tfrac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & b \end{pmatrix},$$

$$V_\pm = \begin{pmatrix} 0 & 0 & \epsilon\sqrt{q\pm q_3+\tfrac{1}{2}} & 0 & 0 & 0 & \epsilon'\sqrt{q\pm q_3+\tfrac{1}{2}} & 0 \\ \pm\alpha\sqrt{q\mp q_3} & 0 & 0 & \tau\sqrt{q\pm q_3} & 0 & 0 & 0 & \tau'\sqrt{q\pm q_3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \pm\zeta\sqrt{q\mp q_3-\tfrac{1}{2}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \epsilon\sqrt{q\pm q_3+\tfrac{1}{2}} & 0 \\ 0 & 0 & 0 & 0 & \pm\alpha\sqrt{q\mp q_3} & 0 & 0 & \tau\sqrt{q\pm q_3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \pm\zeta\sqrt{q\mp q_3-\tfrac{1}{2}} & 0 \end{pmatrix},$$

$$W_\pm = \begin{pmatrix} 0 & \gamma\sqrt{q\pm q_3+\tfrac{1}{2}} & 0 & 0 & 0 & \gamma'\sqrt{q\pm q_3+\tfrac{1}{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \pm\beta\sqrt{q\mp q_3} & 0 & 0 & \omega\sqrt{q\pm q_3} & 0 & 0 & 0 & \omega'\sqrt{q\pm q_3} \\ 0 & \pm\delta\sqrt{q\mp q_3-\tfrac{1}{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma\sqrt{q\pm q_3+\tfrac{1}{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \pm\beta\sqrt{q\mp q_3} & 0 & 0 & \omega\sqrt{q\pm q_3} \\ 0 & 0 & 0 & 0 & 0 & \pm\delta\sqrt{q\mp q_3-\tfrac{1}{2}} & 0 & 0 \end{pmatrix} \qquad (4.3)$$

$\alpha,\ldots,\omega$ obey (2.11). The other constants are

$$\gamma' = 1/2q\alpha, \quad \epsilon' = -1/2q\beta,$$
$$\tau' = -1/2q\delta, \quad \omega' = 1/2q\zeta. \qquad (4.4)$$

Up to equivalence transformations there is only one representation $[b,q] \oplus [b,q]$.

The representation $[b,q] \oplus [b,q] \oplus [b,q]$ clearly exists, all the matrix elements being known from (4.3) and (4.4). The Casimir operators are of course nondiagonalizable.

Thus we have finished the description of typical representations and have convinced ourselves of how similar they are to the representations of semisimple Lie algebras.

For $b = \pm q$ Eq. (4.3) can "degenerate" to $\phi_0 \in [q-\tfrac{1}{2}, q]_\pm \oplus [q-\tfrac{1}{2}, q]_\pm \ni \psi_0$. Our knowledge of nontypical representations with nondiagonal Cartan subalgebra is, however, not complete.

5. COMPARISON BETWEEN THE REPRESENTATIONS OF SPL (2,1) AND THOSE OF THE GENERALIZED SUPERALGEBRA SPL (2,0,11)

spl (2, 1) is equivalent to osp (2, 2):

$$\begin{pmatrix} B \\ V_+ \\ V_- \\ W_+ \\ W_- \end{pmatrix}_{\text{osp}(2,2)} = \tfrac{1}{2}\begin{pmatrix} 4iB \\ V_+ + W_+ \\ V_- + W_- \\ -iV_+ + iW_+ \\ -iV_- + iW_- \end{pmatrix}_{\text{spl}(2,1)}. \tag{5.1}$$

(We retain the same notation for the generators in both cases, as no ambiguity is likely to occur.) The commutation relations for osp (2, 2) are:

$$[Q_3,Q_\pm] = \pm Q_\pm, \quad [Q_+,Q_-]_+ = 2Q_3,$$

$$[Q_\pm,B] = [Q_3,B] = 0, \tag{5.2}$$

$$[Q_3,W_\pm] = \tfrac{1}{2}W_\pm, \quad [Q_\pm,W_\pm] = 0, [Q_\pm,W_\mp] = W_\pm, \tag{5.3}$$

$$[Q_3,V_\pm] = \tfrac{1}{2}V_\pm, \quad [Q_\pm,V_\pm] = 0, \quad [Q_\pm,V_\mp] = V_\pm, \tag{5.4}$$

$$[B,V_\pm] = -W_\pm, \quad [B,W_\pm] = V_\pm,$$

$$\{V_\pm,V_\pm\} = \{W_\pm,W_\pm\} = \pm\tfrac{1}{2}Q_\pm,$$

$$\{V_+,V_-\} = \{W_+,W_-\} = -\tfrac{1}{2}Q_3, \tag{5.5}$$

$$\{V_\pm,W_\pm\} = 0, \quad \{V_\pm,W_\mp\} = \pm\tfrac{1}{4}B. \tag{5.6}$$

$Q_\pm$, Q_3, and $V_\pm$ form an osp (1, 2) superalgebra.[1,9] B and $W_\pm$ form an irreducible tensor of this osp (1, 2), transforming according to the three-dimensional "spin-½" representation. B is even and $W_\pm$ are odd [this assignment is determined by the type of brackets occurring in (5.4) and (5.6)]. On the other hand, $Q_\pm$, Q_3, and $W_\pm$ form an osp (1, 2). B and $V_\pm$ form an irreducible tensor, transforming according to the same representation. B is again even and $V_\pm$ are odd. Now let us make some changes in the commutation relations: (5.4) and (5.6) are replaced by

$$\{B,V_\pm\} = W_\pm, \quad \{B,W_\pm\} = V_\pm, \tag{5.4'}$$

$$[V_\pm,W_\pm] = 0, \quad [V_\pm,W_\mp] = \pm\tfrac{1}{4}B. \tag{5.6'}$$

This amounts to interchanging the grades of B and $W_\pm$ [as tensor of osp (1,2)] and of B and $V_\pm$. The algebraic structure thus obtained has a natural $Z_2 \oplus Z_2$ grading:

generators:	$Q_\pm, Q_3$	$V_\pm$	$W_\pm$	B,	
grade:	(0,0)	(1,0)	(0,1)	(1,1).	(5.7)

It is the generalized superalgebra osp (2, 0, 1, 1).[10,11]

We can compute the representations of both osp (2, 2) and osp (2, 0, 1, 1) using the Wigner–Eckart theorem for osp (1, 2). The matrix elements will be the Clebsch–Gordan coefficients of osp (1, 2)[9] multiplied with some constants (the reduced matrix elements). The relevant coefficients differ [for osp (2, 2) and osp (2, 0, 1, 1)] only by some signs. Thus it is very plausible that there exists a one-to-one correspondence between the representations of osp (2, 0, 1, 1) and those of osp (2, 2) [spl (2, 1)]. We have explicitely [i.e., by computing the representations of osp (2, 0, 1, 1)] checked that this statement is true. During this computation an interesting point emerged: Both the Cartan subalgebra (Q_3 and B) and the step operators ($V_+ \pm W_+, V_- \pm W_-$) have no well-defined grade.

The $Z_2 \oplus Z_2$ gradable irreducible representations are the representations $[0, q]$: B has grade (1, 1), so for $b \neq 0$ the relation $B\phi_0 = b\phi_0$ (ϕ_0 is the highest weight) already destroys the $Z_2 \oplus Z_2$ grading. Any $Z_2 \oplus Z_2$ gradable representation has to be symmetric with respect to the $b = 0$ axis of the root space (remember, b is a complex number in general).

Finally, we note that the representations of osp (2, 0, 1, 1) corresponding to the irreducible nontypical representations of spl (2, 1) are not faithful. We do not understand the origin of this fact.

6. CONCLUSIONS

Basic superalgebras are the superalgebras most closely related to (semi)simple Lie algebras. In their representation theory this similarity is reflected by the typical representations. Nontypical representations are still to be studied. In the example of spl (2, 1) we saw that they are in many respects degenerate: The representation space is different from that of typical representations; the Casimirs have zero diagonal; there are many not fully reducible nontypical representations; the osp (2, 0, 1, 1) equivalent of an irreducible nontypical representations of spl (2, 1) is an unfaithful representation. Many basic superalgebras have a nonsemisimple even part. Thus it is not entirely surprising that identical typical representations of spl (2, 1) may combine into a not fully reducible representation with nondiagonal Cartan subalgebra. This is, however, the only type of representation where a typical representation is not a direct summand. By "degeneration" we get nontypical representations with nondiagonal Cartan subalgebra, but we don't know whether we get all of them.

ACKNOWLEDGMENT

It would like to thank Professor V. Rittenberg for his constant encouragement and many helpful discussions. I also thank Dr. M. Scheunert for pointing out to me the existence of representations with nondiagonal Cartan subalgebra and for other helpful discussions.

[1]A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975); **17**, 598 (1976).
[2]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626, 1640 (1976).
[3]V.G. Kac, Commun. Math. Phys. **53**, 31 (1977); Adv. Math. **26**, 8 (1977).
[4]N.B. Backhouse, J. Math. Phys. **18**, 339 (1977).
[5]V. Rittenberg in *Group Theoretical Methods in Physics*, edited by P. Kramer and A. Rieckers (Springer, Berlin, 1978), p. 3.
[6]V.G. Kac, Commun. Algebra **5**, 889 (1977).
[7]M. Scheunert, private communication.
[8]M. Marcu, J. Math. Phys. **21**, 1277 (1980).
[9]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).
[10]V. Rittenberg and D. Wyler, Nucl. Phys. B **139**, 189 (1978).
[11]V. Rittenberg and D. Wyler, J. Math. Phys. **19**, 2193 (1978).
[12]M. Scheunert, J. Math. Phys. **20**, 712 (1979).
[13]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).

The tensor product of two irreducible representations of the spl(2,1) superalgebra

M. Marcu
Physikalisches Institut der Universitat Bonn, Bonn, West Germany

(Received 27 September 1979; accepted for publication 21 January 1980)

The Clebsch–Gordan series and the unnormalized Clebsch–Gordan coefficients for the tensor product of two irreducible representations of spl(2,1) are given. Special emphasis is laid on the discussion of the structure of the series when it contains several nontypical representations. Thereby a better understanding of the "degenerate" properties of nontypical representations is gained.

1. INTRODUCTION

Recently there have been a number of attempts to use the spl(2,1) superalgebra in elementary particle theory.[1] This naturally raises the problem of the tensor product of two irreducible representation of spl(2,1).

There are two types of irreducible representations of spl(2,1): typical and nontypical ones.[2–4] Typical representations are similar to the irreducible representations of semisimple Lie algebras. In particular, they occur in reducible representations solely as direct summands (provided the Cartan subalgebra is diagonal, which is the case throughout the present paper). Nontypical irreducible representations, however, can combine in various ways into not fully reducible nontypical representations.[3]

The existence of not fully reducible representations of spl(2,1) raises the following question: Is it possible to derive a Clebsch–Gordan series? As opposed to the Lie algebra case [note that spl(2,1) is in many respects similar to su(3)] there is no mathematical scheme that tells us if the Clebsch–Gordan series contains direct sums of irreducible nontypical representations or if it contains not fully reducible nontypical representations. Our main result is that we *can always specify what type of nontypical representations occur in the Clebsch–Gordan series.* As far as we know, this is the first time that tensor products of irreducible representations of a superalgebra have ever been computed.[5] We are also able to give a *graphical rule* for the decomposition of the tensor product.

In Sec. 2 we give the Clebsch–Gordan series for the "nongenerate" product of two irreducible representation of spl(2,1), by "nondegenerate" we are referring to a situation where the product decomposes into a direct sum of irreducible representations. Appendix 1 completes this section by explicitly giving the Clebsch–Gordan coefficients. The coefficients are not normalized because we still lack a satisfactory definition of orthogonality for the basis vectors of the representation spaces of superalgebras (this in turn stems from our inability to find a "canonical" form similar to the Hermitian representations of semisimple Lie algebras).[2,3,5]

Typical representations of spl(2,1) bear two labels: a continuous one (call it b; note that b can be a complex number!) and a discrete one (call it q). Nontypical representations are "degenerate cases of typical representations" ($b = \pm q$). In Sec. 3 we examine what happens if we modify the continuous label, such that some of the typical representations in the nondegenerate Clebsch–Gordan series degenerate into nontypical ones. Using our graphical rule, we show that two such situations can occur: Two typical irreducible representations degenerate into four nontypical ones or four typical irreducible representations into eight nontypical ones. As shown in Ref. 3, we can build many kinds of nontypical representations out of four (or eight) irreducible ones. The first idea would be as follows: Take the formulas for the nondegenerate tensor product and change the value of the continuous label; thus you might obtain the Clebsch–Gordan coefficients for the degenerate product and thereby know what kinds of nontypical representations occur. This does not work however, because some basis vectors of the representation spaces of the two (four) typical representations which degenerate into nontypical ones, become linearly dependent. (Note that this is impossible for Lie algebras, where the basis vectors can always be chosen orthogonal). We have to introduce the missing basis vectors "by hand" (i.e., not using the concept of highest weight.) Thus we are finally able to give the Clebsch–Gordan series: it contains one or two not fully reducible nontypical representations of a rather complicated form (with four irreducible nontypical components); their explicit matrix structure is given in Ref. 3. We call such a tensor product a "degenerate product." In Appendix 2 we give the corresponding Clebsch–Gordan coefficients (unnormalized), thereby proving our results.

Throughout this paper the notation for the representations of spl(2,1) will be that of Ref. 3.

2. NONDEGENERATE TENSOR PRODUCT OF TWO IRREDUCIBLE REPRESENTATIONS OF spl(2,1)

The Cartan subalgebra is diagonal for every irreducible representation. It follows that it is also diagonal for a tensor product of two irreducible representations. Thus the tensor product decomposes into a direct sum of typical $\oplus$ a direct or semidirect sum of nontypical irreducible representations.[3] Every irreducible representation is uniquely characterized by its highest weight.[4] We can easily write down the $\mathrm{su}(2) \oplus \mathrm{u}(1)$ content of the product representation. Using the above-mentioned facts it is very simple to determine what irreducible representations of spl(2,1) occur in the product. It is, however, difficult to see whether a nontypical irreducible representation comes in as a direct or as a semidirect

 0022-2488/80/061284-09$1.00

summand (the typical ones are always direct summands[4]). If there is at most one nontypical irreducible representation, the tensor product is nondegenerate, i.e., a direct sum of irreducible representations. In what follows, we give the Clebsch–Gordan series (and a grpahical rule for them) for the product of two typical, a typical with a nontypical and two nontypical irreducible representations.

A. Nondegenerate tensor product of two typical irreducible representations of spl(2,1)

Let $[b_1,q_1]$ and $[b_2,q_2]$ be two typical representations, of spl(2,1) (so $b_1 \neq \pm q_1$, $b_2 \neq \pm q_2$). We assume $q_1 \neq 0$ and $q_2 \neq 0$. Throughout this paper we make the following notations:

$$b = b_1 + b_2, \quad q = q_1 + q_2, \quad q' = |q_1 - q_2|,$$
$$q_< = \min(q_1,q_2) . \tag{2.1}$$

The results are:

(a) $q_1 = q_2 = \frac{1}{2}$; $b \neq 0$; $b \neq \pm 1$:

$$[b_1,q_1] \otimes [b_2,q_2] = [b,1] \oplus [b + \tfrac{1}{2},\tfrac{1}{2}] \oplus [b - \tfrac{1}{2},\tfrac{1}{2}] . \tag{2.2}$$

(b) $q_1 \geqslant 1; q_2 = \frac{1}{2}$(or $q_1 = \frac{1}{2}$; $q_2 \geqslant 1$); $b \neq \pm q$; $b \neq \pm (q-1)$:

$$[b_1,q_1] \otimes [b_2,q_2] = [b_1,q] \oplus [b_1 q - 1] \oplus [b + \tfrac{1}{2},q - \tfrac{1}{2}] \oplus [b - \tfrac{1}{2},q - \tfrac{1}{2}] . \tag{2.3}$$

(c) $q_1 \geqslant 1$; $q_2 \geqslant 1$; $b \neq \pm (q - n)$ for $n = 0,1,\ldots,2q_<$:

$$[b_1,q_1] \otimes [b_2,q_2] = \sum_{n=0}^{2q_<}{}' \oplus [b,q-n] \oplus \sum_{n=1}^{2q_<-1} \oplus [b,q-n] \oplus \sum_{n=0}^{2q_<-1} \oplus [b + \tfrac{1}{2},q - \tfrac{1}{2} - n] \oplus \sum_{n=0}^{2q_<-1} \oplus [b - \tfrac{1}{2},q - \tfrac{1}{2} - n] . \tag{2.4}$$

$\Sigma_0^{2q_<}{}' \oplus [b,q-n]$ means that for $q_1 = q_2$ we sum only from zero to $2q_< - 1$ {the representation $[b,0]$, for $b \neq 0$ as required, simply does not exist}. Note that for the excluded values of b, the tensor product becomes degenerate.

These formulas can be summarized in a graphical rule: Assume (without any loss of generality) that $q_1 \geqslant q_2$. Draw the B and Q_3 axes in the root space[3] (B is a complex axis, but all weights of the product representation have the same imaginary part, namely Imb). Then draw the weight diagram of $[b_2,q_2]$ centered around the highest weight of $[b_1,q_1]$. If $q_1 = q_2$ eliminate the point on the B axis. Write the value of Imb near the diagram. Each resulting point is the highest weight of a typical representation and $[b_1, q_1] \otimes [b_2,q_2]$ is equivalent to the direct sum of all these typical representations. The restrictions on b say that the product is nondegenerate as long as there is no point on the $\pm 45°$ axis and with Im$b = 0$. We shall call such points: "points on the light cone." As an example consider:

$$|2 + i,\tfrac{3}{2}| \otimes [-1 + i,1] = |1 + 2i,\tfrac{5}{2}| \oplus |1 + 2i,\tfrac{3}{2}| \oplus |1 + 2i,\tfrac{3}{2}| \oplus [1 + 2i,\tfrac{1}{2}] \oplus |\tfrac{3}{2} + 2i,2| \oplus |\tfrac{3}{2} + 2i,1| \oplus [\tfrac{1}{2} + 2i,2] \oplus [\tfrac{1}{2} + 2i,1] . \tag{2.5}$$

Graphically, we can refer to Fig. 1.

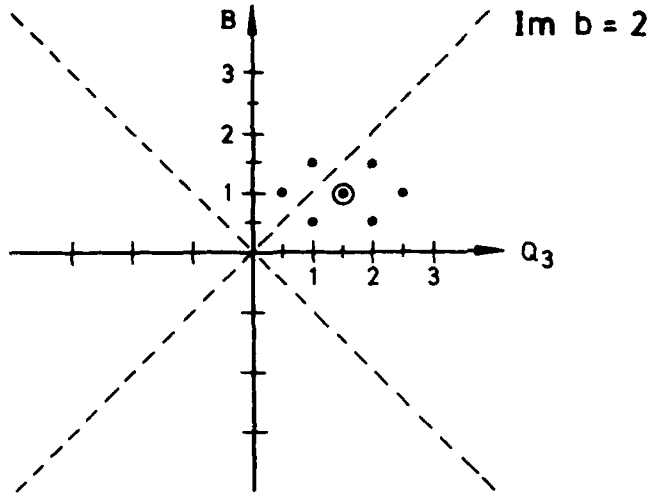

FIG. 1. The Clebsch–Gordan series for $|2 + i,\frac{3}{2}| \otimes [-1 + i,1]$.

B. Nondegenerate tensor product of a typical with a nontypical irreducible representation of spl(2,1)

Let $[b_1,q_1]$ be a typical ($b_1 \neq \pm q_1$) and $[q_2]_\pm$ (see Ref. 3) a nontypical irreducible representation (highest weight: $b_2 = \pm q_2$). The notations (2.1) are maintained, with $b = b_1 \pm q_2$. The Clebsch–Gordan series is:

$$[b_1,q_1] \otimes [q_2]_\pm = \sum_{n=0}^{2q_<-1} \oplus [b,q-n] \oplus \sum_{n=0}^{2q_<-1} \oplus [b \pm \tfrac{1}{2},q - \tfrac{1}{2} - n] \oplus \theta(q_1 - q_2)[b,q'] ,$$

with

$$\theta(q_1 - q_2) = \begin{cases} 1 & \text{if } q_1 > q_2, \\ 0 & \text{if } q_1 \leqslant q_2, \end{cases} \tag{2.6}$$

provided that none of the highest weights of the representations on the left-hand side is on the light cone (i.e., on the $\pm 45°$ axis and with Im$b = 0$).

We can give the following graphical rule: Draw the weight diagram of $[q]_\pm$, centered around the highest weight of $[b_1,q_1]$. Again, all the weights of the product representation will have the same imaginary part, namely Imb. Write Imb near the diagram. If $q_1 = q_2$ eliminate the point on the B axis. If $q_1 < q_2$ eliminate all pairs of points which are symmetrical with respect to the B axis (including the point on the B axis). Each resulting point is the highest weight of a typical representation and the Clebsch–Gordan series is the direct sum of all these typical representations. The condition $b_1 \neq \pm q_1$ implies that the only possibility for the nondegenerate product to degenerate is for the highest weights of two typical representations to get simultaneously on the light cone (thereby four nontypical irreducible representations appear and in Sec. 3 we have to find out whether they are direct or semidirect summands). As a first example consider:

$$[\tfrac{1}{2},1] \otimes [\tfrac{1}{2}]_+ = [1,\tfrac{3}{2}] \oplus [\tfrac{3}{2},1] \oplus [1,\tfrac{1}{2}] . \tag{2.7}$$

This can be graphically represented by Fig. 2. Another example:

$$[i,\tfrac{1}{2}] \otimes [1]_- = [-1 + i,\tfrac{3}{2}] \oplus [-\tfrac{3}{2} + i,1] , \tag{2.8}$$

(See Fig. 3 for a graphical representation. There the "x"

denote the points which have been eliminated after drawing the weight diagram $[1]_-$ centered around the highest weight of $[i,\tfrac{1}{2}]$.)

C. Nondegenerate tensor product of two nontypical irreducible representations of spl(2,1)

(a) Let $[q_1]_\pm$ and $[q_2]_\pm$ be two nontypical irreducible representations. With the notations from (2.1) [here $b = \pm(q_1+q_2) = \pm q$] we obtain

$$[q_1]_\pm \otimes [q_2]_\pm = [q]_\pm \oplus \sum_{n=0}^{2q_<-1} \oplus [\pm(q+\tfrac{1}{2}),\, q-\tfrac{1}{2}-n]\,. \quad (2.9)$$

There is no graphical rule in terms of the weight diagram of one representation centered around the highest weight of the other one. This is another feature that distinguishes nontypical from typical representations. An example:

$$[1]_+ \otimes [\tfrac{3}{2}]_+ = [\tfrac{5}{2}]_+ \oplus [3,2] \oplus [3,1]\,. \quad (2.10)$$

Graphically, one has Fig. 4.

(b) Consider $[q_1]_+$ and $[q_2]_-$ (then $b = q_1 - q_2$). With the notations from (2.1) we obtain:

$$[q_1]_+ \otimes [q_2]_- = [q']_\pm \oplus \sum_{n=0}^{2q_<-1} \oplus [q_1-q_2, q-n]\,. \quad (2.11)$$

The first term on the left-hand side is $[q_1-q_2]_+$ if $q_1>q_2$, $[q_2-q_1]_-$ if $q_1<q_2$ and [0] if $q_1=q_2$. Again, there is no graphical rule similar to that from Secs. 2.A and 2.B. An example:

$$[1]_+ \otimes [\tfrac{3}{2}]_- = [\tfrac{1}{2}]_- \oplus [-\tfrac{1}{2},\tfrac{3}{2}] \oplus [-\tfrac{1}{2},\tfrac{5}{2}]\,. \quad (2.12)$$

Graphically, we have Fig. 5.

We see that in both cases (a) and (b) there is no continuous label left. The product of two nontypical irreducible representations is always nondegenerate.

3. DEGENERATE TENSOR PRODUCT OF TWO IRREDUCIBLE REPRESENTATIONS OF spl(2,1)

In this section we give the results for the degenerate tensor product and explain in principle how they were derived. The actual proof is given in Appendix 2.

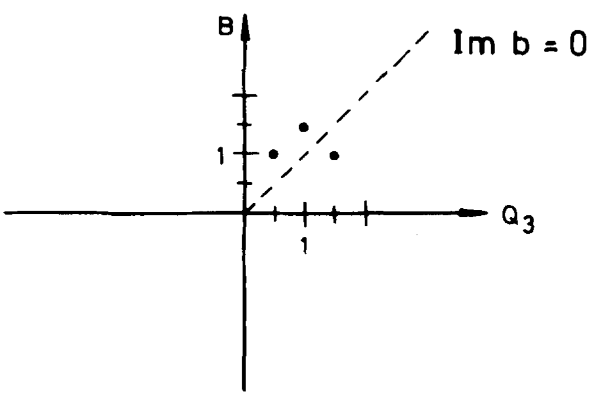

FIG. 2. The Clebsch–Gordan series for $[\tfrac{1}{2},1] \otimes [\tfrac{1}{2}]_+$.

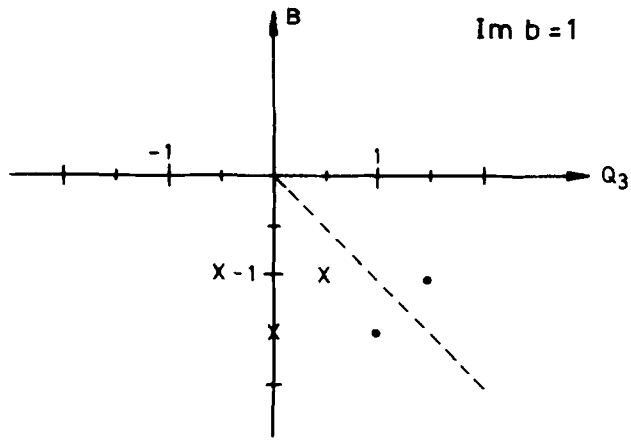

FIG. 3. The Clebsch–Gordan series for $[i,\tfrac{1}{2}] \otimes [1]_-$.

A. Degenerate tensor product of two typical representations

1. $b = \pm q$

The highest weights of $[b,q]$ and $[b \mp \tfrac{1}{2}, q-\tfrac{1}{2}]$ are now the light cone (the $\pm 45°$ line, with real b), so these representations cease to be typical (all the other representations in the Clebsch–Gordan series remain typical). An analysis of the su(2) ⊕ u(1) content of the product representation shows that the series contain the following nontypical irreducible representations: $[q]_\pm$ once, $[q-\tfrac{1}{2}]_\pm$ twice and $[q-1]_\pm$ once. In Appendix 1 we calculate the basis vectors of the typical representations in the decomposition of the nondegenerate tensor product. We take from there the explicit expressions for the $16q-4$ basis vectors of $[b,q] \oplus [b \mp \tfrac{1}{2}, q-\tfrac{1}{2}]$ and replace $b \to \pm q$. As a result, we do not obtain the explicit form of the $16q-4$ dimensional nontypical representation from the decomposition of the degenerate tensor product. What we obtain is the $12q-3$ dimensional representation $[q-\tfrac{1}{2}, q-1, q]_\pm$ $\{ = [q-1]_\pm \oplus [q-\tfrac{1}{2}]_\pm \oplus [q]^3_\pm \}$. Its representation space is invariant because the representation space of $[b,q] \oplus [b \mp \tfrac{1}{2}, q-\tfrac{1}{2}]$ is, trivially, invariant in the nondegenerate case. The second $[q-\tfrac{1}{2}]_\pm$, denoted from now on $[q-\tfrac{1}{2}]'_\pm$, is not here because some basis vectors which were linearly independent for $b \neq \pm q$ become colinear for $b = \pm q$!

In deducing the formulas for the nondegenerate product we have sought and found all vectors with the property that the action of the step-up operators on them gives zero (they were the highest weights of the typical representations in the Clebsch–Gordan series). When $b = \pm q$, these vectors are contained in $[q-\tfrac{1}{2}, q-1, q]_\pm$ and not in $[q-\tfrac{1}{2}]'_\pm$. As a consequence there is at least one step-up operator that takes us from $[q-\tfrac{1}{2}]'_\pm$ to $[q-\tfrac{1}{2}, q-1, q]_\pm$. It follows that the representation space $[q-\tfrac{1}{2}]'_\pm$, is not invariant in any basis!

To sum up, the four nontypical irreducible representations combine into a not fully reducible representation of the form $[q-\tfrac{1}{2}, q-1, q]_\pm \oplus [q-\tfrac{1}{2}]_\pm$. In Ref. 3 we showed that the only representation of this structure is

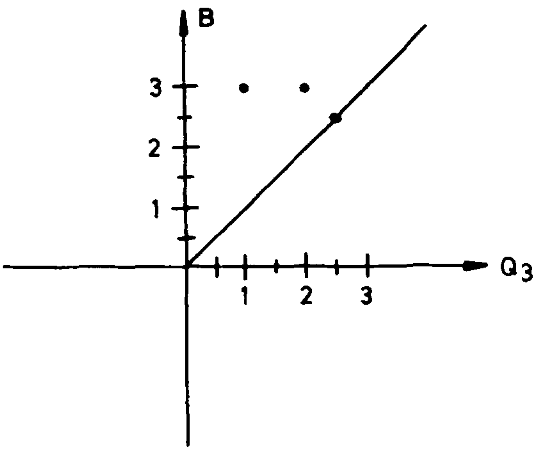

FIG. 4. The Clebsch–Gordan series for $[\frac{1}{2}]_+ \otimes |\frac{3}{2}|_-$.

$[q-\frac{1}{2},q-1,q,q-\frac{1}{2}]_\pm$. Its explicit matrix form is given there. It replaces $[b,q] \oplus [b \mp \frac{1}{2}, q-\frac{1}{2}]$ in (2.2)–(2.4).

2. $b = \pm q' \neq 0$: (so $q_1 \neq q_2$)

Using exactly the same arguments as before, we can show that the Clebsch–Gordan series is given by replacing, in (2.3) and (2.4), $[b,q'] \oplus [b \pm \frac{1}{2}, q' + \frac{1}{2}]$ by $[q',q'-\frac{1}{2},q'+\frac{1}{2},q']_\pm$.

3. $b = q' = 0$: (so $q_1 = q_2$)

By the same arguments, we change (2.2) and (2.4) by replacing $[b+\frac{1}{2},\frac{1}{2}] \oplus [b-\frac{1}{2},\frac{1}{2}]$ with $[0,-\frac{1}{2},\frac{1}{2},0]$ {note that $[0,-\frac{1}{2},\frac{1}{2},0]_+ = [0,-\frac{1}{2},\frac{1}{2},0]_- = [0,-\frac{1}{2},\frac{1}{2},0]$}.

We note that the tensor product $[0,\frac{1}{2}] \otimes [0,\frac{1}{2}]$ was computed in Ref. 2.

A representation of the type $[q,q-\frac{1}{2},q+\frac{1}{2},q]_\pm$ is the most complicated not fully reducible nontypical representation containing $[q]_\pm$ twice, $[q+\frac{1}{2}]_\pm$ and $[q-\frac{1}{2}]_\pm$ (i.e., the only combination these four nontypical irreducible representations in which none of them is a direct summand).

4. $b = \pm(q-n)$ for $n = 1,...,2q_< -1$:

The highest weights of the following four typical representations get simultaneously on the light cone: $[b \pm \frac{1}{2}, q+\frac{1}{2}-n]$ once, $[b,q-n]$ twice, and $[b \mp \frac{1}{2}, q-\frac{1}{2}-n]$ once. From the su(2) ⊕ u(1) content of the product representation, we see that these four typical representations are replaced by eight nontypical irreducible ones: $[q+\frac{1}{2}-n]_\pm$ once, $[q-n]_\pm$ three times, $[q-\frac{1}{2}-n]_\pm$ three times, and $[q-1-n]_\pm$ once. Trying to obtain the same result by changing the value of b $[b \to \pm(q-n)]$ in the formulas for the nondegenerate product we end up with a nontypical representation of smaller dimension: $[q-\frac{1}{2}-n,q-1-n,q-n]_\pm \oplus [q-n,q+\frac{1}{2}-n]_\pm$. Its representation space is invariant in any basis. The representation space of each of the remaining three nontypical irreducible representations is not invariant. These pieces of information, together with the knowledge of all types of nontypical representations with 2, 3, and 4 irreducible components, are sufficient to determine the Clebsch–Gordan series: We have to replace, in (2.4),

$[b \pm \frac{1}{2}, q+\frac{1}{2}-n] \oplus [b,q-n]$
$\oplus [b,q-n] \oplus [b \mp \frac{1}{2},q-\frac{1}{2}-n]$ by
$[q-\frac{1}{2}-n,q-1-n,q-n,$
$q-\frac{1}{2}-n] \oplus [q-n,q-\frac{1}{2}-n,q+\frac{1}{2}-n,q-n]_\pm$.

B. Degenerate tensor product of a typical with a nontypical irreducible representation of spl(2,1)

For simplicity, we shall discuss only the case $[b_1,q_1] \otimes [q_2]_+$. The results for $[b_1,q_1] \otimes [q_2]_-$ are obtained using Eq. (2.4) of Ref. 3.

1. $b = -(q-n)$ for $n = 0,1,...,2q_< -1$

The same thing as in Sec. 3.A happens. As a result, we have to change (2.6) by replacing $[b,q-n] \oplus [b+\frac{1}{2}, q-\frac{1}{2}-n]$ with $[q-\frac{1}{2}-n,q-1-n,q-n,q-\frac{1}{2}-n]_-$.

2. $b = q-n$ for $n = 1,...,2q_<$ ($n = 2q_<$ only for $q_1 > q_2$)

Again, by the same mechanism, we have to change (2.6) by putting $[q-n,q-\frac{1}{2}-n,q+\frac{1}{2}-n,q-n]_+$ instead of $[b,q-n] \oplus [b+\frac{1}{2},q+\frac{1}{2}-n]$.

4. CONCLUSIONS

By now, it is a proven fact that we can write down the Clebsch–Gordan series for the case when the tensor product decomposes into a direct sum of typical representations plus eventually one nontypical irreducible representation[4] (we called this a "nondegenerate tensor product"). We convinced ourselves once again that typical representations are indeed very similar to the representations of semisimple Lie algebras.

The fact that we can write the Clebsch–Gordan series in a closed form even if it contains several irreducible nontypical representations is by all means nontrivial (the "degenerate tensor product"). In order to give the proof, we first had to classify the not fully reducible nontypical representations containing 2, 3, and 4 irreducible nontypical ones and then to do the unpleasant job of explicitly computing the Clebsch–Gordan coefficients.

The way in which we obtained the formulas for the degenerate tensor product, starting from the nondegenerate one, was surprising, because such a phenomenon does not occur in the Lie algebra theory. We believe that this is related to the lack of a viable concept of orthogonality for the basis vectors of the representations of superalgebras.

The not fully reducible nontypical representation occurring in the Clebsch–Gordan series for the degenerate product has the most complicated structure possible. It is

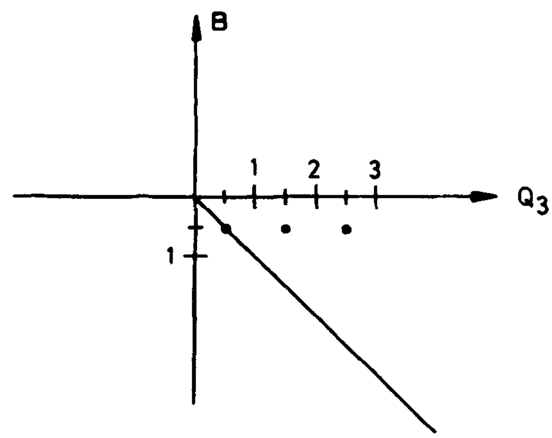

FIG. 5. The Clebsch–Gordan series for $[\frac{1}{2}]_+ \otimes |\frac{3}{2}|_-$.

one representation of the type $[q,q-\frac{1}{2},q+\frac{1}{2},q]_{\pm}$ or a direct sum: $[q-\frac{1}{2},q-1,q,q-\frac{1}{2}]_{\pm} \oplus [q,q-\frac{1}{2},q+\frac{1}{2},q]_{\pm}$ (with the given irreducible nontypical representation content all other alternatives are direct sums of smaller representation).

The even part of spl(2,1) is not semisimple. Representations of nonsemisimple Lie algebras bear both discrete and continuous labels. The (anti)commutation relations involving the odd generators are not strong enough to provide discrete labels for all representations. There is, however, a class of superalgebras with semisimple even part. The labels of their representations are accordingly discrete, but they still have nontypical representations [except osp$(1,2n)$]. It would be interesting to study the properties of these representations and compare them to those of the nontypical representations of spl(2,1).

ACKNOWLEDGMENT

I thank Professor V. Rittenberg for suggesting I investigate this problem and for many interesting discussions.

APPENDIX 1: EXPLICIT DECOMPOSITION OF THE NONDEGENERATE TENSOR PRODUCT OF TWO IRREDUCIBLE REPRESENTATIONS OF spl(2,1)

First we are going to do the decomposition of the product of two typical representations: $[b_1,q_1] \otimes [b_2,q_2]$. Each of them has four su(2) $\otimes$ u(1) multiplets (for $q_1 \geqslant 1$, $q_2 \geqslant 1$), namely:

$$(b_i,q_i), \quad (b_i+\tfrac{1}{2},q_i-\tfrac{1}{2}),$$
$$(b_i-\tfrac{1}{2},q_i-\tfrac{1}{2}), \quad (b_i,q_i-1), \quad i=1,2. \tag{A1.1}$$

The basis vectors of $[b_i,q_i]$ are:

$$|b_i,q_i,m_i\rangle\,, \quad |b_i+\tfrac{1}{2},q_i-\tfrac{1}{2},m_i\rangle\,,$$
$$|b_i-\tfrac{1}{2},q_i-\tfrac{1}{2},m_i\rangle\,, \quad |b_i,q_i-1,m_i\rangle\,, \tag{A1.2}$$

m_i being the third component of isospin. $|b_1,q_1,q_1\rangle$ has the grade λ_1 and $|b_2,q_2,q_2\rangle$ the grade λ_2. For $q_i=\frac{1}{2}$, the multiplet (b_i,q_i-1) is absent ($i=1$ or $i=2$ or both).

The action of the even generators of spl(2,1) is obvious and the action of the odd generators is given by Eq. (3.20) of Ref. 2 [or (2.9) of Ref. 3]. The definition of the tensor product for two representations of a superalgebra is also given in Ref. 2.

A typical representation of spl(2.1) is characterized by 8 constants, 3 of which are independent [see Eqs. (2.9) – (2.11) of Ref. 3]: $\alpha_i, \beta_i, \gamma_i, \delta_i, \epsilon_i, \zeta_i, \tau_i$, and ω_i ($i=1,2$). We do not make a canonical choice for these constants because we don't know how to generalize the concept of Hermitian representation to superalgebras [for spl(2.1) this has been done only for typical representations with b real and $|b|>q$].[2,3]

We do not have to make an extra computation for $q_i=\frac{1}{2}$ ($i=1$ or $i=2$ or both). In this case we simply take the formulas for $q_1 \geqslant 1$, $q_2 \geqslant 1$ (which we are going to present), eliminate all terms containing $|b_i, q_i-1, m_i\rangle$ and put $\delta_i = \zeta_i, = \tau_i, = \omega_i, = 0$.

Let us now consider all su(2) $\oplus$ u(1) tensor products of pairs of su(2) $\oplus$ u(1) multiplets (for $q_1 \geqslant 1$, $q_2 \geqslant 1$ there are 16 different pairs). To do this we use the su(2) Clebsch–Gordan coefficients (with the notation from Goldberger and Watson). The resulting vectors will be labeled by the labels of the two multiplets we started with and the eigenvalues of Q^2 and Q_3 [B and Q_3 have additive eigenvalues and the usual su(2) convention is used for the eigenvalues of Q^2]. For example:

$$|(b_1,q_1),(b_2,q_2),b,q-n,m_1+m_2\rangle = \sum_{m_1,m_2} C^{q-n,m_1+m_2}_{q_1m_1q_2m_2} |b_1,q_1,m_1\rangle \otimes |b_2,q_2,m_2\rangle\,. \tag{A1.3}$$

By analyzing the quantum numbers of the resulting su(2) $\oplus$ u(1) multiplets we can immediately derive (2.2)–(2.4).

The basis vectors of a typical representation on the right-hand side of (2.2)–(2.4) will be labeled by the corresponding representation label {e.g.,$[b,q]$} and the eigenvalues of B, Q^2, and Q_3. These vectors will be linear combinations of vectors of the type (A1.3), having the same eigenvalues for B, Q^2, and Q_3. In order to find them, we first look for the highest weight vectors of each irreducible representation from the Clebsch–Gordan series. Thus the first step is to find all vectors which obey the condition that the action of V_+ and W_+ on them gives zero (everything concerning the action of the even generators is by now settled). After having obtained all highest weight vectors we act on them with V_- and W_- in order to obtain the other basis vectors, exactly as in Eq. (3.20) of Ref. 2. We shall not attempt to normalize the basis vectors because, as said before, we don't have a useful definition of orthogonality. Instead, we shall give the values of the constants $\alpha,\ldots,\omega$ for each irreducible representation [again, these constants obey Eqs. (2.10) and (2.11) of Ref. 3].

Now we give the results of our computation.

1. The representations $[b+\frac{1}{2}, q-\frac{1}{2}-n]$ $(n=0,\ldots,2q_<-1)$

$$\begin{aligned}
|[b+\tfrac{1}{2},q-\tfrac{1}{2}-n],b+\tfrac{1}{2},q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle
&= \gamma_2\sqrt{(2q-n)(2q_2-n)}\,|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle \\
&\quad + (-1)^{\lambda_1+1}\gamma_1\sqrt{(2q-n)(2q_1-n)}\,|(b_1,q_1),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle \\
&\quad + (-1)^{\lambda_1}\delta_1\sqrt{n(2q_2-n)}\,|(b_1,q_1-1),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle \\
&\quad + \delta_2\sqrt{n(2q_1-n)}\,|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2-1),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle\,.
\end{aligned}$$

$$|[b+\tfrac{1}{2},q-\tfrac{1}{2}-n],b+1,q-1-n,q-1-n\rangle = |(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1-n,q-1-n\rangle\,.$$

$$
\begin{aligned}
|[b+\tfrac{1}{2},q-\tfrac{1}{2}-n],b,q-1-n,q-1-n\rangle =\ & \gamma_1\gamma_2\sqrt{(2q-n)(n+1)}\,|(b_1,q_1),(b_2,q_2),q-1-n,q-1-n\rangle \\
& +\delta_1\gamma_2\sqrt{(2q_2-n)(2q_1-1-n)}\,|(b_1,q_1-1),(b_2,q_2),q-1-n,q-1-n\rangle \\
& -\gamma_1\delta_2\sqrt{(2q_1-n)(2q_2-1-n)}\,|(b_1,q_1),(b_2,q_2-1),q-1-n,q-1-n\rangle \\
& +\delta_1\delta_2\sqrt{n(2q-1-n)}\,|(b_1,q_1-1),(b_2,q_2-1),q-1-n,q-1-n\rangle \\
& +(-1)^{\lambda_1+1}2q_2\beta_2\gamma_2|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1-n,q-1-n\rangle \\
& +(-1)^{\lambda_1+1}2q_1\beta_1\gamma_1|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1-n,q-1-n\rangle\,.
\end{aligned}
$$

$$
\begin{aligned}
|[b+\tfrac{1}{2},&q-\tfrac{1}{2}-n],b+\tfrac{1}{2},q-\tfrac{3}{2}-n,q-\tfrac{3}{2}-n\rangle \\
&=\gamma_2\sqrt{(n+1)(2q_1-1-n)}\,|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2),q-\tfrac{3}{2}-n,q-\tfrac{3}{2}-n\rangle \\
&\quad+(-1)^{\lambda_1}\gamma_1\sqrt{(n+1)(2q_2-1-n)}\,|(b_1,q_1),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{3}{2}-n,q-\tfrac{3}{2}-n\rangle \\
&\quad+(-1)^{\lambda_1}\delta_1\sqrt{(2q-1-n)(2q_1-1-n)}\,|(b_1,q_1-1),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{3}{2}-n,q-\tfrac{3}{2}-n\rangle \\
&\quad-\delta_2\sqrt{(2q-1-n)(2q_2-1-n)}\,|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2-1),q-\tfrac{3}{2}-n,q-\tfrac{3}{2}-n\rangle\,.
\end{aligned}
$$

$$
\alpha=(-1)^{\lambda_1+1}(b+q-n)\left(\frac{q-n}{q-\frac{1}{2}-n}\right)^{1/2},\quad \beta=\left(\frac{q-n}{q-\frac{1}{2}-n}\right)^{1/2},\quad \delta=(-1)^{\lambda_1}[2(q-1-n)(2q-1-2n)]^{-1/2}\,. \tag{A1.4}
$$

2. The representations $[b-\frac{1}{2},q-\frac{1}{2}-n]$ $(n=0,...,2q_<-1)$

$$
\begin{aligned}
|[b-\tfrac{1}{2},&q-\tfrac{1}{2}-n],b-\tfrac{1}{2},q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle \\
&=\epsilon_2\sqrt{(2q-n)(2q_2-n)}\,|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle \\
&\quad+(-1)^{\lambda_1+1}\epsilon_1\sqrt{(2q-n)(2q_1-n)}\,|b_1,q_1),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle \\
&\quad+(-1)^{\lambda_1}\zeta_1\sqrt{n(2q_2-n)}\,|(b_1,q_1-1),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle \\
&\quad+\zeta_2\sqrt{n(2q_1-n)}\,|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2-1),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle\,.
\end{aligned}
$$

$$
|[b-\tfrac{1}{2},q-\tfrac{1}{2}-n],b-1,q-1-n,q-1-n\rangle=|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1-n,q-1-n\rangle\,.
$$

$$
\begin{aligned}
|[b-\tfrac{1}{2},&q-\tfrac{1}{2}-n],b,q-1-n,q-1-n\rangle \\
&=\epsilon_1\epsilon_2\sqrt{(2q-n)(n+1)}\,|(b_1,q_1),(b_2,q_2),q-1-n,q-1-n\rangle \\
&\quad+\zeta_1\epsilon_2\sqrt{(2q_2-n)(2q_1-1-n)}\,|(b_1,q_1-1),(b_2,q_2),q-1-n,q-1-n\rangle \\
&\quad-\epsilon_1\zeta_2\sqrt{(2q_1-n)(2q_2-1-n)}\,|(b_1,q_1),(b_2,q_2-1),q-1-n,q-1-n\rangle \\
&\quad+\zeta_1\zeta_2\sqrt{n(2q-1-n)}\,|(b_1,q_1-1),(b_2,q_2-1),q-1-n,q-1-n\rangle \\
&\quad+(-1)^{\lambda_1+1}2q_1\alpha_1\epsilon_1|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1-n,q-1-n\rangle \\
&\quad+(-1)^{\lambda_1+1}2q_2\alpha_2\epsilon_2|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1-n,q-1-n\rangle\,.
\end{aligned}
$$

$$
\begin{aligned}
|[b-\tfrac{1}{2},&q-\tfrac{1}{2}-n],b-\tfrac{1}{2},q-\tfrac{3}{2}-n,q-\tfrac{3}{2}-n\rangle \\
&=\epsilon_2\sqrt{(n+1)(2q_1-1-n)}\,|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2),q-\tfrac{3}{2}-n,q-\tfrac{3}{2}-n\rangle \\
&\quad+(-1)^{\lambda_1}\epsilon_1\sqrt{(n+1)(2q_2-1-n)}\,|(b_1,q_1),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{3}{2}-n,q-\tfrac{3}{2}-n\rangle \\
&\quad+(-1)^{\lambda_1}\zeta_1\sqrt{(2q-1-n)(2q_1-1-n)}\,|(b_1,q_1-1),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{3}{2}-n,q-\tfrac{3}{2}-n\rangle \\
&\quad-\zeta_2\sqrt{(2q-1-n)(2q_2-1-n)}\,|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2-1),q-\tfrac{3}{2}-n,q-\tfrac{3}{2}-n\rangle\,.
\end{aligned}
$$

$$
\alpha=\left(\frac{q-n}{q-\frac{1}{2}-n}\right)^{1/2},\quad \beta=(-1)^{\lambda_1+1}(-b+q-n)\left(\frac{q-n}{q-\frac{1}{2}-n}\right)^{1/2},\quad \zeta=(-1)^{\lambda_1}[2(q-1-n)(2q-1-2n)]^{-1/2}\,. \tag{A1.5}
$$

3. The representations $[b-n,q-n]$ (for $n=0$ and $n=2q_<$ the multiplicity is one and for $n=1,...,2q_<-1$ it is two)

For $n=1,...,2q_<-1$ we could not find a "canonical" way to specify the two equivalent representations $[b-n,q-n]$. So we give the basis vector in terms of two additional constants M_n and N_n. Putting, e.g., $M_n=0$, $N_n=1$, we obtain one $[b-n,q-n]$ and then putting $N_n=0$, $M_n=1$ we obtain the other $[b-n,q-n]$. There is, however, an infinity of choices. We obtain:

$$
\begin{aligned}
|[b,&q-n],b,q-n,q-n\rangle \\
&=a_1|(b_1,q_1),(b_2,q_2),q-n,q-n\rangle+a_2|(b_1,q_1-1),(b_2,q_2),q-n,q-n\rangle \\
&\quad+a_3|(b_1,q_1),(b_2,q_2-1),q-n,q-n\rangle \\
&\quad+a_4|(b_1,q_1-1),b_2,q_2-1),q-n,q-n\rangle+a_5|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-n,q-n\rangle \\
&\quad+a_6|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-n,q-n\rangle
\end{aligned}
$$

for $n=0$:

$$a_1=1,\quad a_2=a_3=a_4=a_5=a_6=0;$$

$n=1,...,2q_<-1$:

$$a_1=-\left(\frac{2q-n+1}{n}\right)^{1/2}(\gamma_1\epsilon_2 M_n+\epsilon_1\gamma_2N_n),\quad a_2=\left(\frac{2q_2-n+1}{2q_1-n}\right)^{1/2}(\delta_1\epsilon_2M_n+\zeta_1\gamma_2N_n),$$

$$a_3=-\left(\frac{2q_1-n+1}{2q_2-n}\right)^{1/2}(\gamma_1\zeta_2M_n+\epsilon_1\delta_2N_n),\quad a_4=\left(\frac{n-1}{2q-n}\right)^{1/2}(\delta_1\zeta_2M_n+\delta_2\zeta_1N_n),$$

$$a_5=(-1)^{\lambda_1}M_n,\quad a_6=(-1)^{\lambda_1}N_n;$$

$n=2q_<;q_1>q_2$:

$$a_1=\left(\frac{2q_1+1}{2q_2}\right)^{1/2}\epsilon_1\omega_1,\quad a_2=-\tfrac{1}{2}[2(q_1-q_2)]^{-1/2}\left(\frac{b_1}{q_1}+\frac{b_2}{q_2}\right),\quad a_3=0,$$

$$a_4=-\left(\frac{2q_2-1}{2q_1}\right)^{1/2}\alpha_2\delta_2,\quad a_5=(-1)^{\lambda_1}\tau_1\beta_2,\quad a_6=(-1)^{\lambda_1+1}\omega_1\alpha_2;$$

$n=2q_<;q_1<q_2$:

$$a_1=-\left(\frac{2q_2+1}{2q_1}\right)^{1/2}\epsilon_2\omega_2,\quad a_2=0,\quad a_3=-\tfrac{1}{2}[2(q_2-q_1)]^{-1/2}\left(\frac{b_1}{q_1}+\frac{b_2}{q_2}\right),$$

$$a_4=\left(\frac{2q_1-1}{2q_2}\right)^{1/2}\alpha_1\delta_1,\quad a_5=(-1)^{\lambda_1}\alpha_1\omega_2,\quad a_6=(-1)^{\lambda_1+1}\beta_1\tau_2;$$

$n=2q_<;q_1=q_2$:

$$a_1=a_2=a_3=a_4=a_5=a_6=0.$$

$$\begin{aligned}|[b,q-n],b+\tfrac{1}{2},q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle &= c_1|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle\\ &+c_2|(b_1,q_1),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle\\ &+c_3|(b_1,q_1-1),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle\\ &+c_4|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2-1),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle;\end{aligned}$$

for $n=0$:

$$c_1=\alpha_1\left(\frac{2q_1}{2q+1}\right)^{1/2},\quad c_2=(-1)^{\lambda_1}\alpha_2\left(\frac{2q_2}{2q+1}\right)^{1/2},\quad c_3=c_4=0;$$

$n=1,...,2q_<1$:

$$c_1=-[n(2q_1-n)]^{-1/2}[\epsilon_2(q_1+b_1-n)M_n+2q_1\alpha_1\epsilon_1\gamma_2N_n],$$
$$c_2=(-1)^{\lambda_1+1}[n(2q_2-n)]^{-1/2}[2q_2\gamma_1\alpha_2\epsilon_2M_n+\epsilon_1(q_2+b_2-n)N_n],$$
$$c_3=(-1)^{\lambda_1}[(2q-n)(2q_1-n)]^{-1/2}[2q_2\delta_1\alpha_2\epsilon_2M_n+\zeta_1(q_2+b_2-n+2q_1)N_n],$$
$$c_4=-[(2q-n)(2q_2-n)]^{-1/2}[\zeta_2(q_1+b_1-n+2q_2)M_n+2q_1\alpha_1\epsilon_1\delta_2N_n];$$

$n=2q_<;q_1>q_2$:

$$c_1=-\tfrac{1}{2}\tau_1(b+q_1-q_2)[q_2(q_1-q_2)]^{-1/2};\quad c_2=0,$$
$$c_3=(-1)^{\lambda_1+1}\tfrac{1}{2}\alpha_2(b+q_1-q_2)[q_1(q_1-q_2)]^{-1/2};\quad c_4=0;$$

$n=2q_<;q_2>q_1$:

$$c_1=0;\quad c_2=(-1)^{\lambda_1}\tfrac{1}{2}\tau_2(b+q_2-q_1)[q_1(q_2-q_1)]^{-1/2},$$
$$c_3=0;\quad c_4=-\tfrac{1}{2}\alpha_1(b+q_2-q_1)[q_2(q_2-q_1)]^{-1/2}.$$

$$\begin{aligned}|[b,q-n],b-\tfrac{1}{2},q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle &= d_1|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle\\ &+d_2|(b_1,q_1),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle\\ &+d_3|(b_1,q_1-1),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle\\ &+d_4|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2-1),q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle;\end{aligned}$$

for $n=0$:

$$d_1=\beta_1\left(\frac{2q_1}{2q+1}\right)^{1/2};\quad d_2=(-1)^{\lambda_1}\beta_2\left(\frac{2q_2}{2q+1}\right)^{1/2};\quad d_3=d_4=0;$$

$n=1,..,2q_<-1$:

$$d_1=-[n(2q_1-n)]^{-1/2}[2q_1\beta_1\gamma_1\epsilon_2M_n+\gamma_2(q_1-b_1-n)N_n],$$

$$d_2 = (-1)^{\lambda_1+1}[n(2q_2-n)]^{-1/2}[\gamma_1(q_2-b_2-n)M_n + 2q_2\epsilon_1\beta_2\gamma_2N_n],$$
$$d_3 = (-1)^{\lambda_1}[(2q-n)(2q_1-n)]^{-1/2}[\delta_1(q_2-b_2-n+2q_1)M_n + 2q_2\zeta_1\beta_2\gamma_2N_n].$$
$$d_4 = -[(2q-n)(2q_2-n)]^{-1/2}[2q_1\beta_1\gamma_1\zeta_2M_n + \delta_2(q_1-b_1-n+2q_2)N_n],$$

$n=2q_<;q_1>q_2$:

$$d_1 = \tfrac{1}{2}\omega_1(-b+q_1-q_2)[q_2(q_1-q_2)]^{-1/2};$$
$$d_2 = 0,$$
$$d_3 = (-1)^{\lambda_1}\tfrac{1}{2}\beta_2(-b+q_1-q_2)[q_1(q_1-q_2)]^{-1/2};$$
$$d_4 = 0;$$

$n=2q_<;q_1<q_2$:

$$d_1 = 0$$
$$d_2 = (-1)^{\lambda_1+1}\tfrac{1}{2}\omega_2(-b+q_2-q_1)[q_1(q_2-q_1)]^{-1/2},$$
$$d_3 = 0,$$
$$d_4 = \tfrac{1}{2}\beta_1(-b+q_2-q_1)[q_2(q_2-q_1)]^{-1/2}q.$$

$$\begin{aligned}|[b,q-n],b,q-1-n,q-1-n\rangle = {} & l_1|(b_1,q_1),(b_2,q_2),q-1-n,q-1-n\rangle\\ &+ l_2|(b_1,q_1-1),(b_2,q_2),q-1-n,q-1-n\rangle\\ &+ l_3|(b_1,q_1),(b_2,q_2-1),q-1-n,q-1-n\rangle\\ &+ l_4|(b_1,q_1-1),(b_2,q_2-1),q-1-n,q-1-n\rangle\\ &+ l_5|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1-n,q-1-n\rangle\\ &+ l_6|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1-n,q-1-n\rangle;\end{aligned}$$

$$l_1 = \epsilon_1d_1[(n+1)(2q_2-n)]^{1/2} + (-1)^{\lambda_1+1}\epsilon_2d_2[(n+1)(2q_1-n)]^{1/2},$$
$$l_2 = \zeta_1d_1[(2q-n)(2q_1-1-n)]^{1/2} + (-1)^{\lambda_1+1}\epsilon_2d_3[n(2q_1-1-n)]^{1/2},$$
$$l_3 = (-1)^{\lambda_1}\zeta_2d_2[(2q-n)(2q_2-1-n)]^{1/2} + \epsilon_1d_4[n(2q_2-1-n)]^{1/2},$$
$$l_4 = (-1)^{\lambda_1}\zeta_2d_3[(2q-1-n)(2q_2-n)]^{1/2} + \zeta_1d_4[(2q-1-n)(2q_1-n)]^{1/2},$$
$$l_5 = \alpha_1d_2[(2q-n)(2q_1-n)]^{1/2} + \tau_1d_3[n(2q_2-n)]^{1/2},$$
$$l_6 = (-1)^{\lambda_1+1}\alpha_2d_1[(2q-n)(2q_2-n)]^{1/2} + (-1)^{\lambda_1}\tau_2d_4[n(2q_1-n)]^{1/2};$$
$$\alpha = \left(\frac{q+\frac{1}{2}-n}{q-n}\right)^{1/2} = \beta,\quad \delta = -\tfrac{1}{2}[(q-\tfrac{1}{2}-n)(q-n)]^{-1/2}. \tag{A1.6}$$

The complete su(2) multiples are obtained by repeatedly acting with Q_- on the vectors written here.

If somebody will ever find a canonical form for the representations of spl(2,1), he can still use our formulas. All he has to do is to shift some multiplicative constants from the basis vectors to $\alpha,...,\omega$.

Let us now consider the case $[b_1,q_1]\otimes[q_2]_\pm$. The explicit decomposition can be obtained from the same formulas as above, with $|\pm(q_2-\frac{1}{2}),q_2-\frac{1}{2}\rangle$ and $|\pm q_2,q_2-1\rangle$ missing and $\alpha_2\gamma_2=1, \beta_2=\delta_2=\epsilon_2=\zeta_2=\tau_2=\omega_2=0$ {for $[q_2]_+$} or $\beta_2\epsilon_2=1, \alpha_2=\gamma_2=\delta_2=\zeta_2=\tau_2=\omega_2=0$ {for $[q_2]_-$}.

Doing the same thing for $[b_1,q_1]\rightarrow[q_1]_\pm$, we can also particularize our formulas for $[q_1]_\pm\otimes[q_2]_\pm$ and $[q_1]_+\otimes[q_2]_-$.

APPENDIX 2: DECOMPOSITION OF THE DEGENERATE TENSOR PRODUCT OF TWO IRREDUCIBLE REPRESENTATIONS OF spl(2,1)

We shall treat the cases discussed in Secs. 3.A.1 and 3.A.4 in order to explain how the degeneration occurs. The fomulas for all other cases can be derived along the same line, starting from the formulas in Appendix 1. As a matter of fact we explicitly did the calculations for all cases, but nothing new can be learned from the ones omitted here. The explicit matrix elements of the representations of the type $[q,q-\frac{1}{2},q+\frac{1}{2},q]_\pm$ which we are going to encounter are in agreement with Ref. 3, so we won't write them once again here.

1. $[b_1,q_1]\otimes[b_2,q_2]$ for $b=q$

The following pairs of vectors [and with them their whole su(2) multiplets] become colinear:

$$|[b,q],b-\tfrac{1}{2},q-\tfrac{1}{2},q-\tfrac{1}{2}\rangle \text{ with } |[b-\tfrac{1}{2},q-\tfrac{1}{2}],b-\tfrac{1}{2},q-\tfrac{1}{2},q-\tfrac{1}{2}\rangle,\quad |[b,q],b,q-1,q-1\rangle, \text{ with } |[b-\tfrac{1}{2},q-\tfrac{1}{2}],b,q-1,q-1\rangle. \tag{A2.1}$$

Using (A1.5) and (A1.6) we obtain the basis vectors for $[q-\frac{1}{2},q-1,q,q-\frac{1}{2}]_+$ {$[q-\frac{1}{2}]'_+$ is as in Sec. 3.A.1)}:

$$|[q-\tfrac{1}{2}]_{+},q-\tfrac{1}{2},q-\tfrac{1}{2},q-\tfrac{1}{2}\rangle=\epsilon_2\sqrt{2q_2}\,|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2),q-\tfrac{1}{2},q-\tfrac{1}{2}\rangle$$
$$+(-1)^{\lambda_1+1}\epsilon_1\sqrt{2q_1}\,|(b_1,q_1),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{1}{2},q-\tfrac{1}{2}\rangle\,.$$

$$|[q-\tfrac{1}{2}]_{+},q,q-1,q-1\rangle$$
$$=\epsilon_1\epsilon_2\sqrt{2q}\,|(b_1,q_1),(b_2,q_2),q-1,q-1\rangle$$
$$+\zeta_1\epsilon_2\sqrt{2q_2(2q_1-1)}\,|(b_1,q_1-1),(b_2,q_2),q-1,q-1\rangle-\epsilon_1\zeta_2\sqrt{2q_1(2q_2-1)}\,|(b_1,q_1),(b_2,q_2-1),q-1,q-1\rangle$$
$$+(-1)^{\lambda_1+1}2q_1\alpha_1\epsilon_1|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1,q-1\rangle$$
$$+(-1)^{\lambda_1+1}2q_2\alpha_2\epsilon_2|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1,q-1\rangle\,.$$

$$|[q-1]_{+},q-1,q-1,q-1\rangle=|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1,q-1\rangle\,.$$

$$|[q-1]_{+},q-\tfrac{1}{2},q-\tfrac{3}{2},q-\tfrac{3}{2}\rangle=\epsilon_2\sqrt{2q_1-1}\,|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2),q-\tfrac{3}{2},q-\tfrac{3}{2}\rangle$$
$$+(-1)^{\lambda_1}\epsilon_1\sqrt{2q_2-1}\,|(b_1,q_1),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{3}{2},q-\tfrac{3}{2}\rangle$$
$$+(-1)^{\lambda_1}\zeta_1\sqrt{(2q-1)(2q_1-1)}\,|(b_1,q_1-1),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{3}{2},q-\tfrac{3}{2}\rangle$$
$$-\zeta_2\sqrt{(2q-1)(2q_2-1)}\,|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2-1),q-\tfrac{3}{2},q-\tfrac{3}{2}\rangle\,.$$

$$|[q]_{+},q,q,q\rangle=|(b_1,q_1),(b_2,q_2),q,q\rangle\,.$$

$$|[q]_{+},q+\tfrac{1}{2},q-\tfrac{1}{2},q-\tfrac{1}{2}\rangle=\alpha_1\sqrt{2q_1}\,|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2),q-\tfrac{1}{2},q-\tfrac{1}{2}\rangle$$
$$+\alpha_2\sqrt{2q_2}\,|(b_1,q_1),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{1}{2},q-\tfrac{1}{2}\rangle\,.$$

$$|[q-\tfrac{1}{2}]'_{+},q-\tfrac{1}{2},q-\tfrac{1}{2},q-\tfrac{1}{2}\rangle=\epsilon_2\sqrt{2q_2}\,|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2,q_2),q-\tfrac{1}{2},q-\tfrac{1}{2}\rangle$$
$$+(-1)^{\lambda_1}\epsilon_1\sqrt{2q_1}\,|(b_1,q_1),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-\tfrac{1}{2},q-\tfrac{1}{2}\rangle\,.$$

$$|[q-\tfrac{1}{2}]'_{+},q,q-1,q-1\rangle=2\epsilon_1\epsilon_2(q_1-q_2)(\sqrt{2q})^{-1}|(b_1,q_1),(b_2,q_2),q-1,q-1\rangle$$
$$-\zeta_1\epsilon_2\sqrt{2q_2(2q_1-1)}\,|(b_1,q_1-1),(b_2,q_2),q-1,q-1\rangle$$
$$-\epsilon_1\zeta_2\sqrt{2q_1(2q_2-1)}\,|(b_1,q_1),(b_2,q_2-1),q-1,q-1\rangle$$
$$+(-1)^{\lambda_1+1}2q_1\alpha_1\epsilon_1|(b_1+\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2-\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1,q-1\rangle$$
$$+(-1)^{\lambda_1}2q_2\alpha_2\epsilon_2|(b_1-\tfrac{1}{2},q_1-\tfrac{1}{2}),(b_2+\tfrac{1}{2},q_2-\tfrac{1}{2}),q-1,q-1\rangle \qquad \text{(A2.2)}$$

2. $[b_1,q_1]\otimes[b_2,q_2]$ for $b=q-n, n=1,\ldots,2q_<-1$

The following basis vectors become colinear:

$$|[b+\tfrac{1}{2},q+\tfrac{1}{2}-n],b,q-n,q-n\rangle$$

with one of the

$$|[b,q-n],b,q-n,q-n\rangle\,,$$

$$|[b+\tfrac{1}{2},q+\tfrac{1}{2}-n],b+\tfrac{1}{2},q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle$$

with one of the

$$|[b,q-n],b+\tfrac{1}{2},q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle\,,$$

$$|[b-\tfrac{1}{2},q-\tfrac{1}{2}-n],b-\tfrac{1}{2},q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle$$

with both

$$|[b,q-n],b-\tfrac{1}{2},q-\tfrac{1}{2}-n,q-\tfrac{1}{2}-n\rangle\,,$$

$$|[b-\tfrac{1}{2},q-\tfrac{1}{2}-n],b,q-1-n,q-1-n\rangle$$

with both

$$|[b,q-n],b,q-1-n,q-1-n\rangle\,. \qquad \text{(A2.3)}$$

Using Appendix 1 we found that $b\to q-n$ implies: $[b+\tfrac{1}{2},q+\tfrac{1}{2}-n]\oplus[b,q-n]\oplus[b,q-n]\oplus[b-\tfrac{1}{2},q-\tfrac{1}{2}-n]\to[q-n,q+\tfrac{1}{2}-n]_{+}\oplus[q-\tfrac{1}{2}-n,q-1-n,q-n]_{+}$. Its representation space is invariant whereas those of the remaining $[q-\tfrac{1}{2}-n]_{+}$ (twice) and $[q-n]_{+}$ (once) are not. By explicitly considering all representations of this form and all possible basis transformations we have been able to show that the result is indeed $[q-\tfrac{1}{2}-n,q-1-n,q-n,q-\tfrac{1}{2}-n]_{+}\oplus[q-n,q-\tfrac{1}{2}-n,q+\tfrac{1}{2}-n,q-n]_{+}$.

[1]Y. Ne'eman, Phys. Lett. B **81**, 190 (1979); Tel-Aviv preprint TAUP 727/79; D.B. Fairlie, J. Phys. G **5**, L55 (1979); Phys. Lett. B **82**, 97 (1979), P.H. Dondi and P.D. Jarvis, Southhampton preprints THEP 78/9-8, 78/9-13; R. Ecclestone, Southhampton preprint THEP 78/9-15.

[2]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).

[3]M. Marcu, J. Math. Phys. **21**, 1277 (1980).

[4]V.G. Kac, Commun. Alg. **5**, 889 (1977).

[5]The tensor product of two irreducible representations of osp(1,2) has been computed in Ref. 2. However, osp(1,2) has no nontypical representations.

[6]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).

Fock space representations of the Lie superalgebra $A(0,n)$

Tchavdar D. Palev[a]

Institute für Theoretische Physik der Technischen Universität Clausthal, Clausthal-Zellerfeld, West Germany

(Received 31 July 1979; accepted for publication 11 January 1980)

An infinite class of finite-dimensional irreducible representations and one particular infinite-dimensional representation of the special linear superalgebra of an arbitrary rank is constructed. For every representation an orthonormal basis in the corresponding representation space is found, and the matrix elements of the generators are calculated. The method we use is similar to the one applied in quantum theory to compute the Fock space representations of Bose and Fermi operators. For this purpose we first introduce a concept of creation and annihilation operators of a simple Lie superalgebra and give a definition of Fock-space representations.

1. INTRODUCTION

After the concept of supersymmetry was introduced in particle physics,[1] its physical consequences[2] as well as the underlying mathematical structure were investigated in several studies. It was soon recognized that the new symmetry was based on a mathematically known, however almost unstudied, algebraical structure, the so-called graded Lie algebras or Lie superalgebras (LS's).[3] During the last years much attention was devoted to the classification[4-10] and the representation theory[11-18] of the LS's. The biggest success is the full classification of all finite-dimensional irreducible representations of the so-called basic classical LS's.[16] In contrast to the simple Lie algebras, however, these representations do not exhaust all finite-dimensional representations. The simple LS's possess in general finite-dimensional representations that are not completely reducible. The theory of these representations, as well as of the infinite-dimensional representations, is still far from being complete. The study of the known representations, including those with highest weight (and hence all finite-dimensional irreducible representations), is also not yet on the corresponding level for the Lie algebras. In particular, the physically important problem of computing the matrix elements of the generators within a fixed basis of the classified modules have been solved only for the lowest rank LS's (see Refs. 5, 15, 18).

In the prsent paper we study one class of finite-dimensional irreducible representations and one particular infinite-dimensional representation of the basic classical Lie superalgebra[10] $A(0,n)$ for any value of n. The results hold also for the infinite-rank algebra $A(0,\infty)$. In every irreducible module we construct an orthonormal basis and calculate explicitly the matrix elements of the generators. The method we use is similar to the one applied in quantum theory for finding the Fock space representations of the Bose and Fermi operators. The guiding idea comes from the observation that for a given simple LS $\mathscr{A}$ one can choose a finite set $a_1^\pm,\ldots,a_m^\pm\in\mathscr{A}$ of root vectors such that, on one hand, they generate through multiplications and linear space operations the algebra $\mathscr{A}$ and, on the other hand, $a_1^\pm,\ldots,a_m^\pm$ allow construction of Fock-type representations in the usual way for quantum physics, so that the corresponding representations of $\mathscr{A}$ are irreducible. Because of the last property, we call the elements $a_1^\pm,\ldots,a_m^\pm$ creation and annihilation operators (CAO's). The Bose and Fermi operators and also their generalization, the paraoperators,[20] fit into this scheme. Any n pairs of para-Bose operators generate the simple LS $B(0,n)^2$; similarly n pairs of para-Fermi operators generate the simple Lie algebra B_n of the orthogonal group $SO(2n+1)$.[22,23]

In the terminology of Ref. 16 the finite-dimensional representations we obtain are typical representations induced by trivial representations of the subalgebra

$P=A_0(0,n)+A_1(0,n)$. Here

$$A(0,n)=A_{-1}(0,n)+A_0(0,n)+A_1(0,n) \tag{1}$$

is the distinguished Z-gradation. In general, the finite-dimensional irreducible $A(0,n)$- modules are labelled with $n+1$ numbers $(a_s,a_1,\ldots,a_n)$, where $a_s\in\mathbb{C}$ and $a_1,\ldots,a_n$ are arbitrary nonnegative integers, $a_i\in Z_+$. We study a sequence of $A(0,n)$- modules with signature $(p,0,\ldots,0), p\in Z_+$. The restrictions on the allowed modules come from the definition of the Fock space, which requires that the whole space be generated out of the highest weight (= the vacuum) by means of only odd negative roots, (i.e., of the creation operators). The infinite-dimensional representation of $A(0,2n-1)$ corresponds to a representation with highest weight induced from a trivial representation of a subalgebra generated from $gl(n)+gl(n)\subset A_0(0,2n-1)$ and a certain system of positive root vectors [whose linear envelope is noncommutative and hence different from $A_1(0,2n-1)$].

The representations we consider are star representations (introduced in Ref. 15) corresponding to an adjoint operation which is natural for the quantum theory:

$$(a_i^\pm)^*=a_i^\mp. \tag{2}$$

We introduce a Hilbert-space structure in the representation spaces in such a way that the $*$-operation (2) is a Hermitian conjugation. In principle one can construct the irreducible modules without defining a metric. For this purpose, however, one has first to determine the maximal submodule I in the $A(0,n)$-module $\bar{V}$ induced by the trivial representations of the subalgebra P and then past to the quotient $V=\bar{V}/I$.[16] In

[a] Present address: Institute of Nuclear Research and Nuclear Energy, boul. Lenin 72, 1184 Sofia, Bulgaria.

 0022-2488/80/061293-06$1.00

our case the metric singles out automatically an irreducible submodule of $\overline{V}$ which is isomorphic to V and thus simplifies all calculations considerably.

We wish to point out that the CAO's of $A(0,n)$ can be defined also from purely physical considerations as an alternative way for quantization of a scalar field in the framework of the Lagrangian quantum field theory which leads to a generalization of the quantum statistics different from the parastatistics.[24] In this respect it is important that the results hold for infinite and even continuum number of CAO's as this is the case in the quantum field theory. In this article we follow an algebraical approach. The CAO's are defined in a way that is more suitable for construction of representations. We first give a definition of CAO's and their Fock space representations (Sec. 2) and point out that the definition does not determine the operators uniquely within the LS. In Sec. 3 with one possible realization of CAO's we construct a class of finite-dimensional representations of the LS $A(0,n-1)$ and calculate the matrix elements of the generators in a proper basis. In Sec. 4 we write down the matrix elements of an infinite-dimensional representation of $A(0,2n-1)$ that is due to another realization of the creation and annihiliation operators.

2. DEFINITION OF CAO'S AND THEIR FOCK-SPACE REPRESENTATIONS

Definition 1: The minimal set of root vectors $a_1^\xi,...,a_m^\xi$ $\xi = \pm$, from the (semi) simple LS $\mathscr{A}$ with product $[\![\ ,\ ]\!]$ are said to be creation ($\xi = +$) and annihilation ($\xi = -$) operators of $\mathscr{A}$ if the following conditions are fulfilled.

1. The second order polynomials of $a_1^\pm,...,a_m^\pm$ with respect to $[\![\ ,\ ]\!]$ generate $\mathscr{A}$, i.e., (lin.env. = linear envelope),

$$\mathscr{A} = \text{lin.env.}\{a_i^\xi,[\![a_j^\eta,a_k^\epsilon]\!] | i,j,k = 1,...,m;\xi,\eta,\epsilon = \pm\}. \quad (3)$$

2. The ordering of the basis in the Cartan subalgebra H of $\mathscr{A}$ can be chosen such that the root of a_i^ξ is ξh^i, where h^i is a negative root [i.e., all $a_1^+,...,a_m^+$ ($a_1^-,...,a_m^-$) are negative (positive) root vectors].

This definition makes sense. It is known that the second order polynomials of para-Fermi (and hence of Fermi) operators $b_1^\pm,...,b_n^\pm$ generate the simple Lie algebra B_n of the orthogonal group.[22,23] To show that the second condition holds we recall that by definition[20] these operators satisfy the relation

$$[[b_i^\xi,b_j^\eta],b_k^\epsilon] = \tfrac{1}{2}(\eta-\epsilon)^2\delta_{jk}b_i^\xi - \tfrac{1}{2}(\xi-\epsilon)^2\delta_{ik}b_j^\eta. \quad (4)$$

Here and throughout the paper $\xi,\eta,\epsilon,\delta = \pm$ or ± 1, $[x,y] = xy - yx$; and $\{x,y\} = xy + yx$. By a space and an operator we always mean a linear space and linear operator. An ordered basis in the Cartan subalgebra $H \subset B_n$ can be chosen to be[25]

$$h_i = \tfrac{1}{2}[b_i^-,b_i^+],\quad i = 1,...,n. \quad (5)$$

From (4) we have

$$[h_i,b_j^\xi] = -\xi\delta_{ij}b_j^\xi. \quad (6)$$

Here b_j^ξ is a root vector and the corresponding root $\omega^{*\xi}_j$ in the dual basis $h^{*1},...,h^{*n}$ of the conjugate space H^* of H is

$$\omega^{*\xi}_j = \sum_{i=1}^n(-\xi\delta_{ij})h^{*i}.$$

Therefore the first nonzero coordinate of $\omega^{*\pm}_j$ is ∓ 1 and b_j^+ (b_j^-) is a negative (positive) root vector.

For another example take the Bose or, more generally, para-Bose operators $b_1^\pm,...,b_n^\pm$. They span a basis in the odd part of the LS $B(0,n)$.[21] In this case

$$B(0,n) = \text{lin.env.}\{b_i^\xi,\{b_j^\eta,b_k^\epsilon\} | i,j,k = 1,...,n;\xi,\eta,\epsilon = \pm\} \quad (7)$$

and the product between the odd and even elements is determined through the three linear relations of the para-Bose operators

$$[\{b_i^\xi,b_j^\eta\},b_k^\epsilon] = (\epsilon-\xi)\delta_{ik}a_j^\eta + (\epsilon-\eta)\delta_{kj}a_i^\xi. \quad (8)$$

If the basis in the Cartan subalgebra is chosen to be

$$h_i = -\tfrac{1}{2}\{b_i^+,b_i^-\},\quad i = 1,...,n, \quad (9)$$

then (6) holds and hence both conditions of Def. 1 hold. The examples show that the Bose and Fermi operators and also their generalization, the paraoperators, are CAO's in the sense of Def. 1.

There are different possible choices of CAO's for a given LS. Without going into a detailed discussion we remark that up to transformations from the Weyl group the roots corresponding to the CAO's are defined uniquely. For instance, in the case of classical Lie algebras the only possible choice of the roots so that the corresponding root vectors are CAO's is the following[24] ($i = 1,...,n,\eta = \pm$):

A_n $\quad a_i^\pm \leftrightarrow \mp(h^0 - h^i)$,

B_n $\quad b_i^\pm \leftrightarrow \mp h^i$,

C_{n-1} $\quad c_{\eta_i}^\pm \leftrightarrow \mp(h^0 - \eta h^i)$,

D_{n-1} $\quad d_{\eta_i}^\pm \leftrightarrow \mp(h^0 + \eta h^i)$.

For a construction of representations via CAO's we define a Fock space through some of the main properties of the ordinary Fock space of Bose and Fermi CAO's.

Definition 2: The irreducible $\mathscr{A}$-module (= irreducible representation space of the LS $\mathscr{A}$) W is called a Fock space of $\mathscr{A}$ if CAO's can be chosen such that

(1) W contains a vector $|0\rangle$, called a vacuum, so that

(a) $a_i^-|0\rangle = 0,\quad i = 1,...,n,$ (10)

(b) $a_i^-a_j^+|0\rangle = \delta_{ij}p_i|0\rangle,\quad p_i$ are constants; (11)

(2) W is a Hilbert space with metric (,) defined in the usual way (see also Lemma 1),

$$\text{(a) } (a_{i_1}^+a_{i_2}^+\cdots a_{i_m}^+|0\rangle, a_{j_1}^+a_{j_2}^+\cdots a_{j_r}^+|0\rangle) = \langle 0|a_{i_m}^-\cdots a_{i_2}^-a_{i_1}^-a_{j_1}^+a_{j_2}^+\cdots a_{j_r}^+|0\rangle, \quad (12)$$

(b) For any polynomial P of CAO's

$\langle 0|a_i^+P|0\rangle = 0,\quad i = 1,...,n,$

(c) $(|0\rangle,|0\rangle) \equiv \langle 0|0\rangle = 1.$

For the Bose and Fermi operators all p_i in (11) are equal to 1. In the parastatistics, $p_1 = p_2 = \cdots = p_n = p$ is a positive integer called an order of the statistics. For different values of p one obtains different irreducible representations of B_n or $B(0,n)$. For instance, the representation of B_n with order of the statistics p has a signature [that is, coordinates of the highest weight $|0\rangle$ in the basis (5)] $(p/2, p/2,..., p/2)$.[25]

The following property of the Fock space of $\mathscr{A}$ is a consequence of (10) and (11) only.

Lemma 1: A Fock space W is the closure of the linear envelope W' of all possible vectors

$$a_{i_1}^+ a_{i_2}^+ \cdots a_{i_m}^+ |0\rangle, \quad m = 0,1,2,\cdots. \tag{13}$$

Proof: Let, for some values of j,k,η,ϵ, a_i^ξ, $i = 1,...,n$, and $[\![a_j^\eta, a_k^\epsilon]\!]$ be the generators of $\mathscr{A}$. The proof is based on the Poincaré–Birkhoff–Witt theorem,[26] stating in our case that the basis in the universal enveloping algebra of $\mathscr{A}$ is given by all monomials

$$\begin{aligned} P(A,...,F) = &\prod_{i,j} [\![a_i^+, a_j^+]\!]^{A_{ij}} \prod_k (a_k^+)^{n_k} \\ &\times \prod_l [\![a_l^-, a_l^+]\!]^{C_l} \prod_{p\neq q} [\![a_p^-, a_q^+]\!]^{D_{pq}} \\ &\times \prod_{r,s} [\![a_r^-, a_s^-]\!]^{E_{rs}} \prod_m (a_m^-)^{F_m}. \end{aligned} \tag{14}$$

Here A_{ij}, B_k, C_l, D_{pq}, E_{rs} and F_m are arbitrary nonnegative integers; the products are to be taken over all possible values of $i, j,...,m$. Any vector $|x\rangle \in W'$ is a linear combination of vectors $P(A,...,F)|0\rangle$. The latter is different from zero only for $D_{pq} = E_{rs} = F_m = 0$. To complete the proof it remains to remark that due to (11) the non-zero vectors $\Pi_l [\![a_l^-, a_l^+]\!]^{C_l} \times |0\rangle$ are proportional to $|0\rangle$.

Lemma 2: Up to a multiplicative constant the vacuum is unique.

Proof: Suppose there is another vacuum $|0_1\rangle$. Without loss of generality we can accept that $|0_1\rangle$ is orthogonal to $|0\rangle$ ($\alpha|0_1\rangle + \beta|0\rangle$ is also a vacuum). Let $|x\rangle$ be an arbitrary vector from the dense domain W'. According to Lemma 1 there exists a polynomial $P(a^+)$ of the creation operators such that $|x\rangle = P(a^+)|0\rangle$. Since $\langle 0_1|a_i^+ Q|0\rangle = 0$, for every $|x\rangle \in W'$

$$\langle 0_1|P(a^+)|0\rangle = \langle 0_1|x\rangle = 0.$$

Therefore $|0_1\rangle = 0$.

For the Hermitian conjugate L^* of L in W we obtain

Corollary: On the dense subspace $W' \subset W (a_i^+)^* = a_i^-$.

3. FINITE-DIMENSIONAL FOCK REPRESENTATIONS OF $A(0,n-1)$

In order to choose CAO's of the Lie superalgebra $A(0,n-1)$, it is convenient to introduce first the general linear superalgebra $gl(1,n)$. The latter is given by the set of all $(n+1)$-dimensional square matrices. Let $e_{\alpha\beta}$, $\alpha,\beta = 0,1,...,n$ be a matrix from $gl(1,n)$ with 1 in the αth row and βth column and zero elsewhere. The even and the odd parts G_0 and G_1 of $gl(1,n) = G_0 + G_1$ are

$$G_1 = \text{lin.env.}\{e_{0i}, e_{i0} | i = 1,2,...,n\}, \tag{15}$$

$$G_0 = \text{lin.env.}\{e_{00}, e_{ij} | i,j = 1,...,n\}. \tag{16}$$

The multiplication in $gl(1,n)$ is obtained by linear extension of

$$[\![a,b]\!] = ab - (-1)^{\alpha\beta} ba, \quad a \in G_\alpha,\ b \in G_\beta,\ \alpha,\beta = 0,1, \tag{17}$$

where ab is the usual matrix product of a and b.

The LS $gl(1,n)$ is not simple since it contain as nontrivial ideal the subalgebra of all matrices that are proportional to the unit matrix. The special linear Lie superalgebra

$$A(0,n-1) = A_0(0,n-1) + A_1(0,n-1) \tag{18}$$

is subalgebra of $gl(1,n)$ with

$$A_1(0,n-1) = \text{lin.env.}\{e_{0i}, e_{i0} | i = 1,2,...,n\}, \tag{19}$$

$$A_0(0,n-1) = \text{lin.env.}\{e_{00} + e_{kk}, e_{ij} | i \neq j, i,j,k = 1,...,n\}. \tag{20}$$

The Cartan subalgebra H of $A(0,n-1)$ can be chosen to be

$$H = \{e_{00} + e_{kk} | k = 1,...,n\}. \tag{21}$$

Now we define the CAO's of $A(0,n-1)$. To obtain a representation independent choice we follow the method of Ref. 27. Take the algebra E of formal polynomials of indeterminates $a_1^\pm,...,a_n^\pm$ with additional algebraical relations

$$\begin{aligned} &[\{a_i^+, a_j^-\}, a_k^+] = \delta_{jk} a_i^+ - \delta_{ij} a_k^+, \\ &[\{a_i^+, a_j^-\}, a_k^-] = -\delta_{ik} a_j^- + \delta_{ij} a_k^-, \\ &\{a_i^+, a_j^+\} = \{a_i^-, a_j^-\} = 0. \end{aligned} \tag{22}$$

Consider the linear subspaces A_0 and A_1 from E,

$$\begin{aligned} &A_1 = \text{lin.env.}\{a_i^\xi | \xi = \pm, i = 1,...,n\}, \\ &A_0 = \text{lin.env.}\{\{a_i^+, a_j^-\} | i,j = 1,...,n\}, \end{aligned} \tag{23}$$

and let $A = A_0 + A_1 \subset E$ be the direct space sum of A_0 and A_1. For any elements $a_\alpha, b_\alpha \in A_\alpha$, $\alpha = 0,1$, we define a product $[\![\ ,\ ]\!]$ in A as a linear extension of the relations

$$[\![a_1,b_1]\!] = \{a_1,b_1\}, \quad [\![a_0,b_\alpha]\!] = [a_0,b_\alpha]. \tag{24}$$

Proposition 1: The above elements $a_1^\xi,...,a_n^\xi$ are creation ($\xi = +$) and annihiliation ($\xi = -$) operators of $A(0,n-1)$.

Proof: Because of (22) and (24) A is a LS with an even part A_0 and an odd part A_1. Let Θ be a one-to-one linear map of $A(0,n-1)$ onto A,

$$\begin{aligned} &\Theta: e_{i0} \to a_i^+;\ e_{0i} \to a_i^-, \\ &e_{ij} + \delta_{ij} e_{00} \to \{a_i^+, a_j^-\}, \end{aligned} \qquad i,j = 1,...,n. \tag{25}$$

A simple, however, somewhat lengthy calculation shows

$$[\![\Theta(x),\Theta(y)]\!] = \Theta([\![x,y]\!]). \tag{26}$$

Therefore Θ is an isomorphism. Thus the second order polynomials (3) of n pairs $a_1^\pm,...,a_n^\pm$ generate $A(0,n-1)$. To show that the last requirement of Def. 1 holds, we obtain

$$[h, a_i^\xi] = \omega^{*\xi}_i(h) a_i^\xi, \quad h \in H, \xi = \pm, i = 1,...,n. \tag{27}$$

Therefore, $a_1^\pm,...,a_n^\pm$ are root vectors of $A(0,n-1)$ with roots $\omega^{*\xi}_i$ that are linear functionals (i.e., elements from the dual to the Cartan subalgebra H space H^*, $\omega^{*\xi}_i \in H^*$). It is possible to consider $\omega^{*\xi}_i$ always as a negative ($\xi = +$) or a positive ($\xi = -$) root by choosing the ordered basis in H as

$$\begin{aligned} &h_1 = \{a_1^+, a_1^-\} + \{a_2^+, a_2^-\}, \\ &h_i = \{a_1^+, a_1^-\} - \{a_i^+, a_i^-\} \end{aligned} \qquad i = 2,3,...,n. \tag{28}$$

From (22) we have

$$[h_1, a_k^\xi] = \xi(\delta_{1k} + \delta_{2k} - 2) a_k^\xi. \tag{29}$$

Therefore

$$\omega^{*\xi}_{k}(h_1) = \xi(\delta_{1k} + \delta_{2k} - 2). \tag{30}$$

Since $\omega^{*\xi}_{i}(h_1)$ is the first co-ordinate of the root $\omega^{*\xi}_{i}$ in the dual basis, from (30) we have

$$\omega^{*+}_{i}(h_1) \leqslant -1, \quad \omega^{*-}_{i}(h_1) \geqslant 1. \tag{31}$$

Hence a_i^+ (a_i^-), are negative (positive) root vectors. This completes the proof.

Now we give a class of finite-dimensional representations of $A(0,n-1)$. Because of Def. 2 we postulate that

$$a_i^- a_j^+ |0\rangle = \delta_{ij} p |0\rangle, \tag{32}$$

(i.e., we consider the case $p_1 = p_2 = \cdots = p_n = p$ with p[28] being positive integer, $p = 1,2,\cdots$). In this case the metric in the Fock space is positive definite. For para-Bose and para-Fermi statistics (32) also holds for the representations of the CAO's of $B(0,n)$ and B_n, respectively. Therefore in this paper we call the integer p an order of the (A-super) statistics.

Let us denote by $W(n,p)$ the Fock space of $A(0,n-1)$ with an order of the statistics p. Since $\{a_i^+, a_i^+\} = 0$, the square of every operator $(a_i^+)^2 = 0$. A related statement is given in the following lemma.

Lemma 3: Let $q = \min(n,p)$. The product of arbitrary $q+1$ creation operators is the zero operator in $W(n,p)$, i.e.,

$$a_{i_1}^+ a_{i_2}^+ \cdots a_{i_{q+1}}^+ = 0. \tag{33}$$

Proof: (1) If $p > n$, then $q = n$ and the product (33) has to contain at least one operator, say a_i^+, twice. Since the creation operators anticommute and for $n > 1$ $(a_i^+)^n = 0$, (33) holds. So the representation space $W(n;p)$ is finite for any p.

(2) For the case $n > p$ (i.e., $q = p$) we first show that

$$J_m \equiv a_i^- a_{j_1}^+ \cdots a_{j_k}^+ \cdots a_{j_m}^+ |0\rangle = (p - m + 1) \sum_{k=1}^{m} (-1)^{k+1} \delta_{ij_k} a_{j_1}^+ \cdots a_{j_{k-1}}^+ a_{j_{k+1}}^+ \cdots a_{j_m}^+ |0\rangle. \tag{34}$$

We prove by induction. For $m = 1$ the above equation reduces to the Fock space defining relation (32). Suppose (34) is true for $m < p+1$ and consider J_{m+1},

$$J_{m+1} = a_i^- a_{j_0}^+ a_{j_1}^+ \cdots a_{j_m}^+ |0\rangle = [\{a_i^-, a_{j_0}^+\}, a_{j_1}^+ \cdots a_{j_m}^+]|0\rangle + a_{j_1}^+ \cdots a_{j_m}^+ \{a_i^-, a_{j_0}^+\}|0\rangle - a_{j_0}^+ a_i^- a_{j_1}^+ \cdots a_{j_m}^+ |0\rangle.$$

Using (34) and the identity

$$[\{a_i^-, a_j^+\}, a_{j_1}^+ \cdots a_{j_k}^+ \cdots a_{j_m}^+] = \sum_{k=1}^{m} \delta_{ij_k} a_{j_1}^+ \cdots a_{j_{k-1}}^+ a_j^+ a_{j_{k+1}}^+ \cdots a_{j_m}^+ - m\delta_{ij} a_{j_1}^+ \cdots a_{j_m}^+, \tag{35}$$

we obtain

$$J_{m+1} = \sum_{k=1}^{m} \delta_{ij_k} a_{j_1}^+ \cdots a_{j_{k-1}}^+ a_{j_0}^+ a_{j_{k+1}}^+ \cdots a_{j_m}^+ |0\rangle - (p - n + 1) \sum_{k=1}^{m} (-1)^{k+1} \delta_{ij_k} a_{j_0}^+ a_{j_1}^+ \cdots a_{j_{k-1}}^+ a_{j_{k+1}}^+ \cdots a_{j_m}^+ |0\rangle + (p-n)\delta_{ij_0} a_{j_1}^+ \cdots a_{j_m}^+ |0\rangle$$

$$= (p-n) \sum_{k=0}^{m} (-1)^{k+1} \delta_{ij_k} a_{j_0}^+ a_{j_1}^+ \cdots a_{j_{k-1}}^+ a_{j_{k+1}}^+ \cdots a_{j_m}^+ |0\rangle.$$

Thus Eq. (34) holds also for $m+1$ and hence for any $m \leqslant p+1$. For $m = p+1$ Eq. (34) gives

$$J_{p+1} = a_i^- a_{j_1}^+ \cdots a_{j_{p+1}}^+ |0\rangle = 0. \tag{36}$$

(3) For any vector $a_{i_1}^+ \cdots a_{i_m}^+ |0\rangle$, $m = 1,2,\cdots$,

$$K \equiv (a_{i_1}^+ \cdots a_{i_m}^+ |0\rangle, a_{j_1}^+ \cdots a_{j_{p+1}}^+ |0\rangle) = 0. \tag{37}$$

This is evident for $m = 0$ because of (12b). If $m > 0$, we have from (12) and (36)

$$K = \langle 0| a_{i_m}^- \cdots a_{i_1}^- a_{j_1}^+ \cdots a_{j_{p+1}}^+ |0\rangle = 0.$$

Since an arbitrary vector is a linear combination of vectors $a_{i_1}^+ a_{i_2}^+ \cdots a_{i_m}^+ |0\rangle$ we find that for any $x \in W(n;p)$

$$(x, a_{j_1}^+ \cdots a_{j_{p+1}}^+ |0\rangle) = 0. \tag{38}$$

Therefore,

$$a_{j_1}^+ \cdots a_{j_{p+1}}^+ |0\rangle = 0, \tag{39}$$

This immediately implies that for arbitrary vector $a_{i_1}^+ \cdots a_{i_r}^+ |0\rangle$

$$a_{j_1}^+ \cdots a_{j_{p+1}}^+ a_{i_1}^+ \cdots a_{i_r}^+ |0\rangle = 0,$$

Hence for all $|y\rangle \in W(n;p)$

$$a_{j_1}^+ \cdots a_{j_{p+1}}^+ |y\rangle = 0 \text{ and therefore } a_{j_1}^+ \cdots a_{j_{p+1}}^+ = 0, \tag{40}$$

which completes the proof.

We remark that if a_i^+ is interpreted as an operator creating a particle in a state "i", Eq. (39) means that an arbitrary particle ensemble cannot contain more than $q = \min(n,p)$ particles, which because of $(a_i^+)^2 = 0$ must be in different states. To calculate the matrix elements, we first prove the following theorem.

Theorem 1: The set of all vectors

$$|i_1, i_2, \ldots, i_m\rangle = (p!)^{-1/2}((p-m)!)^{1/2} a_{i_1}^+ a_{i_2}^+ \cdots a_{i_m}^+ |0\rangle, \tag{41}$$

where $i_1 < i_2 < \cdots < i_m$, $m = 0,1,2,\ldots,\min(n;p)$ build an orthonormal basis in $W(n;p)$.

Proof: Suppose that in

$$S \equiv (a_{i_1}^+ \cdots a_{i_m}^+ |0\rangle, a_{j_1}^+ \cdots a_{j_r}^+ |0\rangle)$$

the index i_k is not contained in the index set $(j_1, \ldots, j_r)$. Then

$$S = (-1)^{k-1} \langle 0| a_{i_m}^- \cdots a_{i_{k+1}}^- a_{i_{k-1}}^- \cdots a_{i_k}^- a_{j_1}^+ \cdots a_{j_r}^+ |0\rangle = 0,$$

since $a_{i_k}^- a_{j_1}^+ \cdots a_{j_r}^+ |0\rangle = 0$. Thus $S \neq 0$ if and only if $m = r$ and $(i_1, \ldots, i_m)$ and $(j_1, \ldots, j_r)$ coincide. To calculate the norm of $a_{i_1}^+ \cdots a_{i_m}^+ |0\rangle$ we use Eq. (34) repeatedly.

$$(a_{i_1}^+ \cdots a_{i_m}^+ |0\rangle, a_{i_1}^+ \cdots a_{i_m}^+ |0\rangle) = \langle 0| a_{i_m}^- \cdots a_{i_1}^- a_{i_1}^+ \cdots a_{i_m}^+ |0\rangle = (p-m+1)\langle 0| a_{i_m}^- \cdots a_{i_2}^- a_{i_2}^+ \cdots a_{i_m}^+ |0\rangle = (p-m+1)(p-m+2)\cdots(p-1)p = p!/(p-m)!.$$

This proves the theorem.

To calculate the matrix elements of the CAO's it is convenient to use the occupation number representation. Let

$$(\theta_1, \theta_2, \ldots, \theta_n) = |i_1, i_2, \ldots, i_m\rangle. \tag{42}$$

where $\theta_{i_1} = \theta_{i_2} = \cdots = \theta_{i_m} = 1$ and all other $\theta_i = 0$. From (22) and (34) one obtains

$$a_k^- |\ldots,\theta_k,\cdots) = \theta_k(-1)^{\theta_1 + \cdots + \theta_{k-1}} \Big(p - \sum_i \theta_i + 1\Big)^{1/2} |\ldots,\theta_k - 1,\cdots), \tag{43}$$

$$a_k^+ |\ldots,\theta_k,\cdots) = (1-\theta_k)(-1)^{\theta_1 + \cdots + \theta_{k-1}} \Big(p - \sum_i \theta_i\Big)^{1/2} |\ldots,\theta_k + 1,\cdots).$$

This yields immediately the matrix elements of the other generators of $A(0,n-1)$.

It is interesting to note that the structure relations (22), as well as Lemma 3, Theorem 1, and Eq. (43), are valid also for infinite number of CAO's and hence for $A(0,\infty)$. In this case any finite subset $a_{i_1}^\pm,\ldots,a_{i_n}^\pm$ of CAO's determines, through (43), an irreducible representation of $A(0,n-1)$.

The following theorem characterize the representations obtained so far.

Theorem 2: The representation of the LS $A(0,n-1)$ in $W(n;p)$ and $W(n;p')$ are equivalent if and only if $p = p'$.

Proof: We first calculate the dimension of $W(n;p)$. Call, for simplicitly, the basis vector $|\theta_1,\theta_2,\ldots,\theta_n)\in W(n;p)$ an m-state if $\theta_1 + \theta_2 + \cdots + \theta_n = m$. Since in $(\theta_1,\ldots,\theta_n)$, with $\theta_i = 0$ or 1 and the restriction $\Sigma_i \theta_i = m$, the $\theta_1,\ldots,\theta_n$ can be distributed in $\binom{n}{m}$ different ways, the subspace $W_m(n;p)$ of all m-states has a dimension

$$\dim W_m(n;p) = \binom{n}{m}. \tag{44}$$

Taking into account that $m = 0,1,2\ldots,q = \min(n,p)$, we find

$$\dim W(n;p) = \sum_{m=0}^{q} \binom{n}{m}. \tag{45}$$

Hence all $W(n;p)$ with $p = 1,2,\ldots,n$ have different dimensions and the corresponding representations are inequivalent. Spaces with $p \geqslant n$ have the same dimension,

$$\dim W(n;p) = \sum_{m=0}^{n} \binom{n}{m} = 2^n, \quad p \geqslant n \tag{46}$$

So we have n irreducible representations with different dimensions and an infinite sequence of irreducible representations realized in a space of dimension 2^n. To show that the latter are pairwise inequivalent, we calculate the character $\mathrm{Tr}I$ of the central element from the even part $\mathrm{gl}(n)$ of $A(0,n-1)$. In terms of the CAO's, I reads as

$$I = \sum_{k=1}^{n} \{a_k^+, a_k^-\}, \tag{47}$$

From (43) we obtain that for any m-state $|\theta_1,\ldots,\theta_n)_m$

$$\{a_k^+, a_k^-\} |\theta_1,\ldots,\theta_n)_m = (p - m + \theta_k) |\theta_1,\ldots,\theta_n)_m.$$

Therefore

$$I |\theta_1,\ldots,\theta_n)_m = \sum_{k=1}^{n} \{a_k^+, a_k^-\} |\theta_1,\ldots,\theta_n)_m = [np - (n-1)m] |\theta_1,\ldots,\theta_n)_m. \tag{48}$$

Since the number of all m-states is $\binom{n}{m}$ and $0 \leqslant m \leqslant n$, for $p \geqslant n$, we obtain

$$\mathrm{Tr}I = \sum_{m=0}^{n} \binom{n}{m} (np - nm\, m) = C_1(n)\,p + C_2(n), \tag{49}$$

where $C_1(n) = 2^n n$ and $C_2(n) = -(n-1)\Sigma_{m=0}^n m\binom{n}{m}$. So any two representations realized in $W(n;p)$ and $W(n;p')$ have different characters for $p \neq p'$ and are inequivalent.[29]

As an example we consider the Fock space representations $W(2;p)$ of $A(1,0)$ for $p \geqslant 2$. All such representations are four-dimensional. If we order the basis in $W(2;p)$ to be $|0,0)$, $|1,1)$, $|1,0)$, $|0,1)$, we obtain the following matrix realization of the generators.

Odd generators:

$$a_1^- = p^{1/2} e_{13} + (p-1)^{1/2} e_{42}, \quad a_1^+ = p^{1/2} e_{31} + (p-1)^{1/2} e_{24},$$
$$a_2^- = p^{1/2} e_{14} - (p-1)^{1/2} e_{32}, \quad a_2^+ = p^{1/2} e_{41} - (p-1)^{1/2} e_{23}, \tag{50}$$

Even generators:

A_1-generators,

$$H_3 = \tfrac{1}{2}\{a_1^+, a_1^-\} - \tfrac{1}{2}\{a_2^+, a_2^-\} = \tfrac{1}{2}(e_{33} - e_{44}),$$
$$H_+ = \{a_1^+, a_2^-\} = e_{34}, \quad H_- = \{a_2^+, a_1^-\} = e_{43}, \tag{51}$$

Center:

$$I = \{a_1^+, a_1^-\} + \{a_2^+, a_2^-\} = 2pe_{11} + (2p-2)e_{22} + (2p-1)e_{33} + (2p-1)e_{44}. \tag{52}$$

The above expressions show that for $p \geqslant 2$ the Fock space contains two spin-zero and one spin-1/2 representations, i.e.,

$$W(2;p)|_{A_1} = (0) + (0) + (1/2). \tag{53}$$

This is the well known four-dimensional boson-fermion representation, which accomodated for the first time within one irreducible multiplet particles with integer- and half-integer-spin and later on led to the invention of the supersymmetry.[1]

4. INFINITE-DIMENSIONAL FOCK SPACE REPRESENTATIONS OF *A(0,2n − 1)*

Within a given LS the definition of the CAO's is not unique. With a rearrangement of the basis in the Cartan subalgebra it is always possibe to turn any of the positive root vectors into a negative one. If we choose a new basis to be $h_i' = \eta_i h_i$, $\eta_i = \pm 1$ (see Eq. 28) then the operators $a_i^{\eta_i}$ become positive root vectors. It turns out that even such a simple replacement of some of the creation operators through annihilation ones and vice versa leads in general to nonequivalent Fock representations of the corresponding algebra. As an example of this kind we shall review (without giving any proofs) results obtained in Ref. 30.

We consider the even rank LS's $A(0,2n-1)$ for $n = 1,2,\cdots$. The CAO's are labelled with three indices $a_{\eta i}^{\xi}$, where $\eta = \pm$, $i = 1,2,\ldots,n$ and as before $\xi = +$ ($\xi = -$) corresponds to creation (annihilation) operator. The structure relation between the CAO's reads as

$$[\{a_{-\epsilon\xi i}^{\xi}, a_{\epsilon\eta j}^{\eta}\}, a_{\mu\nu k}^{\nu}] = \tfrac{1}{2}(\nu - \eta)\delta_{-\epsilon j,\mu k} a_{-\epsilon\xi i}^{\xi}$$

$$+\tfrac{1}{2}(\nu-\xi)\delta_{\epsilon i,\mu k}a^{\eta}_{\epsilon\eta j}+\tfrac{1}{2}(\xi-\eta)\mu\epsilon\delta_{ij}a^{\nu}_{\mu\nu k},$$

$$\{a^{\xi}_{\epsilon\xi i},a^{\eta}_{\epsilon\eta j}\}=0. \tag{54}$$

The operators $a^{\xi}_{\eta i}$ satisfy all the requirements of the Definition 1. The Fock spaces $W(n;p,q)$ in this case are labelled with two nonnegative integers $p,q=0,1,2,\cdots$ and are completely determined from the conditions

$$a^{-}_{\eta i}|0\rangle=0,\quad a^{-}_{i}a^{+}_{j}=p\delta_{ij}|0\rangle,\quad a^{-}_{-i}a^{+}_{-j}|0\rangle=q\delta_{ij}|0\rangle. \tag{55}$$

Here we consider only the simples representation $W(n;1,0)$, i.e., we put $p=1$, $q=0$.

Theorem 3: The product $a^{+}_{\eta i}a^{+}_{\eta j}$, $\eta=\pm$, $i,j=1,\ldots,n$ is a zero operator in $W(n;1,0)$. The representation space is infinite-dimensional. It is spanned on all possible vectors $(i_k=1,\ldots,n)$

$$a^{+}_{\xi i_m}a^{+}_{-\xi i_{m-1}}\cdots a^{+}_{-i_4}a^{+}_{-i_3}a^{+}_{-i_2}a^{+}_{i_1}|0\rangle,\quad \xi=(-1)^{m+1}. \tag{56}$$

The state (56) is symmetric with respect to arbitrary permutations of creation operators $a^{+}_{\eta i}$ with η only $+$ or $-$.

Because of this symmetry every vector (56) is completely determined by the number of the creation operators $a^{+}_{\eta i}$ in the state ηi. This justifies the notation

$$|z;p_1,p_2,\ldots,p_n;q_1,q_2,\ldots,q_n\rangle \tag{57}$$

for every vector (56), generated from $|0\rangle$ by means of a monomial which is a homogeneous function of order p_i(resp.q_i) of a^{+}_{i} (resp.a^{+}_{-i}) and z is

$$z=\sum_{i=1}^{n}(p_i-q_i)=0\quad\text{or}\quad 1. \tag{58}$$

It is not difficult to calculate the matrix elements of the CAO's in the basis (57). For this, represent the space $W(n;1,0)$ as a direct sum of its subspaces with $z=0$ and $z=1$, denoted as W_0 and W_1 resp. $W(n;1,0)=W_0+W_1$. Then

$$\begin{aligned}
&a^{+}_{-i}W_0=a^{-}_{i}W_0=a^{+}_{i}W_1=a^{-}_{-i}W_1=0,\\
&a^{+}_{i}|0;\ldots,p_i,\cdots\rangle=|1;\ldots,p_i+1,\cdots\rangle,\\
&a^{+}_{-j}|0;\ldots,q_j,\cdots\rangle=|0;\ldots,q_j+1,\cdots\rangle,\\
&a^{-}_{i}|1;\ldots,p_i\cdots\rangle=p_i|0;\ldots,p_i-1,\cdots\rangle,\\
&a^{-}_{-j}|0,\ldots,q_j,\cdots\rangle=q_j|1;\ldots,q_j-1,\cdots\rangle
\end{aligned} \tag{59}$$

holds. The orthonormalized basis in $W(n;1,0)$ is

$$|z;p_1,\ldots,q_n)=(p_1!\cdots p_n!q_1!\cdots q_n!)^{-1/2}|z;p_1,\ldots,q_n\rangle. \tag{60}$$

The representation is obviously irreducible. It is infinite-dimensional since p_i and q_i are arbitrary nonnegative integers. The only constraint on them is that $z=0$ or 1.

5. CONCLUDING REMARKS

The method outlined in this paper can be applied for construction of representation also for other Lie superalgebras. For the simple Lie algebras an explicit construction was carried out in Ref. 24.

The mathematical frame of CAO's presented above can be motivated also on more physical background. In fact, the realization (54) was derived in Ref. 30 as a possible generalization of the statistics of charged tensor fields. In this case z is the charge of the ensemble of particles being zero or 1.

Definition 1 reflects some of the properties of the Bose and Fermi operators. For the representation theory of the LS's it could be also useful to postulate some of the other features of the ordinary CAO's and this could lead in general to other irreducible representations. For instance, one can require that the creation operators $a_1^+,\ldots,a_n^+$ (and hence the annihilation operators $a_1^-,\ldots,a_n^-$) are (anti) commuting root vectors generating a LS of rank n. This is compatible with the multiplication of any simple LS, however, in general it may lead to higher than three-linear defining relations for the CAO's. We shall study elsewhere this possibility as well as the physical consequences of the considered statistics.

ACKNOWLEDGMENT

The author is thankful to Professor H.D. Doebner for the valuable discussions, critical reading of the manuscript, and the several suggestions he made, which improved considerably the initial version of the paper.

[1] Y.A. Gol'fand and E.P. Likhtman, JETP Lett. **13**, 323 (1972); D.V. Volkov and V.P. Akulov, Phys. Lett. B **46**, 109 (1973); J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974); Phys. Lett. B **49**, 52 (1974).
[2] See, for instance, the review articles B. Zumino, Proceedings of the 17th International Conference on High Energy Physics, London 1974; L. Corwin, J. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975); V.G. Ogievetsky and L. Mezincescu, Usp. Fiz. Nauk **117**, 637 (1975); S. Ferrara, Nuovo Cimento **6**, 105 (1975); P. Fayet and S. Ferrara, Phys. Rep. C **32**, 249 (1977); Abdus Salam and J. Strathdee, Fortschr. Phys. **26**, 57 (1978).
[3] See, for instance, F.A. Berezin and G.I. Kac, Math. USSR Sb. **11**, 311 (1970); M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626, 1640 (1976).
[4] V.G. Kac, Adv. Math. **26**, 8 (1977).
[5] A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975).
[6] P.G.O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976).
[7] W. Nahm, V. Rittenberg, and M. Scheunert, Phys. Lett. B **61**, 383 (1976).
[8] D.Z. Djokovic, J. Pure, Appl. Alg. **7**, 217 (1976).
[9] W. Nahm and M. Scheunert, J. Math. Phys. **17**, 868 (1976).
[10] N. Backhouse, J. Math. Phys. **19**, 2400 (1978).
[11] L. Ross, Trans. Amer. Math. Soc. **120**, 17 (1965).
[12] G. Hochschild, Ill. J. Math. **20**, 107 (1976).
[13] D.Z. Djokovic and G. Hochschild, Ill. J. Math. **20**, 134 (1976).
[14] D.Z. Djokovic, J. Pure Appl. Alg. **9**, 25 (1976).
[15] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146, 155 (1977).
[16] V.G. Kac, "Representations of classical Lie superalgebras," Lecture Notes in Math. **676**, 597 (1978).
[17] V.G. Kac, Proceedings of the Group Theoretical Methods in Physics, Austin 1978, p. 441.
[18] M. Bednář and V.F. Šachl, J. Math. Phys. **20**, 367 (1979).
[19] Throughout the paper we use the Kac notation for the simple LS's from Ref. 4.
[20] H.S. Green, Phys. Rev. **90**, 270 (1953).
[21] A. Ganchev and T. Palev, J. Math. Phys. **21**, 797 (1980).
[22] S. Kamefuchi and Y. Takahashi, Nucl. Phys. **36**, 177 (1960).
[23] C. Ryan and E.C.G. Sudarshan, Nucl. Phys. **47**, 207 (1963).
[24] T. Palev, Thesis, Institute of Nuclear Research and Nuclear Energy, Sofia 1976, Preprints JINR E17-10550 (1977); E2-11904 and E2-11905 (1978), to appear in Rep. Math. Phys.; E2-11929 and E2-11942 (1978), C.R. Acad. Bulg. Sci. **30**, 993 (1977); Proc. XV-th ICCR, **11**, 546 (1977); Proc. III School of Elementary Particles, Promorsco, Bulgaria, 418 (1977).
[25] T. Palev, Ann. Inst. Henri Poincaré **23**, 49 (1975).
[26] J. Milnor and J. Moore, Ann. Math. **81**, 211 (1965).
[27] H.D. Doebner and T. Palev, Preprint IC/71/104 (1971).
[28] This is not a restriction. One can show that the $A(0,n-1)$-module is a Fock space if and only if all p_i are equal.
[29] Note that for semisimple Lie algebras only a finite number of inequivalent irreducible representations will exist in a given finite-dimensional space.
[30] T. Palev, Preprints JINR, E2-11942 and E2-11943 (1978).

The theory of spinors via involutions and its application to the representations of the Lorentz group

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

(Received 22 October 1979; accepted for publication 21 January 1980

A simple theory of the spinor representations of the complex orthogonal group O(d,$\mathbb{C}$) in the d-dimensional Euclidean space $V^{(d)}$ is presented via a basic lemma on involutional transformations and Cartan's theorem on O(d,$\mathbb{C}$). The arbitrary gauge factors of the representations are reduced to $\pm$ signs by introducing appropriate phase conventions. The concept of an axial involution is introduced. The plane rotations in $V^{(d)}$ are introduced and used to construct the representations of the proper orthogonal group $O^{+}(d,\mathbb{C})$.

The Lorentz group is treated as a subgroup of O(4,$\mathbb{C}$). The general expression for the basic 2×2 irreducible representations $A(L_0)$ of the proper orthochronous Lorentz group $G(L_0)$ is obtained by direct reduction of the 4×4 spinor representation $S(L_0)$ by means of the basic lemma on the involutional transformations. It is completely parameterized by the angle and the axis of the spacial rotation and by the velocity of the pure Lorentz transformation. The finite dimensional irreducible representations of the Lorentz group $G(L)$ are discussed. The transformations of electro-magnetic field under $G(L)$ are discussed in the most general form.

1. INTRODUCTION

Recently, the author has developed a general theory of matrix transformations[1,2] which gives an explicit form of the transformation matrix which connects two square matrices A and B satisfying a given polynomial equation. In a special case where B is a diagonal matrix equivalent to A, the theory yields a general theory of matrix diagonalization. The theory takes a particularly simple form when the matrices A and B are involutional matrices, which are defined to satisfy a quadratic equation of the form[3]

$$x^2 = \text{const}\times \mathbf{1}, \qquad (1.1)$$

where $\mathbf{1}$ is the unit matrix. The Pauli spin matrices and Dirac γ-matrices[4] are simplest examples. The general $d\times d$ matrix solution of the equation has also been studied.[3] It has been recognized that involutional matrices have deep roots in various problem of mathematical physics.

We shall state the simplest special case of the transformation theory developed in Ref. 1 as a lemma, since it will be used throughout of this work.

Lemma: Let A and B be involutional matrices of a given order satisyfing $A^2 = B^2 = \mathbf{1}$. If their anticommutator is a c-number,

$$[A,B]_{+} = AB + BA = 2c\mathbf{1}, \quad c\neq -1, \qquad (1.2)$$

then,

(i) there exists an involutional transformation which interchanges A and B via

$$YAY = B;\ Y^2 = 1, \qquad (1.3)$$

where

$$Y = (A+B)/(2+2c)^{1/2}. \qquad (1.4)$$

(ii) There exists an additional transformation which connects A and B through a similarity transformation

$$A = V_{AB} B V_{AB}^{-1}, \qquad (1.5)$$

where

$$V_{AB} = AY = YB;\ V_{AB}^{-1} = V_{BA}. \qquad (1.6)$$

The direct proof of the Lemma is also very simple. When $c=-1$, it is obvious that A can be transformed into $-B$. A general condition[1] which allows to transform $-B$ into B has also been discussed. The effectiveness of the lemma in the Dirac theory of an electron can be seen from the fact that any linear combination of the generalized Dirac γ-matrices in d-dimensions is involutional. It has been shown[1] that almost all the existing transformations involved with the γ-matrices are special cases of the lemma. It is also noted here that there exists an involutional transformation which interchanges two commuting sets of spin like matrices (see 5.35).

In the present work, the lemma will be used to construct the spinor representations of the group of orthogonal transformations O(d,$\mathbb{C}$) in a d-dimensional Euclidean space $V^{(d)}$ over the complex field. Then the result will be specialized to the Lorentz group $G(L)$ which may be regarded as a subgroup of O(4,$\mathbb{C}$). It will be shown that the 4×4 spinor representation of the proper orthochronous Lorentz group is directly reduced to a direct sum of 2×2 irreducible representations by means of the lemma. This then gives the complete parametrization of the 2×2 representations. In this respect, the present approach is different from the ordinary methods which are based on the infinitesimal operators in constructing the irreducible representations.[5] Despite the numerous works on spinors by many workers, especially by Cartan,[6] by Brauer and Weyl,[7] by van der Waerden,[8] and by Bargman and Wigner,[9] it seems that there still exist some simple aspects of the representation theory of the Lorentz group which are not well recognized.

The basic transformation introduced in the present work may be called the axial involution $R(h)$ about a unit vector h in $V^{(d)}$ (see 2.6). It describes the inversion of the

 0022-2488/80/061299-13$1.00

$d-1$ dimensional subspace orthogonal to h; for example, it describes a coordinate transformation $x_i \to -x_i$ ($i \neq d$), $x_d \to x_d$. Thus, it is proper or improper according as d is odd or even. One may call $R(h)$ the two-fold rotation about h in $V^{(d)}$. The author, however, prefers the term "axial involution" over the term "two-fold rotation". The author recognized the basic importance of the axial involution $R(h)$ as a member of O(d,ℂ) from the fact that its spinor representation is given by the linear form of the vector defined by $\gamma_h = \Sigma h_\nu \gamma_\nu$ where $\gamma_1, \gamma_2 \cdots \gamma_d$ are the generalized Dirac γ-matrices.

The axial involution $R(h)$ followed by the total inversion describes a reflection $\bar{R}(h)$ in the hyperplane π_h orthogonal to the vector h of $V^{(d)}$. It is simply given by $\bar{R}(h) = -R(h)$, and always an improper rotation. It was Cartan who first recognized the basic importance of the reflection as a fundamental brickstone of the group of orthogonal transformations O(d,ℂ). His basic theorem may be stated as follows[10]: "Any proper (improper) rotation $R \in$ O(d,ℂ) is given by a product of an even (odd) number ($\leqslant d$) of reflections." Here a reflection means Cartan's reflection $\bar{R}(h)$ defined above. The basic importance of the theorem in the spinor representations is obvious. One can construct whole spinor representations of O(d,ℂ) from the spinor representation of $\bar{R}(h)$. In particular, the proper rotation group O$^+$(d,ℂ) in any dimensions or the rotation group in an even dimension may be represented by a product of axial involutions instead of Cartan's reflections. Cartan adapted a convention that γ_h represents $\bar{R}(h)$ which, however, does not leave the Dirac equation covariant.[6]

Following the Dirac procedure for a spinning electron generalized by Brauer and Weyl[7] we shall first define the spinor representations for O(d,ℂ). Then, based on the lemma we shall show that there exist two elementary transformations, an axial involution and a plane rotation, which transform a given unit vector in $V^{(d)}$ into another unit vector (Theorem 1, Sec. 2).

In Sec. 3, the mathematical properties of an axial involution $R(h)$ and its spinor representation $S(h)$ will be discussed. Based on Cartan's theorem we shall achieve the double valued spinor representation $S(R)$ of $R \in$ O(d,ℂ) by introducing the proper phase convention to $S(R)$. In Sec. 4, the general properties of the plane rotations in $V^{(d)}$ will be presented and used to construct the representations of the proper orthogonal groups O$^+$(d,ℂ) and O$^+$(d,ℝ) over the complex and real fields respectively.

In Sec. 5, the finite dimensional irreducible representations of the Lorentz group will be discussed. The direct reduction of a 4×4 spinor representation of the proper Lorentz group to a direct sum of two 2×2 irreducible representations gives the complete parametrization of the latter in terms of the angle and axis of the spacial rotation and the three components of the velocity of parallel translation. The transformation of the electromagnetic field[11] under the full Lorentz transformations will be discussed.

The theory developed here applies in any dimensions over the complex field in general even though important applications are on the Lorentz group. One of the points stressed in this work is the parametrization of the elementary transformations (axial involutions and plane rotations) by two vectors connected by the transformations. Another point is to emphasize the fundamental role played by involutional transformations in the theory of group representations. Further applications of the involutional transformation to physical problems will be presented in forthcoming papers.[12]

2. BASIC THEOREMS

Let O(d,ℂ) be the group of orthogonal transformations R,

$$x_\nu' = \sum_{\mu=1}^{d} R_{\nu\mu} x_\mu, \quad \nu = 1,2,\ldots,d, \tag{2.1}$$

of the d-dimensional Euclidean space $V^{(d)}$. We shall first operate within the continuum of all complex numbers until we apply the theory to the Lorentz group with certain restriction. The general Dirac procedure is to turn the scalar square $x^2 = \Sigma x_\nu^2$ of a vector $x \in V^{(d)}$ into the square of the linear form of x defined by

$$\gamma_x = x_1\gamma_1 + x_2\gamma_2 + \cdots + x_d\gamma_d, \tag{2.2}$$

where $\gamma_1, \gamma_2, \ldots, \gamma_d$ must satisfy the anticommutation relations,

$$[\gamma_\nu, \gamma_\mu]_+ \equiv \gamma_\nu\gamma_\mu + \gamma_\mu\gamma_\nu = 2\delta_{\nu\mu}, \quad \nu,\mu = 1,2,\ldots,d, \tag{2.3}$$

$\delta_{\nu\mu}$ being the Kronecker delta. Brauer and Weyl[7] gave the general Hermitian expression of the matrix representation of $\{\gamma_\nu\}$ which, however, may not be needed in the basic development. It is customary to use 1 for the unit matrix in the spinor space.

Under the influence of the orthogonal transformation R of (2.1), the linear form γ_x transforms according to

$$\gamma_{x'} = S_R \gamma_x S_R^{-1} \quad (S_R\text{, independent of } x) \tag{2.4}$$

or

$$\sum_\sigma \gamma_\sigma R_{\sigma\nu} = S_R \gamma_\nu S_R^{-1},$$

and a spinor $\psi(x)$ as a function of x transforms through $\psi'(x') = S_R\,\psi(x)$. The existence of the spinor transformation S_R which represents R has been shown by Brauer and Weyl. The general expression S_R due to Pauli exists for the Lorentz group.[13,14] This expression, however, is not very convenient to determine S_R for a given R.

It is evident from (2.4) that the transformation R is uniquely determined by a given S_R through

$$R_{\nu\mu} = \tfrac{1}{2}[\gamma_\nu, S_R\gamma_\mu S_R^{-1}]_+, \quad \nu,\mu = 1,2,\ldots,d, \tag{2.5}$$

while S_R is determined up to an arbitrary gauge factor for a given R. This should be kept in mind until we reduce the gauge factor up to $\pm$ signs by a proper phase convention. It is an immediate consequence of (2.5) that the linear form γ_h of a unit vector $h \in V^{(d)}$ represents an axial involution $R(h)$,

$$R(h) = 2hh - \mathbf{1}, \tag{2.6}$$

where hh is a symmetric tensor written in the diadic notation. Then the group property yields that any product $S(h_1,h_2,\ldots,h_n) = \gamma(h_1)\gamma(h_1)\cdots\gamma(h_n)$ where $\gamma(h_i)$ is the linear form of h_i, represents $R(h_1,h_2\cdots h_n) = R(h_1)R(h_2)\cdots R(h_n)$, which is an orthogonal matrix belonging to O(d,C) since each factor is an orthogonal matrix. According to Cartan's theorem stated in Sec. 1 this then completes the basic spinor representations of O(d,C) when $d =$ even. When $d =$ odd, we need the spinor representation of the total inversion U_i which, however, will be discussed in the next section.

In stating the basic theorem, a scalar product of two vectors u and v in $V^{(d)}$ is denoted by $(u\cdot v)$ ($= \Sigma\mu_\nu v_\nu$), and the second rank tensor constructed by two vectors u and v is denoted by uv using the diadic notation. The following theorem may be regarded as a geometric interpreation of the lemma introduced in Sec. 1.

Theorem 1: Let u and v be two unit vectors in $V^{(d)}$ with $(u\cdot v) \neq -1$, and h be a unit vector in the u-v plane which bisects the angle θ between u and v. Let γ_u, γ_v and γ_h be their respective linear forms with the set $\{\gamma_\nu\}$. Then there exist two elementary transformations which bring γ_v and γ_u via similarity transformations.

$$\gamma_u = S_i \gamma_v S_i^{-1}, \quad i = 1,2, \tag{2.7}$$

where

$$S_1 \equiv S(h) = \gamma_h, \tag{2.8}$$

$$S_2 \equiv S(u,v) = \gamma_u \gamma_h = \gamma_h \gamma_v. \tag{2.9}$$

Regarded as the spinor representations of the orthogonal transformations which bring v into u, $S(h)$ represents the axial involution $R(h)$ about h and $S(u,v)$ represents the plane rotation in the u-v plane given by two successive involutions,

$$R_2 \equiv R(u,v) = R(u)R(h) = R(h)R(v). \tag{2.10}$$

The proof is elementary, except for the geometric interpretation of R_2 as a plane rotation, which will be given separately in Sec. 4. To prove (2.7) we need the definition of the unit vector h given by

$$h = (u + v)/(2\cos(\theta/2)), \tag{2.11}$$

where $u^2 = v^2 = 1$ and the "angle" θ is defined by

$$\cos\theta = u\cdot v = \sum_{\nu=1}^{d} u_\nu v_\nu \neq -1. \tag{2.12}$$

It should be noted here that there exsits no ambiguity in the definition of $\cos\theta$ if one takes the same normalization constant for u and v as one should while h has double values for the given vectors u and v which affect S_1 and S_2 but not $R(h)$ and R_2. In the limit $u + v \to 0$, h is given by any unit vector which is orthogonal to u or v. With the definition of h, equation (2.7) follows immediately from the lemma since γ_u, γ_v and γ_h are all involutional. Then the group property of S_R leads to (2.10).

The orthogonal transformations $R_1 = R(h)$ and $R_2 = R(u,v)$ constitute a kind of a complete set in the sense that the second rank tensor uv is given by

$$uv = \tfrac{1}{2}[R(h) + R(u,v)], \tag{2.13}$$

which is rewritten in the form

$$R(u,v) = \mathbf{1} + 2(uv - hh), \tag{2.14}$$

and may be used to express the matrix elements of $R(u,v)$.

3. AXIAL INVOLUTIONS AND THE PHASE CONVENTION

A. Basic properties

We shall first discuss some of the elementary properties of the axial involution $R(h)$ of (2.6) represented by $S(h) = \gamma_h$, then propose a phase convention which may reduce the arbitrary gauge factors of the spinor representation $S(R)$ up to the signs.

We may quote (2.6) here for convenience,

$$2(h) = 2hh - \mathbf{1}. \tag{3.1}$$

Then, it is immediately obvious that $R(h)$ is an IOS (involutional, orthogonal and symmetric) matrix. Its characteristic roots are -1 and 1 with the degeneracies $d-1$ and 1, respectively. The trace and the determinant are

$$\mathrm{tr}R(h) = 2 - d, \quad \det R(h) = (-1)^{d-1}. \tag{3.2}$$

Accordingly, $R(h)$ is proper (improper) rotation when $d =$ odd (even). Cartan's reflection $\bar{R}(h)(= -R(h))$ in the hyperplane π_h orthogonal to the vector h is always an improper rotation, a point which is of the basic importance in Cartan's theorem stated in the introduction. It is evident from the definition that all the axial involutions (Cartan's reflections) belong to the same conjugate class of O(d,C).

If we introduce a new coordinate system by means of a complete set of orthonormalized vectors $h^{(1)}, h^{(2)},\ldots,h^{(d)}$ in $V^{(d)}$, we can define a set of axial involutions $\{R(h^{(j)})\}$ about the new coordinate axes. It can be easily shown that

$$(-1)^{p-1}\prod_{j=1}^{p} R(h^{(j)}) = (p-1)\mathbf{1} + \sum_{j=1}^{p} R(h^{(j)}), \tag{3.3}$$

regardless of the order of the factors on the left-hand side. Thus, all $R(h^{(j)})$ in the set commute with each other. In particular when $p = d$, we have

$$\prod_{j=1}^{d} \bar{R}(h^{(j)}) = -\mathbf{1}. \tag{3.4}$$

Accordingly the product of d-Cartan's reflections $\bar{R}(h^{(j)})$ describes the total inversion U_i, while the product of d-axial involutions $R(h^{(j)})$ describes U_i only when $d =$ even.

It is obvious that the above results may be described analogously by their spinor representations. Let $\{\gamma^{(j)}\}$ be γ-matrices referred to the new coordinate system defined by

$$\gamma^{(j)} = \sum_\nu h_\nu^{(j)}\gamma_\nu. \tag{3.5}$$

Then these satisfy the same set of anticommutation relations as (2.3),

$$\gamma^{(i)}\gamma^{(j)} = -\gamma^{(j)}\gamma^{(i)}, \quad i \neq j; (\gamma^{(j)})^2 = 1. \tag{3.6}$$

The anticommutation of $\gamma^{(i)}$ and $\gamma^{(j)} (i \neq j)$ corresponds to the commutation of $R(h^{(i)})$ and $R(h^{(j)})$. As a result, the double valuedness of the representation γ_h of $R(h)$ is essential. The products of $\{\gamma^{(j)}\}$ satisfy

$$\gamma^{(1)}\gamma^{(2)}\ldots\gamma^{(p)} = \frac{1}{p!}\sum_{i_1,i_2\cdots i_p} \alpha(i_1,i_2,\ldots,i_p)\gamma_{i_1}\gamma_{i_2}\cdots\gamma_{i_p} \tag{3.7}$$

where the coefficients $\alpha(i_1,i_2,\ldots,i_p)$ are the elements of a skew symmetric tensor composed of the components of the p vector$(h^{(1)},h^{(2)},\ldots,h^{(p)})$. In particular when $p=d$ with $d=2n$ or $2n+1$, we have

$$\pm\gamma^{(1)}\gamma^{(2)}\cdots\gamma^{(d)}=\gamma_1\gamma_2\cdots\gamma_d\equiv\iota\gamma_{d+1}, \qquad (3.8)$$

where ι is 1 or i according as n is even or odd. By definition, γ_{d+1} is involutional, $(\gamma_{d+1})^2=1$. The proper correspondence between (3.4) and (3.8) will lead us to the spinor representation of the total inversion U_i as well as the phase convention for the double valued representation.

When $d=2n$, γ_{d+1} anticommutes with all γ_ν and hence from (2.5) we may represent U_i by $g\gamma_{d+1}$ where g is a constant factor limited to ± 1 or $\pm i$, if $(g\gamma_{d+1})^2=\pm 1$.

When $d=2n+1$, γ_{d+1} commutes with all γ_ν so that it is a constant equal to ± 1. For definiteness we take a representation such that $\gamma_{d+1}=1$. Then we have

$$\pm\gamma^{(1)}\gamma^{(2)}\cdots\gamma^{(d)}=\gamma_1\gamma_2\cdots\gamma_d=\iota;\quad (d=2n+1). \qquad (3.9)$$

This identity reduces the order of the algebra engendered by the set $\{\gamma_\nu\}$ from 2^d to $2^{d-1}=2^n\times 2^n$ which enables us to introduce $2^n\times 2^n$ matrix representations for $\{\gamma_\nu\}$ as it has been achieved by Brauer and Weyl.[7] At the same time, it leads to the conclusion that there exists no member of the algebra which anticommutes with all $\{\gamma_\nu\}$. To remedy this, Brauer and Weyl simply extended the spinor representation by assigning ± 1 to the total inversion U_i, which commutes with all the members of the group.

Now, we can state in general ($d=$ even or odd) that the total inversion U_i is represented by $g\gamma_{d+1}$ with $g=\pm 1$ or $\pm i$, if $(g\gamma_{d+1})^2=\pm 1$.

B. The phase convention for O(d,ℂ)

Based on the foregoing arguments, we shall determine the phase convention of the spinor representations $S(R)\in O(d,\mathbb{C})$ which ensures the double valued representations such that

$$S(R_1R_2)=\pm S(R_1)S(R_2),\quad R_1,R_2\in O(d,\mathbb{C}), \qquad (3.10)$$

for any members of the group. This is simply achieved by assigning a proper phase for the spinor representations of a Cartan's reflection on account of Cartan's theorem stated in Sec. 1. In view of the identity (3.9) we propose the following convention:

$$\overline{R}(h)\leftrightarrow\pm\iota\gamma_{d+1}\gamma_h, \qquad (3.11)$$

where $\iota\gamma_{d+1}$ is defined by (3.8). Then, from (3.4) and (3.8), U_i and hence $R(h)$ are represented by

$$U_i\leftrightarrow\pm\iota\gamma_{d+1},\quad R(h)\leftrightarrow\pm\gamma_h\quad (d=2n),$$
$$U_i\leftrightarrow\pm 1,\quad R(h)\leftrightarrow\pm\iota\gamma_h\quad (d=2n+1). \qquad (3.12)$$

It should be noted here that the factor ι is essential only when $d=$ odd on account of the identity (3.9). Thus, it is possible to introduce an alternative convention for $d=$ even. It is also noted that when $d=$ even, the total inversion U_i is proper and can be given by a product of axial involutions. Thus, one can make the basic phase convention like (3.11) in terms of axial involutions instead of Cartan's reflections for this case.

The convention (3.11) is in accordance with the well-known convention in three dimensions that the reflection in the π_x plane and the two-fold rotation $R(x)$ about the x axis are represented by $\pm i\sigma_x$ where σ_x is a Pauli spin matrix. In four dimensions, Cartan's reflection $\overline{R}(h)$ and the axial involution $R(h)$ are represented by

$$\overline{R}(h)\leftrightarrow\pm\gamma_5\gamma_h,\quad R(h)\leftrightarrow\pm\gamma_h, \qquad (3.13)$$

and the total inversion U_i, the time inversion U_t, and the spacial inversion U_s are represented by

$$U_i\leftrightarrow\pm\gamma_5=\mp\rho_1,$$
$$U_t\leftrightarrow\pm\gamma_1\gamma_2\gamma_3=\pm i\rho_2, \qquad (3.14)$$
$$U_s\leftrightarrow\pm\gamma_4=\pm\rho_3,$$

provided that the first three coordinates describe space and the fourth describes time. These Dirac matrices $\rho_1, i\rho_2, \rho_3$ are all real in the Dirac standard representation. As a result, these commute with the Wigner time reversal operation $\tau(=-i\Sigma_2K$ with K being complex conjugation). Alternative conventions will be discussed with respcet to the Lorentz group.

C. Axial involutions and the phase convention of the Lorentz group G(L)

In the spirit of the present formalism the full Lorentz group G(L) may be regarded as a subgroup of O(4,ℂ) as it is frequently encountered.[11,14] It is defined by the group of orthogonal transformations which leave invariant the scalar square of a vector x in $V^{(4)}$ with three real space components x_1,x_2,x_3 and an imaginary time component $x_4=ict$ where c is the light velocity and t is the real time. Since x_1,x_2,x_3 and t are all real, the scalar square x^2 is positive or negative unless it is isotropic (i.e., $x^2=0$). The vector x is called space-like if $x^2>0$ and time-like if $x^2<0$. As a result, one can show from Theorem 1 that a unit vector h which defines an involutional transformation $R(h)\in G(L)$ must be either a space-like unit vector h^s or a time-like unit vector h^t defined by

$$h^s=(\mathbf{h}=\text{ real}, h_4=\text{ imaginary or zero}), h^s\cdot h^s=1, \qquad (3.15)$$
$$h^t=(\mathbf{h}=\text{ imaginary or zero}), h_4=\text{ real}), h^t\cdot h^t=1,$$

h being the respective spacial components. The unit vector h^t is not a Lorentz space vector owing to the normalization, which is introduced as a matter of convenience to use the lemma and Theorem 1 freely. Actually, three is no need for h^t to be a Lorentz space vector since it is introduced to describe the transformations of the Lorentz space vectors. The same comment applies for the unit vectors u and v in Theorem 1. The axial involutions corresponding to (3.15) are characterized by

$$R(h^s)_{44}\leqslant -1,\quad R(h^t)_{44}\geqslant 1, \qquad (3.16)$$

and hence $R(h^s)$ is antichronous and $R(h^t)$ is orthochronous. These may be called the space-like and the time-like axial involutions respectively. Analogously, $\overline{R}(h^s)$ and $\overline{R}(h^t)$ are called the space-like and time-like reflections respectively. It has been stated that all the axial involutions (reflections) belong to the same conjugate class of O(d,ℂ). In the Lorentz group G(L), however, the space-like and time-like axial involutions (reflections) belong to two different conjugate classes.

Following Cartan[15] one can show that any member of G(L) is given by a product of the space-like axial involutions $R(h^s)$ and the time-like involutions $R(h^t)$. Thus, it is possible to give different phase conventions for $R(h^s)$ and $R(h^t)$. There exist three other phase conventions besides (3.13) which ensures the double valued spinor representations of G(L). For convenience, we may summarize here all of them including (3.13):

$$\begin{aligned}&\text{Conv. 1}\quad R(h)\leftrightarrow \pm\gamma_h,\\ &\text{Conv. 2}\quad R(h)\leftrightarrow \pm i\gamma_h,\\ &\text{Conv. 3}\quad R(h^s)\leftrightarrow \pm\gamma_{h^s},\quad R(h^t)\leftrightarrow \pm i\gamma_{h^t},\\ &\text{Conv. 4}\quad R(h^s)\leftrightarrow \pm\gamma_{h^s},\ R(h^t)\leftrightarrow \pm i\gamma_{h^t}.\end{aligned} \tag{3.17}$$

Convention 1 has already been discussed. Convention 2 is also applicable to O(d,ℂ). Convention 3 is interesting since it is only one which gives the spinor representations which commute with the charge conjugation $C(=\gamma_2 K$ in the standard representation with K being complex conjugation). The rest of the conventions (1,2, and 4) give the representations which commute or anticommute with C. This should not create any difficulty since the spinor representations are double valued anyway. According to Convention 3, the total inversion U_i, the time inversion U_t and the spacial inversion U_s are represented by

$$U_i\leftrightarrow \pm i\rho_1,\quad U_t\leftrightarrow \pm i\rho_2,\quad U_s\leftrightarrow \pm i\rho_3. \tag{3.18}$$

These have previously been introduced based on unsatisfactory physical ground[16,17] that they should commute with the charge conjugation C. A convention similar to Convention 3 has also been proposed by Watanabe[18] in a pseudo-Euclidean space. Conventions 1–4 will be discussed again in connection with the higher-dimensional single-valued representations (see Sec. 5C). The author prefers Convention 1 for its simplicity and its general nature. Hereafter, this will be used unless otherwise stated.

D. Examples

For later use, let us consider some simple examples of the spinor representations $S(h)$ of $R(h)\in$G(L).

1. An axial involution about a space-like vector h in x_1-x_2 plane. Let

$$u=(\sin\theta,\cos\theta,0,0),\quad v=(0,1,0,0).$$

Then, from Theorem 1,

$$h=(\sin(\theta/2)\cos(\theta/2),0,0). \tag{3.19}$$

With a real θ, the vectors u, v and h are all space-like unit vectors in the Lorentz frame. We have

$$\begin{aligned}S(h)=\gamma_h&=\sin(\theta/2)\gamma_1+\cos(\theta/2)\gamma_2\\ &=\exp((\theta/2)\gamma_1\gamma_2)\gamma_2,\end{aligned} \tag{3.20}$$

$$R(h)=\begin{bmatrix}-\cos\theta & \sin\theta & 0 & 0\\ \sin\theta & \cos\theta & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\end{bmatrix}. \tag{3.21}$$

We did not put $\pm$ signs in $S(h)$ since γ_h is double valued anyway for a given $R(h)$. It is seen from (3.20) that the product $\gamma_h\gamma_2$ of two space-like involutions (reflections) describes a spacial rotation of the Lorentz frame.

2. An improper Lorentz transformation. Let $\hat{\mathbf{p}}=(\hat p_1,\hat p_2,\hat p_3)$ be a unit spacial vector and set

$$u=(\sin\theta\hat{\mathbf{p}},\cos\theta),\quad v=(0,0,0,1),$$

with an obvious abbreviation for the spacial components of u. We take any imaginary angle for θ such that

$$\theta=i\chi=i\tan^{-1}(v_0/c),$$

where v_0 is the velocity of the parallel translation of the moving coordinate frame along the direction of $\hat{\mathbf{p}}$. Then, the vectors u, v and h are all time-like unit vectors and

$$h=(i\sinh(\tfrac{1}{2}\chi)\hat{\mathbf{p}},\cosh(\tfrac{1}{2}\chi)). \tag{3.22}$$

Accordingly,

$$\begin{aligned}S(h)&=\gamma_h=\exp[-\tfrac{1}{2}\chi(\hat{\mathbf{p}}\cdot\boldsymbol{\alpha})]\gamma_4,\\ R(h)&=2hh-\mathbf{1},\end{aligned} \tag{3.23}$$

where $\alpha_j=-i\gamma_j\gamma_4, j=1,2,3$, and $\hat{\mathbf{p}}\cdot\boldsymbol{\alpha}=\hat p_1\alpha_1+\hat p_2\alpha_2+\hat p_3\alpha_3$. It is seen from (3.23) that a product of two time-like involutions $\gamma_h\gamma_4$ represents the pure Lorentz transformation, which will be discussed again in Sec. 5 by a more general consideration.

Further important examples of involutional transformations linear in γ matrices have been considered in the previous work[1] in connection with the transformation of the Dirac Hamiltonian for an electron, and the field Hamiltonians in solid state physics.

4. PLANE ROTATIONS IN d-DIMENSIONS

A. Basic properties

It has been stated in Theorem 1 that a product of two axial involutions given by $R(u,v)$ of (2.10) describes a plane rotation in the u-v plane which brings a unit vector v into another unit vector u,

$$u=R(u,v)\cdot v. \tag{4.1}$$

We shall now prove this statement from the fact that the spinor representation $S(u,v)$ of (2.9) and $R(u,v)$ of (2.10) can be written in the forms,

$$S(\theta,\omega)\equiv\pm S(u,v)=\pm\exp\left[\frac{i}{2}\theta\sum_{\nu<\mu}\omega_{\nu\mu}\sigma_{\nu\mu}\right], \tag{4.2}$$

$$\begin{aligned}R(\theta,\omega)\equiv R(u,v)&=\exp[\theta\omega]\\ &=\mathbf{1}+\omega\sin\theta+\omega^2(1-\cos\theta),\end{aligned} \tag{4.3}$$

where $\sigma_{\nu\mu}$ are the generalized Pauli spin matrices[7,19] and ω is a skew symmetric tensor with elements $\omega_{\nu\mu}$;

$$\sigma_{\nu\mu}=(\gamma_\nu\gamma_\mu-\gamma_\mu\gamma_\nu)/(2i),\ \omega_{\nu\mu}=(u_\nu v_\mu-v_\nu u_\mu)/\sin\theta,\quad \nu,\mu=1,2,\ldots,d.$$

It is more convenient to express ω in terms of the diadic notation,

$$\begin{aligned}&\omega=(uv-vu)/\sin\theta=\sum_{\nu<\mu}\omega_{\nu\mu}I^{(\nu,\mu)};\\ &I^{(\nu,\mu)}=i^{(\nu)}i^{(\mu)}-i^{(\mu)}i^{(\nu)};\ i_\lambda{}^{(\nu)}=\delta_{\lambda\nu},\\ &\nu,\mu,\lambda=1,2,\ldots,d,\end{aligned} \tag{4.4}$$

where $i^{(\nu)}$ is the unit vector in the νth coordinate axis, and $I^{(\nu,\mu)}$ describes the infinitesimal rotation in the $i^{(\nu)}$-$i^{(\mu)}$ plane, while ω describes the infinitesimal rotation in the u-v plane since ω is independent of θ as one can see immediately. Thus one can show that $S(\theta,\omega)$ and $R(\theta,\omega)$ represent a plane rotation in a plane defined by ω through an angle θ where θ can be complex in general.

The proof of (4.2) follows from (2.9) with use of the identities,

$$\gamma_u \gamma_v = (u\cdot v) + i \sum_{\nu<\mu} (u_\nu v_\mu - v_\nu u_\mu)\, \sigma_{\nu\mu}, \tag{4.5}$$

$$\Big(\sum_{\nu<\mu} \omega_{\nu\mu}\sigma_{\nu\mu}\Big)^2 = 1. \tag{4.6}$$

The proof of (4.3) can be achieved by substituting (4.2) into (2.4) or directly from (2.10) with use of the fact that the skew symmetric tensor ω of (4.4) satisfies

$$\omega^3 + \omega = 0, \tag{4.7}$$

$$\mathrm{tr}\omega^2 = -2, \tag{4.8}$$

$$\omega_{\nu\mu}\omega_{\kappa\lambda} + \omega_{\kappa\nu}\omega_{\mu\lambda} + \omega_{\mu\kappa}\omega_{\nu\lambda} = 0, \quad \nu,\mu,\kappa,\lambda \neq. \tag{4.9}$$

The last two equations follow directly from (4.4) or from (4.6) and the commutation relations of $\sigma_{\nu\mu}$.

The skew symmetric tensor ω is a familiar one in three dimensions,[20] eventhough it has never been parameterized by two unit vectors as in (4.4). One can easily show that the form of ω given by (4.4) is invariant for any two unit vectors on the u-v plane and hence is independent of θ except for the sign.[21] Thus, one may conclude that ω defines a plane. Further analytical properties of ω follows from (4.7) and (4.8). The characteristic roots of the secular determinant of ω are i, $-i$ and 0 with multiplicities 1, 1 and $(d-2)$ respectively. The multiplicity of the zero eigenvalues signifies that there exist $(d-2)$ independent directions normal to the ω plane in d-dimensions. Since (4.7) can be regarded as the reduced characteristic equation of ω with three distinct roots, there exist three projection operators[1], for which it is most convenient to take ω, $-\omega^2$, and $\mathbf{1}+\omega^2$. Their matrix ranks are 2, 2, and $(d-2)$, respectively. Let x be an arbitrary vector in $V^{(d)}$. Then, $x' = \omega\cdot x$, $x_\parallel = -\omega^2\cdot x$ and $x_\perp = (\mathbf{1}+\omega^2)\cdot x$ define a set of three orthogonal vectors where x' is a vector on the ω plane orthogonal to x, the vectors $x_\parallel$ and $x_\perp$ are the projections of x onto and normal to the ω plane. In a special case where x is an eigenvector of ω, x becomes self-orthogonal or isotropic. The geometric significance of these properties is evident: Suppose that the vector x transforms according to $R(\theta,\omega)$ of (4.3), i.e., $x = R(\theta,\omega)\cdot x_0$ with a constant vector x_0, then the normal component $x_\perp$ is invariant for the rotation and so is $(x_\parallel)^2 = (x')^2$. In fact, x', $x_\parallel$ and $x_\perp$ provide the rate, arm and axis of rotation for x respectively. It should be noted that the direction of $x_\perp$ depends on x except in three dimensions.

Next, we shall show that all the plane rotations defined by $R(\theta,\omega)$ through the same angle θ belong to the same conjugate class of $\mathrm{O}(d,\mathbb{C})$. Firstly, under an orthogonal transformation $O\in\mathrm{O}(d,\mathbb{C})$, $R(\theta,\omega)$ transforms as follows.

$$OR(\theta,\omega)\tilde{O} = R(\theta,\omega'), \tag{4.10}$$

where $\tilde{O}$ is the transpose of O and ω' defines the transformed plane of rotation,

$$\omega' = O\,\omega\tilde{O} = (u'v' - v'u')/\sin\theta = \sum_{\nu<\mu} \omega_{\nu\mu} I^{(\nu,\mu)'} = \sum_{\nu<\mu} \omega'_{\nu\mu} I^{(\nu,\mu)}, \tag{4.11}$$

with $u' = O\cdot u$, $v' = O\cdot v$, $i^{(\nu)'} = O\cdot i^{(\nu)}$. Conversely for a given pair of ω and ω', one can always find an orthogonal transformation which connects them (see below).

Frequently, it is necessary to bring the plane rotation $R(\theta,\omega)$ into its canonical form. To this end, let $e^{(1)}, e^{(2)}, \ldots, e^{(d)}$ be an orthonormalized complete set of vectors in $V^{(d)}$. The first two vectors are constructed from the column vectors of ω, then these are on the ω plane. The remaining vectors are constructed from the columns of the idempotent matrix $\mathbf{1}+\omega^2 = (= \mathbf{1} - e^{(1)}e^{(1)} - e^{(2)}e^{(2)})$. We choose $e^{(1)}$ and $e^{(2)}$ such that $e^{(2)} = -\omega\cdot e^{(1)}$, then

$$\omega = e^{(1)}e^{(2)} - e^{(2)}e^{(1)} \equiv \omega^{(1,2)}, \tag{4.12}$$

which is a special case of (4.4). Now, we define an orthogonal matrix O by its transpose $\tilde{O}$ in terms of the column vectors as follows

$$\tilde{O} = (e^{(1)}, e^{(2)}, \ldots e^{(d)});\ O_{\mu\nu} = e_\nu^{(\mu)}. \tag{4.13}$$

Then, we have the canonical form of ω by

$$O\omega^{(1,2)}\tilde{O} = I^{(1,2)}, \tag{4.14}$$

where $I^{(1,2)}$ represents the infinitesimal rotation on $i^{(1)}$-$i^{(2)}$ plane defined by (4.4). It is an immediate consequence of (4.14) that the orthogonal transformation O defined by (4.13) brings $R(\theta\omega^{(1,2)}) = \exp[\theta\omega^{(1,2)}]$ into a canonical form of a plane rotation,

$$C_{12}(\theta) \equiv O\exp[\theta\omega^{(1,2)}]\tilde{O} = \exp[\theta I^{(1,2)}] = \mathrm{diag}(\Phi(\theta), [1]^{d-2}), \tag{4.15}$$

where $[1]^{d-2}$ means that the diagonal element 1 appears $(d-2)$ times consecutively and

$$\Phi(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{4.16}$$

One may also rewrite the spinor representations $S(\theta,\omega^{(1,2)})$ in the form, analogous to (4.15),

$$S(\theta,\omega^{(1,2)}) = \exp[(i/2)\theta\sigma^{(1,2)}], \quad \sigma^{(1,2)} = -i\gamma^{(1)}\gamma^{(2)};\ \gamma^{(j)} = \sum_\nu e_\nu^{(j)}\gamma_\nu. \tag{4.17}$$

The analytical definition of a plane rotation in any dimensions may be given by a matrix which can be transformed into a canonical form of (4.15) through an orthogonal transformation. Then, one may conclude that any plane rotation can be written in the exponential form of (4.3) and hence its spinor representation is given by (4.2). The required skew symmetric tensor ω and the angle θ of a plane rotation R are determined from the following relations

$$2\sin\theta\omega = R - \tilde{R},\ \mathrm{tr}\,R = d - 2 + 2\cos\theta, \tag{4.18}$$

inverse to (4.3) and (4.8). The matrix $R = R(\theta,\omega)$ satisfies a cubic equation,

$$(R-1)(R-e^{i\theta})(R-e^{-i\theta}) = 0, \tag{4.19}$$

corresponding to (4.7). To obtain the factorized form of

$R(\theta,\omega)$ given by (2.10) as a product of two axial involutions, we construct two orthogonal unit vectors $e^{(1)}$ and $e^{(2)}$ from ω as in (4.12). Then, the required unit vectors u, v, and h for the factorization are given by

$$u = e^{(2)}\cos\theta + e^{(1)}\sin\theta, \quad v = e^{(2)},$$
$$h = e^{(2)}\cos(\theta/2) + e^{(1)}\sin(\theta/2). \tag{4.20}$$

We may summarize some of the results obtained in this section in the following theorem:

Theorem 2: A plane rotation in $V^{(d)}$ is represented by an orthogonal matrix R in $V^{(d)}$ whose reduced characteristic equation is cubic with three distinct roots, one of which is 1 with degeneracy $d-2$. It can be written in the exponential form of (4.3) or in the form of a product of two axial involutions.

This theorem is not trivial even in three dimensions, where any proper orthogonal matrix (real or complex) is a plane rotation. One can state that any orthogonal matrix in three dimensions (proper or improper, real or complex) can be brought into a canonical form by an orthogonal transformation. For later use we shall write down (4.2) in the special case of three dimensions, where some minor simplifications set in

$$S(\theta,\hat{\boldsymbol{\omega}}) = \pm\exp[(i/2)\theta\hat{\boldsymbol{\omega}}\cdot\boldsymbol{\sigma}],$$
$$R(\theta,\hat{\boldsymbol{\omega}}) = \exp[\theta,\hat{\boldsymbol{\omega}}\cdot\mathbf{I}]$$
$$= \mathbf{1}\cos\theta + \hat{\boldsymbol{\omega}}\cdot\mathbf{I}\sin\theta + \hat{\boldsymbol{\omega}}\,\hat{\boldsymbol{\omega}}(1-\cos\theta), \tag{4.21}$$

where

$$\hat{\omega}_i = \omega_{jk}, I_i = I^{(j,k)}, \; i,j,l \neq, \quad \text{cyclic.}$$

It is noted here that the unit vector $\hat{\boldsymbol{\omega}}$ may be defined alternatively by the vector product $\hat{\boldsymbol{\omega}} = [\mathbf{u}\times\mathbf{v}]/\sin\theta$ of the unit vectors $\mathbf{u}$ and $\mathbf{v}$ and also that $[\hat{\boldsymbol{\omega}}\cdot\mathbf{I}]\cdot\mathbf{x} = -[\hat{\boldsymbol{\omega}}\times\mathbf{x}]$.

B. The representations of orthogonal group O(d,$\mathbb{C}$)

We shall discuss some of the basic aspects of the group of orthogonal transformations in a d-dimensional vector space $V^{(d)}$ over the complex field based on the plane rotations developed in this section. For simplicity, we shall restrict the discussion to the proper orthogonal group $O^+(d,\mathbb{C})$ since any member of $O(d,\mathbb{C})$ is trivially related to $O^+(d,\mathbb{C})$.

From Cartan's theorem stated in Sec. 1 and Theorem 1 we can represent any member R of $O^+(d,\mathbb{C})$ by a product of a number ($\leqslant d/2$) of plane rotations. Let k th plane rotation be characterized by any angle (complex) θ_k, and a skew symmetric tensor $\omega^{(k)}$ given by two orthogonal vectors on the plane analogous to (4.12)

$$\omega^{(k)} = e^{(2k-1)}e^{(2k)} - e^{(2k)}e^{(2k-1)},$$
$$(e^{(2k-1)}\cdot e^{(2k)}) = 0, \quad k = 1,2,\ldots,[d/2], \tag{4.22}$$

where $[d/2]$ is the integral part of $d/2$. The set $\{e^{(\nu)}\}$ as a whole is not an orthogonal set in general except for $d=3$. It should be noted that $\{e^{(\nu)}\}$ depends on R. From (4.2) and (4.3) we can write for R and its spinor representation $S(R)$ as follows

$$R = \prod_{k=1}^{[d/2]} \exp[\theta_k\omega^{(k)}],$$

$$S(R) = \pm\prod_{k=1}^{[d/2]} \exp[(i/2)\theta_k\sigma^{(k)}], \tag{4.23}$$

where

$$\sigma^{(k)} = (1/2i)(\gamma^{(2k-1)}\gamma^{(2k)} - \gamma^{(2k)}\gamma^{(2k-1)}),$$
$$\gamma^{(j)} = \sum_{\nu=1}^{d} e_\nu{}^{(j)}\gamma_\nu,$$
$$[\gamma^{(i)},\gamma^{(j)}]_+ = (e^{(i)}\cdot e^{(j)}).$$

It is understood that a definite order of the factors in the products of (4.23) has been assumed since different $\omega^{(k)}$'s ($\sigma^{(k)}$'s) do not commute in general. In a special case of $d=3$, (4.23) reduces to (4.21).

In the case of the real orthogonal group O(d,$\mathbb{R}$) a further simplification sets in since the set $\{e^{(\nu)}\}$ can be taken to be an orthogonal set and all the ω's (σ's) commute with each other. Consequently (4.23) can be rewritten as

$$R = \exp\left\{\sum_{k=1}^{[d/2]}\theta_k\omega^{(k)}\right\},$$
$$S(R) = \pm\exp\left\{(i/2)\sum_{k=1}^{[d/2]}\theta_k\sigma^{(k)}\right\}, \tag{4.24}$$

where θ_k are real angles of rotations. Thus $S(R)$ is a unitary representation of R with a Hermitian set $\{\gamma_\nu\}$. It is also noted that R of (4.24) can be reduced into a product of canonical forms of plane rotations as given by (4.15) through an orthogonal transformation with the orthogonal matrix given in the form of (4.13). We shall next specialize these general results to the representations of the Lorentz group.

5. THE REPRESENTATION OF THE LORENTZ GROUP G(L)

A. Introduction

We shall describe the finite-dimensional irreducible representations of the Lorentz group G(L), in particular, the subgroup G(L_0) of the proper orthochronous Lorentz transformations L_0. The basic 2×2 irreducible representation $A(L_0)$ will be introduced by directly reducing the 4×4 spinor representation $S(L_0)$, or the Lorentz matrix L_0 using the lemma introduced in Sec. 1. This then gives the complete parameterization of $A(L_0)$ by the angle and axis of the spacial rotation and the three components of the velocity of the parallel translation of the Lorentz frame.

To begin with, let us consider a plane rotation belonging to the complex orthogonal group O(4,$\mathbb{C}$) in four dimensions. We set

$$\sigma_{ij} = \Sigma_k, I^{(i,j)} = I^{(k)}, \omega_{ij} = \omega_k; \quad i,j,k \neq; \quad 1,2,3 \text{ cylic,}$$
$$\sigma_{k4} = \alpha_k, I^{(k,4)} = J^{(k)}, \omega_{k4} = p_k, k = 1,2,3, \tag{5.1}$$

where $I^{(k)}$ and $J^{(k)}$ are 4×4 tensors of infinitesimal rotations. Then, from (4.2) and (4.3), the general plane rotation in $V^{(4)}$ is represented by

$$S(\theta,\omega) = \pm\exp[(i/2)\theta(\boldsymbol{\omega}\cdot\boldsymbol{\Sigma} + \mathbf{p}\cdot\boldsymbol{\alpha})],$$
$$R(\theta,\omega) = \exp[\theta(\boldsymbol{\omega}\cdot\mathbf{I} + \mathbf{p}\cdot\mathbf{J})], \tag{5.2}$$

where $\omega\cdot\mathbf{I} = \Sigma_k \omega_k I^{(k)}$, $\mathbf{p}\cdot\mathbf{J} = \Sigma_k p_k J^{(k)}$, and ω and $\mathbf{p}$ satisfy

$$\omega\cdot\omega + \mathbf{p}\cdot\mathbf{p} = 1, \ \omega\cdot\mathbf{p} = 0, \tag{5.3}$$

corresponding to (4.8) and (4.9).

When $\mathbf{p} = 0$, ω become a unit vector $\hat{\omega}$ and (5.2) represents a spacial rotation about the axis $\hat{\omega}$ through an angle θ. When $\omega = 0$, $\mathbf{p}$ becomes a unit vector $\hat{\mathbf{p}}$ and we can write the skew symmetric tensor ω in the form of (4.12)

$$\omega = \hat{\mathbf{p}}\cdot\mathbf{J} = e^{(p)}e^{(4)} - e^{(4)}e^{(p)}, \tag{5.4}$$

with

$$e^{(p)} = (\hat{p}_1, \hat{p}_2, \hat{p}_3, 0), \ e^{(4)} = (0,0,0,1).$$

Thus, (5.2) describes the plane rotation in the $e^{(p)} - e^{(4)}$ plane through an angle θ. We may specialize this to the Lorentz transformation in the space–time coordinate frame, x_1, x_2, x_3 and $x_4 = ict$ by introducing an imaginary angle of rotation,

$$\theta = i\chi = i \tanh^{-1}(v_0/c).$$

Then we arrive at the representation of the pure Lorentz transformation in the direction of the unit vector $\hat{\mathbf{p}}$,

$$S_{\rm Lor} = \pm \exp[-\tfrac{1}{2}\chi(\hat{\mathbf{p}}\cdot\boldsymbol{\alpha})] = \pm \gamma_h \gamma_4,$$
$$R_{\rm Lor} = \exp[i\chi(\hat{\mathbf{p}}\cdot\mathbf{J})] = R(h)R_4, \tag{5.5}$$

where use has been made of (2.10) and h is a time-like unit vector given by (3.22) and $R_4 = \text{diag}(-1,-1,-1,1)$. The factored forms of $S_{\rm Lor}$ and $R_{\rm Lor}$ are very convenient for the algebraic operations of the transformations. These coincide with the previous result (3.23).

Now, according to Cartan,[6] any member L_0 of the proper orthochronous Lorentz group $G(L_0)$ is given by a product of an even number ($\leqslant 2$) of space-like reflections (or axial involutions) and an even number ($\leqslant 2$) of time like reflections (or axial involutions). Alternatively, one can state that L_0 can be represented by a product of a spacial rotation and a pure Lorentz transformation.[5] Thus, we can write

$$S(L_0) = \pm \exp[(i/2)\theta(\hat{\omega}\cdot\boldsymbol{\Sigma})] \exp[-(1/2)\chi(\hat{\mathbf{p}}\cdot\boldsymbol{\alpha})],$$
$$L_0 = \exp[\theta(\hat{\omega}\cdot\mathbf{I})] \exp[i\chi(\hat{\mathbf{p}}\cdot\mathbf{J})], \tag{5.6}$$

where $\hat{\omega}$ and $\hat{\mathbf{p}}$ are three-dimensional unit vectors. These two expressions may be regarded as special cases of (4.23) and are the bases of the whole arguments which follow. For the simple explicit proof that (5.6) satisfies the group property, see Appendix.

In the theory of matrices,[20] the above form of $S \equiv S(L_0)$ is called the polar form of the matrix S since the first factor is unitary while the second factor is positively Hermitian. It is well known that the factorization is unique for any given nonsingular matrix. The two factors of S are essential in the sense that they cannot be reduced to a single plane rotation even as a member of O(4,C). Thus, the equivalence of two Lorentz transformations require the same set of angles θ and $i\chi$ as well as the same mutual orientation of $\hat{\omega}$ and $\hat{\mathbf{p}}$. The equality of the traces gives only the necessary condition for the equivalence. In a special case when $(\hat{\omega}\cdot\hat{\mathbf{p}}) = 0$, S equals two successive pure Lorentz transformations which leads to Einstein's addition formula for the velocities as well as the general expression for the accompanying spacial rotation, a limiting case of which gives the Thomas half (see Appendix). It is evident that the argument given above holds for the Lorentz matrix L_0 as well.

Some of the elementary properties of $S \equiv S(L_0) = S(\theta\hat{\omega}, \chi\hat{\mathbf{p}})$ which stem from (5.6) are as follows:

1. $\text{tr}S = \pm 4\cos(\theta/2)\cosh(\chi/2)$, $\det S = 1$; (5.7.1)

2. S commutes with the charge conjugation C and $\pm\rho_1$ which represents the space–time inversion U_i, (3.14);

3. Let S^τ be the time reversed S in the Wigner sense, S^* the complex conjugate, $\check{S}$ the contragradient, $\widetilde{S}$ the transverse, and $S^\dagger$ the Hermitian conjugate. Then, we have the following equivalences,

$$S^\tau = \Sigma_2 S^* \Sigma_2 = \rho_2 S \rho_2 = \rho_3 S \rho_3 = \check{S}^* = S(\theta\hat{\omega}, -\chi\hat{\mathbf{p}}). \tag{5.7.2}$$

Accordingly, $S^\dagger$, $\widetilde{S}$, and S^{-1} are also equivalent.

$$\Sigma_2 \widetilde{S} \Sigma_2 = \rho_2 S^\dagger \rho_2 = \rho_3 S^\dagger \rho_3 = S^{-1}. \tag{5.7.3}$$

The corresponding properties of $L_0 = L_0(\theta\hat{\omega}, \chi\hat{\mathbf{p}})$ are

1′. $\text{tr}\, L_0 = 4[\cos^2(\theta/2)\cosh^2(\chi/2) + \sin^2(\theta/2)\sinh^2(\chi/2)(\hat{\omega}\cdot\hat{\mathbf{p}})^2]$; (5.7.4)

2′. L_0 commutes with U_i;

3′. $\check{L}_0 = L_0$, and L_0 and L_0^* are equivalent:

$$U_t L_0 U_t = U_s L_0 U_s = L_0^* = L_0(\theta\hat{\omega}, -\chi\hat{\mathbf{p}}). \tag{5.7.5}$$

Accordingly, $L_0^\dagger$ and L_0^{-1} $(= \widetilde{L}_0)$ are also equivalent.

It is noted from (5.7.2) and (5.7.5) that L_0^* is represented by S^τ instead of S^* if one follows the basic correspondence $L_0 \leftrightarrow S(L_0)$. The properties given above are basic to understand the relations between the various subgroups of the full (homogeneous) Lorentz group $G(L)$. Let $G(L_{0,i})$, $G(L_{0,t})$ and $G(L_{0,s})$ be the proper, antichronous and orthochronous Lorentz groups consisting of L_0 combined with U_i, U_t and U_s, respectively. Then, from 2′ and 3′ it follows that $G(L_0)$ is an invariant subgroup of these subgroups as well as of $G(L)$. In particular, the property 3′ establishes the isomorphism of two subgroups $G(L_{0,t})$ and $G(L_{0,s})$. The representations of these subgroups and $G(L)$ will be discussed in Sec. 5.C.

B. The 2×2 irreducible representations of $G(L_0)$ based on $S(L_0)$

The spinor representation $S(L_0)$ of (5.6) is easily reduced if one uses the lemma introduced in Sec. 1. To this end, let us assume the standard representations for $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}$ given by Dirac and note the fact that $\alpha = \rho_1 \Sigma$. Let $\mathbf{1}_2$ be the 2×2 unit matrix. Then, the involutional transformation

$$Y_\rho = (1/\sqrt{2})(\rho_1 + \rho_3) = (1/\sqrt{2})\begin{pmatrix} \mathbf{1}_2 & \mathbf{1}_2 \\ \mathbf{1}_2 & -\mathbf{1}_2 \end{pmatrix}, \tag{5.8}$$

which interchanges ρ_1 and ρ_3, reduces $S(L)$ into a direct sum of a 2×2 matrix A and its time reversal A^τ

$$Y_\rho S(L_0) Y_\rho = \pm \begin{pmatrix} A & 0 \\ 0 & A^\tau \end{pmatrix}, \tag{5.9}$$

where

$$A = \exp[(i/2)\theta\hat{\omega}\cdot\boldsymbol{\sigma}] \exp[-\tfrac{1}{2}\chi\hat{\mathbf{p}}\cdot\boldsymbol{\sigma}] \equiv A(\theta\hat{\omega}, \chi\hat{\mathbf{p}}), \tag{5.10}$$

$$A^{\tau} = \sigma_2 A^* \sigma_2 = A(\theta\hat{\omega}, -\chi\hat{\mathbf{p}}). \tag{5.11.1}$$

Let $\check{A}$ be the contragradient matrix of A, then it follows that

$$\check{A} = \sigma_2 A \sigma_2, \quad A^{-1} = \sigma_2 \tilde{A} \sigma_2, \tag{5.11.2}$$

which are valid for any matrices belonging to SL(2,ℂ), the 2×2 unimodular group over the complex field with six parameters. The 2×2 matrices $\pm A$ or $\pm A^{\tau}$ exhaust all members of SL(2,ℂ) having six parameters and give the double valued irreducible representations of the proper orthochronous Lorentz group $G(L_0)$. The form of A explicitly parameterized by $\theta,\hat{\omega},\chi$ and $\hat{\mathbf{p}}$ has never been reported previously and gives the starting point of the whole finite dimensional irreducible representations of the full Lorentz group $G(L)$.

Traditionally,[8,20] the representation theory of the Lorentz group is based on $A \in SL(2,\mathbb{C})$ and its complex conjugate A^*, without giving the parametrized forms of A and A^*. According to (5.11.1), A^{τ} and A^* are equivalent so that $S(L_0)$ can directly be reduced into $A \oplus A^*$. We prefer, however, to base our arguments on A and A^{τ} because of the simplicity of the involutional transformation Y_ρ and also simple physical meaning of A^{τ}. It should be noted here that the involutional transformation Y_ρ is a real IOS matrix and it brings the four-component Dirac theory of a neutrino into the two component Weyl theory, i.e., $Y_\rho(\boldsymbol{\alpha}\cdot\mathbf{p})Y_\rho = \beta(\boldsymbol{\Sigma}\cdot\mathbf{p})$.

The parameterization of the matrix elements of A in terms of θ, $\hat{\omega}$, χ, and $\hat{\mathbf{p}}$ may explicitly be given as follows,

$$\begin{aligned} A &= r\mathbf{1}_2 + \mathbf{q}\cdot\boldsymbol{\sigma}, \\ r &= \cos(\theta/2)\cosh(\chi/2) - i(\hat{\omega}\cdot\hat{\mathbf{p}})\sin(\theta/2)\sinh(\chi/2), \\ \mathbf{q} &= i\hat{\omega}\sin(\theta/2)\cosh(\chi/2) - \hat{\mathbf{p}}\cos(\theta/2)\sinh(\chi/2) \\ &\quad + [\hat{\omega}\times\hat{\mathbf{p}}]\sin(\theta/2)\sinh(\chi/2). \end{aligned} \tag{5.11.3}$$

When $\chi = 0$, it reduces to the Euler–Olinde–Rodrigues parametrization,[6] and when $\theta = 0$, it reduces to the parametrization of the pure Lorentz transformation. In practical application, it may be better to use the Caley–Klein parametrization[14,22] for the part of the spacial rotation $\exp[i/2)\theta\hat{\omega}\cdot\boldsymbol{\sigma}]$. The determinant and the trace of A and A^{τ} are given by

$$\begin{aligned} \det A &= \det A^{\tau} = r^2 - \mathbf{q}^2 = 1, \\ \operatorname{tr} A &= 2r, \quad \operatorname{tr} A^{\tau} = 2r^*. \end{aligned} \tag{5.11.4}$$

Accordingly, A is not equivalent to A^{τ} in general. In a special case when $\operatorname{tr}A =$ real, A and A^{τ} are equivalent; in fact, $A^{\tau} = (\hat{\omega}\cdot\boldsymbol{\sigma})A(\hat{\omega}\cdot\boldsymbol{\sigma})$ with $(\hat{\omega}\cdot\hat{\mathbf{p}}) = 0$. According to Theorem 4 of Ref. 1, the 2×2 matrix A can be brought into a triangular form by a similarity transformation with an IUH (involutional, unitary and Hermitian) matrix. This property becomes important when we discuss the eigenwert problem of the higher dimensional representations of A.

In view of (5.9) the basis of the representation $A \oplus A^{\tau}$ of the proper orthochronous Lorentz group $G(L_0)$ is given by

$$\xi_\nu = Y_\rho \chi_\nu, \quad \nu = 1,2,3,4, \tag{5.12}$$

where Y_ρ is the involutional transformation and χ_ν is the Dirac spinor basis for the standard representation; $(\chi_\nu)_\mu = \delta_{\nu\mu}$. Explicitly,

$$\xi_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\\0\end{bmatrix}, \quad \xi_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\\0\\1\end{bmatrix}, \quad \xi_3 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\\0\end{bmatrix}, \quad \xi_4 = \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\\0\\-1\end{bmatrix}. \tag{5.13}$$

The symmetry properties of the basis are as follows: Under the charge conjugation C and Wigner time reversal τ, ξ_ν transforms as

$$\xi_1{}^C = \xi_4, \quad \xi_2{}^C = -\xi_3, \tag{5.14.1}$$

$$\xi_1{}^\tau = \xi_2, \quad \xi_3{}^\tau = \xi_4. \tag{5.14.2}$$

Under the space inversion U_s, the time inversion U_t and the total inversion U_i represented by $\pm\rho_3$, $\pm i\rho_2$ and $\pm\rho_1$ respectively it transforms as

$$\rho_3\xi_1 = i\rho_2\xi_1 = \xi_3, \quad \rho_3\xi_2 = i\rho_2\xi_2 = \xi_4, \tag{5.14.3}$$

$$\begin{aligned} &\rho_1\xi_1 = \xi_1, \quad \rho_1\xi_2 = \xi_2, \quad \rho_1\xi_3 = -\xi_3, \\ &\rho_1\xi_4 = -\xi_4. \end{aligned} \tag{5.14.4}$$

These transformations can be shown to be compatible with the transformation properties of $S(L_0)$ given by (5.7.2) and (5.7.4). As one can see, the charge conjugation, the space inversion and the time inversion mix two subspaces (ξ_1,ξ_2) and (ξ_3,ξ_4) belonging to A and A^{τ} respectively, while the Wigner time reversal mixes the bases in each subspace and the total inversion affects only the signs of ξ_ν. As a result, the 4×4 spinor representation $\{S(L_0), \pm\rho_3 S(L)\}$ of the orthochronous Lorentz group $G(L_{0s})$ is irreducible. This can be seen alternatively from the fact that it is impossible to bring more than four of the Dirac matrices ρ_1,ρ_2,ρ_3, and $\Sigma_1,\Sigma_2,\Sigma_3$ into the form of an even operator in the Dirac sense. A similar statement holds for the extended Lorentz groups $G(L_{0,t})$ and $G(L)$.

C. The finite dimensional representations of $G(L)$

Once 2×2 irreducible representation A is properly parametrized, the construction of the higher dimenisonal finite representations are a matter of routine, except for their relation with the phase convention. Let $\mathbf{r}$ be a vector with components, (x,y) and let

$$f_\nu^{(j)}(\mathbf{r}) = x^{2j-\nu}y^\nu/[(2j-\nu)!\nu!]^{1/2}, \quad \nu = 0,1,\ldots,2j, \tag{5.15}$$

be a set of $2j+1$ monomials of two variables x and y with j being an integer or a half integer. We define the $(2j+1)$ dimensional irreducible representation, $D^{(j)}(A) = D^{(j)}(\theta\hat{\omega},\chi\hat{\mathbf{p}})$ of $A \in SL(2,\mathbb{C})$ by

$$f_\nu^{(j)}(A\,\mathbf{r}) = \Sigma_\mu\, D^{(j)}(A)_{\nu\mu} f_\mu^{(j)}(\mathbf{r}). \tag{5.16}$$

This representation[3] previously proposed by the author is different from the conventional representation due to Wigner,[22] which is the inverse transpose of the present representation. Evidently, both representations are equivalent and connected by an involutional transformation since $\check{A} = \sigma_2 A \sigma_2$ from (5.11.2). The present representation $D^{(j)}(A)$ is more convenient in general since it does not require the calculation of A^{-1}. A convenient general expression of

$D^{(j)}(A)$ is available in terms of a hypergeometric function.[3] The symmetry properties, the eigenwert problem of $D^{(j)}(A)$ are also worked out. It has been shown that if A is triangular, then its representation $D^{(j)}(A)$ is also triangular in shape similar to A. Thus, in general, $D^{(j)}(A)$ can be brought into a triangular form by an IUH (involutional unitary and Hermitian) matrix.[1] We shall give here[3] the trace of $D^{(j)}(A)$

$$\mathrm{tr}D^{(j)}(A) = \begin{cases}(\epsilon_1^{2j+1} - \epsilon_2^{2j+1})/(\epsilon_1 - \epsilon_2), & \epsilon_1 \neq \epsilon_2,\\ (2j+1)\epsilon_1^{2j+1}, & \epsilon_1 = \epsilon_2,\end{cases} \tag{5.17}$$

where ϵ_1 and ϵ_2 are the characteristic roots of A. In special cases where $\theta = 0$ or $\chi = 0$, we have the well known result:

$$\mathrm{tr}D^{(j)}(A) = \begin{cases}\sin(j+\frac{1}{2})\theta/\sin(\theta/2), & \text{for } \chi = 0,\\ \sinh(j+\frac{1}{2})\chi/\sinh(\chi/2), & \text{for } \theta = 0.\end{cases}$$

It is also well known that $D^{(j)}(\pm A) = (\pm 1)^{2j} D^{(j)}(A)$.

The $(2j+1)$ dimensional representation of $A^{\tau} = A(\theta\hat{\omega}, -\chi\hat{\mathbf{p}})$ is given by $D^{(j)}(A^{\tau}) = D^{(j)}(\theta\hat{\omega}, -\chi\hat{\mathbf{p}})$. Accordingly, the general $(2j+1)\times(2j'+1)$ dimensional irreducible representation of $G(L_0)$ is given by

$$D^{(jj')}(L_0) = D^{(j)}(\theta\hat{\omega}, -\chi\hat{\mathbf{p}}) \otimes D^{(j')}(\theta\hat{\omega}, -\chi\hat{\mathbf{p}}). \tag{5.18}$$

Here it is noted again that this representation is different from the ordinary representations[8,23] which are based on A and A^*. Both representations are, however, equivalent. Thus, $\mathrm{tr}D^{(j)}(A^{\tau}) = [\mathrm{tr}D^{(j)}(A)]^*$ and hence the trace of $D^{(jj)}(L_0)$ is real and positive. For example,

$$\mathrm{tr}D^{(1/2\ 1/2)}(L_0) = 4rr^*, \tag{5.19}$$

where r is given by (5.11.3). From this result, $\mathrm{tr}L_0$ has been obtained and given by (5.7.4) since L_0 is equivalent to $D^{(1/2\ 1/2)}(L_0)$.

We may express the basis of the representation $D^{(jj')}$ in terms of the basis $\{\xi_\nu\}$ of $A \oplus A^{\tau}$ given by (5.15) formally as follows,

$$\Psi^{jj'}_{mm'} = \xi_1^{j+m}\xi_2^{j-m}\xi_3^{j'+m'}\xi_4^{j'-m'}/ \times[(j+m)!(j-m)!(j'+m')!(j'-m')!]^{1/2}, \tag{5.20}$$

$$m = j, j-1, \ldots, -j, \quad m' = j', j'-1, \ldots, -j'.$$

This formal expression is particularly convenient to discuss the representations of the full Lorentz group $G(L)$. Under the space–time inversion U_i, the time inversion U_t and the spacial inversion U_s represented by $\pm\rho_1$, $\pm i\rho_2$ and $\pm\rho_3$ respectively with Convention 1 given by (3.13), the basis vectors transforms as follows,

$$\begin{aligned}\Psi^{jj'}_{mm'} &\xrightarrow{\pm\rho_1} (-1)^{2j'}(\pm 1)^{2(j+j')}\Psi^{jj'}_{mm'},\\ \Psi^{jj'}_{mm'} &\xrightarrow{\pm i\rho_2} (-1)^{2j'}(\pm 1)^{2(j+j')}\Psi^{j'j}_{m'm},\\ \Psi^{jj'}_{mm'} &\xrightarrow{\pm\rho_3} (\pm 1)^{2(j+j')}\Psi^{j'j}_{m'm},\end{aligned} \tag{5.21}$$

where use has been made of (5.14.3) and (5.14.4). It follows immediately that the sets $\{\Psi^{jj}_{mm'}\}$ and $\{\Psi^{jj'}_{mm'}, \Psi^{j'j}_{m'm}; j \neq j'\}$ provide the bases of the irreducible representations $D^{(jj)}(L)$ and $D^{(jj'+j'j)}$ of $G(L)$ with the dimensions $(2j+1)^2$ and $2(2j+1)(2j'+1)$, respectively. The representations $D^{(jj)}(L)$ are always single valued while $D^{(jj'+j'j)}(L)$ are singled valued when and only when $j+j' =$ integers. According to (3.17) and (5.21), these single valued representations for U_i, U_t and U_s do depend on the arbitrary phase convention when $j+j' =$ odd [or $2j =$ odd for $D^{jj}(L)$]. This unexpected feature is easily understood from the fact that there exist additional irreducible[17,18,23] representations of $G(L)$ and the set of the single valued irreducible respresentations as a whole is invariant for the different choice of the phase conventions. We shall discuss this point in some detail, for it has never been properly recognized.

The additional irreducible representations of $G(L)$ stem from the fact that the four group V ($E =$ identity, U_i, U_t and U_s) is isomorphic to the factor group of $G(L)$ with respect to $G(L_0)$. The four one-dimensional representations D_0, D_1, D_2, D_3, and one double valued (2×2) representation D_4 of V are given by

	E	U_i	U_t	U_s
D_0	1	1	1	1
D_1	1	-1	-1	1
D_2	1	-1	1	-1
D_3	1	1	-1	-1
D_4	$\pm\mathbf{1}_2$	$\pm\sigma_1$	$\pm i\sigma_2$	$\pm\sigma_3$

(5.22)

where D_4 is obtained by reducing the 4×4 representation $(\mathbf{1}_4, \pm\rho_1, \pm i\rho_2, \pm\rho_3)$ of V with Convention 1 by the involutional transformation P of (5.35) which interchanges ρ_i and Σ_i. Each D_r defines an irreducible representation $D_r(L)$ of $G(L)$. These are trivial representations by themselves but their direct products with $D^{(jj)}(L)$ or $D^{(jj'+j'j)}(L)$ yield further representations of $G(L)$. Let us define a set of direct products,

$$D^{(\alpha,r)}(L) = D_r(L) \otimes D^{(\alpha)}(L), \tag{5.23}$$

where $r = 0, 1, 2, 3$, and $D^{(\alpha)}(L)$ stands for $D^{(jj)}(L)$ or $D^{(jj'+j'j)}(L)$ with $j+j' =$ integers. Then, these give four types of single valued irreducible representations for a given $D^{(\alpha)}(L)$. As it has been mentioned before $D^{(\alpha)}(L)$ depends on the phase convention introduced in (3.17) when $j+j' =$ odd [or $2j =$ odd for $D^{jj}(L)$]. However, this dependence gives a trivial consequence since different conventions give the same set of four types as a whole. In fact, one can show from (3.17) and (5.21) that

$$\begin{aligned}D^{(\alpha,r)}_{c-1}(L) &= D_3(L) \otimes D^{(\alpha,r)}_{c-2}(L)\\ &= D_2(L) \otimes D^{(\alpha,r)}_{c-3}(L) = D_1(L) \otimes D^{(\alpha,r)}_{c-4}(L),\end{aligned} \tag{5.24}$$

where $r = 0,1,2,3$, and the suffices $(c-1)$–$(c-4)$ denote Conventions 1–4.

It should be noted here that one can show from (5.21) and (5.23) that the four $D^{(jj,r)}(L)$ are all inequivalent while only two of $D^{(jj'+j'j,r)}$ are inequivalent for a given convention. For the latter the ones with $r = 0$ and 3 are equivalent and so are the ones with $r = 1$ and 2. This then leads to the equivalence of Conventions 1 and 2 and also of Conventions 3 and 4. It is also noted here that it is impossible to choose any single phase convention which makes all the basic representations $D^{(\alpha,0)}(L)$ equivalent to the regular tensor representations; for example, it can be shown that the 4×4 representation based on $(x_1,x_2,x_3,x_4 = ict)$ is equivalent to $D^{(1/2,1/2,0)}_{c-2}(L)$ while the 6×6 regular tensor representation is equivalent to $D^{(10+01,0)}_{c-4}(L)$ (see Sec. 5.E). Thus, one may conclude that there seems no essential difference between different choices of the phase conventions.

Finally, from the direct product of the 2×2 double valued representation $D_4(L)$ of (5.22) and $\{D^{(j,j)}(L), D^{(jj'+j'j)}(L); j\neq j'\}$ one obtains further double valued representations of G(L).

D. The Infinitesimal Method

There exists an alternative approach to the irreducible representations which is more traditional than the one presented hitherto. It is based on the generalized spin matrices $\mathbf{\Sigma}$ and $\boldsymbol{\alpha}$ corresponding to the infinitesimal operators **I** and **J** respectively. Let us define two commuting spin-like vector matrices **a** and **b** by

$$\mathbf{a} = \tfrac{1}{2}(\mathbf{\Sigma} + \boldsymbol{\alpha}) = \tfrac{1}{2}(1+\rho_1)\mathbf{\Sigma},$$
$$\mathbf{b} = \tfrac{1}{2}(\mathbf{\Sigma} - \boldsymbol{\alpha}) = \tfrac{1}{2}(1-\rho_1)\mathbf{\Sigma}. \tag{5.25.1}$$

Then, by definition, **a** and **b** are the projections of $\mathbf{\Sigma}$ belonging to the eigenvalues 1 and -1 of ρ_1 respectively.[2] Thus,

$$a_i b_j = b_j a_i = 0, \quad i,j = 1,2,3. \tag{5.25.2}$$

The involutional transformation Y_ρ of (5.8) connects **a** and **b** with the Pauli spin matrices as follows:

$$Y_\rho \mathbf{a} Y_\rho = \tfrac{1}{2}(1+\rho_3)\mathbf{\Sigma} = \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & 0 \end{pmatrix},$$
$$Y_\rho \mathbf{b} Y_\rho = \tfrac{1}{2}(1-\rho_3)\mathbf{\Sigma} = \begin{pmatrix} 0 & 0 \\ 0 & \boldsymbol{\sigma} \end{pmatrix}, \tag{5.25.3}$$

Hence, the set of simultaneous eigenvectors of $\mathbf{a}^2$, a_z, $\mathbf{b}^2$, and b_z are again given by $\{\xi_\nu\}$. One may easily extend this to the higher dimensional basis given by (5.20). This approach, however, may not lead to the parameterization of the basic 2×2 matrix A as given by (5.10).

E. Three-dimensional representation $R(3)\in$G(L_0)

As an example of the higher dimensional representation, we shall explicitly work out the 3×3 matrix representation $R(3)$ [$= R(3,L_0)$] of the proper Lorentz group, equivalent to $D^{(1,0)}(A)$. This is particularly important in the transformation of the field quantities of electricity and magnetism. Suppose that there exists a three dimensional vector $\mathbf{X} = (X_1,X_2,X_3)$ whose scalar square is invariant under a proper Lorentz transformation. Then one can write

$$X^2 = X_1^2 + X_2^2 + X_3^2$$
$$= (\sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3)^2 = \text{ invariant}, \tag{5.26}$$

where σ_ν are the Pauli spin matrices. Then a 3×3 orthogonal transformation $R(3)$ which leaves X^2 invariant must be related to the corresponding spinor transformation S_R by a special case of (2.4) for $d = 3$,

$$\sum_\nu \sigma_\nu R(3,L_0)_{\nu\mu} = S_R \sigma_\mu S_R^{-1}, \quad \mu = 1,2,3. \tag{5.27}$$

Substituting A of (5.10) into S_R of (5.27) we have

$$R(3;L_0) = \exp[\theta\hat{\omega}\cdot\mathbf{I}]\exp[i\chi\hat{\mathbf{p}}\cdot\mathbf{I}], \tag{5.28.1}$$

where $\mathbf{I} = (I^{(1)},I^{(2)},I^{(3)})$ describes the infinitesimal rotations in 3-D. More explicitly, with use of (4.21) one can write,

$$\mathbf{X}' = R(3,L_0)\cdot\mathbf{X}$$
$$= \exp[\theta\hat{\omega}\cdot\mathbf{I}]\;\{\mathbf{X}_\parallel + \cosh\chi(\mathbf{X}_\perp - i[\mathbf{v}_0\times\mathbf{X}])\}, \tag{5.28.2}$$

where $\mathbf{X}_\parallel$ and $\mathbf{X}_\perp$ are the parallel and perpendicular components of **X** to the direction $\hat{\mathbf{p}}$ of motion of the coordinate frame respectively and $\mathbf{v}_0 = c\tanh\chi\hat{\mathbf{p}}$.

Any antisymmetric tensor of second rank which transforms under the Lorentz transformation gives rise to a vector in 3-D which satisfies the invariance condition (5.26). The well known examples of such tensors[11] are provided by the field tensors in electricity and magnetism; e.g., $^2\mathbf{F} = (\mathbf{B}, -(i/c)\mathbf{E})$ and $^2\mathbf{G} = (\mathbf{H}, -ic\mathbf{D})$ where **B** the magnetic induction, **E** the electric field, **H** the magnetic field, and **D** the electric displacement. One may take $\mathbf{X} = \mathbf{B} - (i/c)\mathbf{E}$ or $\mathbf{H} - ic\mathbf{D}$. Then, (5.28.2) describes their transformations under the full Lorentz transformations **L**. To see this we separate the real and imaginary parts of (5.28.2) and describe the transformations of **B** and **E**, say, in terms of 6×6 matrix,

$$\begin{pmatrix} \mathbf{B}' \\ \mathbf{E}'/c \end{pmatrix} = \begin{pmatrix} R'(3) & R''(3) \\ -R''(3) & R'(3) \end{pmatrix}\begin{pmatrix} \mathbf{B} \\ \mathbf{E}/c \end{pmatrix}, \tag{5.28.3}$$

where $R(3,L_0) = R'(3) + iR''(3)$ and $R'(3)$ and $R''(3)$ are the real and imaginary parts respectively,

$$R'(3) = \exp[\theta\hat{\omega}\cdot\mathbf{I}]\,[\cosh\chi\,\mathbf{1}_3 + \hat{\mathbf{p}}\hat{\mathbf{p}}(1-\cosh\chi)],$$
$$R''(3) = \exp[\theta\hat{\omega}\cdot\mathbf{I}]\sinh\chi\,(\hat{\mathbf{p}}\cdot\mathbf{I}). \tag{5.28.4}$$

The transformation (5.28.3) is the most general form of the well known transformation law[11] of **B** and **E**: It can be shown that the 6×6 tensor of (5.28.3) is equivalent to $D^{(10+01,1)}_{c-1}(L)$ defined by (5.24) and hence the above transformation law of (5.28.3) holds for the full Lorentz group. It should be noted that if one takes A^τ of (5.11.1) for S_R of (5.27), one obtaines the time reversed transforamtion described by $R^\tau(3) = R(3)^*$. In a special case when $\hat{\omega}$ is parallel to $\hat{\mathbf{p}}$, $R(3,L_0)$ takes a particularly simple form

$$R(3,L_0) = \exp[(\theta + i\chi)\hat{\omega}\cdot\mathbf{I}]. \tag{5.28.5}$$

This special case has previously been derived by Kurşunoğlu.[24]

F. The 2×2 irreducible representations based on $R(L_0)$

It is well known that the Lorentz matrix L_0 given by (5.6) is irreducible. It is equivalent to $D^{(1/2,1/2)}(\theta\hat{\omega},\chi\hat{\mathbf{p}}) \equiv A\otimes A^\tau$ so that one may arrive at 2×2 irreducible representations A and A^τ by establishing the equivalence explicitly. To this end, we introduce two commuting sets of vector matrices $\mathbf{N} = (N_1,N_2,N_3)$ and $\mathbf{M} = (M_1,M_2,M_3)$ by the linear combinations of the infinitesimal transformations **I** and **J**, analogous to (5.25.1),

$$\mathbf{N} = -i(\mathbf{I} + \mathbf{J}) = (\rho_2\Sigma_1, -\rho_2\Sigma_3, \Sigma_2),$$
$$\mathbf{M} = -i(\mathbf{I} - \mathbf{J}) = (-\rho_1\Sigma_2, -\rho_2, \rho_3\Sigma_2), \tag{5.29}$$

where on the RHS the three components of **N** and **M** are expressed by the Dirac matrices in the standard representation. It can be easily shown that **N** and **M** commute with each other and are equivalent to the Dirac spin matrices $\mathbf{\Sigma}$

(see below). This property can also be used to construct all the finite dimensional irreducible bases of $G(L_0)$ by means of the ladder operator technique.[8]

In order to achieve the equivalence of L_0 to $A \otimes A^{\tau}$ we may rewrite L_0 in terms of **N** and **M**,

$$L_0 = \exp[\tfrac{1}{2}i\theta\hat{\omega}\cdot\mathbf{N}]\exp[-\tfrac{1}{2}\chi\hat{\mathbf{p}}\cdot\mathbf{N}]\exp[\tfrac{1}{2}i\theta\hat{\omega}\cdot\mathbf{M}] \times\exp[\tfrac{1}{2}\chi\hat{\mathbf{p}}\cdot\mathbf{M}]. \quad (5.30)$$

Then, using the basic lemma of Sec. 1 repeatedly we may bring **N** and **M** into $\boldsymbol{\Sigma}$ and $\boldsymbol{\rho}$ respectively through unitary transformations

$$T\mathbf{N}T^{\dagger} = \boldsymbol{\Sigma}, \quad T\mathbf{M}T^{\dagger} = \boldsymbol{\rho}, \quad (5.31)$$

where T is a unitary matrix defined by

$$T = 2^{3/2}\rho_3(\rho_2+\rho_3)\begin{pmatrix}\sigma_3 & 0\\ 0 & \mathbf{1}\end{pmatrix}(\Sigma_2+\Sigma_3)(\rho_2+\rho_3). \quad (5.32)$$

From (5.10), (5.30), and (5.31) and using that $\boldsymbol{\Sigma} = \mathbf{1}\otimes\boldsymbol{\sigma}$, $\boldsymbol{\rho} = \boldsymbol{\sigma}\otimes\mathbf{1}$, we establish the equivalence

$$TL_0T^{\dagger} = A^{\tau}\otimes A. \quad (5.33)$$

We can easily interchange A and A^{τ} in the direct product by a further transformation through a matrix P

$$PTL_0T^{\dagger}P = A\otimes A^{\tau}, \quad (5.34)$$

where P is an IUH (involutional, unitary, and Hermitian) operator defined by

$$P = \tfrac{1}{2}[\mathbf{1} + \boldsymbol{\rho}\cdot\boldsymbol{\Sigma}], \quad (5.35)$$

which interchanges $\boldsymbol{\rho}$ and $\boldsymbol{\Sigma}$ through $P\boldsymbol{\rho}P = \boldsymbol{\Sigma}$. The explicit form of the overall transformation matrix PT is remarkably simple

$$PT = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -i & 0 & 0\\ 0 & 0 & -1 & i\\ 0 & 0 & -1 & -i\\ -1 & -i & 0 & 0\end{bmatrix}, \quad (5.36)$$

which is unitary. Thus, from (5.34) and (5.36) the coordinate basis which transforms according to $A\otimes A^{\tau}$ is given by

$$PTx = (x_1 - ix_2, -x_3 + ix_4, -x_3 - ix_4, -x_1 - ix_2), \quad (5.37)$$

where x and the RHS should be read as column vectors. This basis could be obtained directly from the differential operators corresponding to **N** and **M** and their eigenwert problem. Starting from (5.34) and (5.37) one can reconstruct whole irreducible representations of $G(L_0)$ in the analogous manner as the classical work of Murnaghan.[20] This approach, however, may not lead to the parametrization of the 2×2 matrix A as given by (5.10).

6. CONCLUDING REMARKS

It has been shown that involutional transformations play a fundamental role in the theory of group representations on account of the lemma introduced in Sec. 1 and Cartan's basic theorem on the orthogonal transformations. The effectiveness of the lemma has been exhibited by the proof of the basic Theorem 1 and the reduction of 4×4 spinor representation of the Lorentz matrix itself (Secs. 5.B and 5.F). Further applications of the involutional transformations to physical problems will be presented in forthcoming papers. In the present work it has been stressed that any $d\times d$ orthogonal matrix can be expressed by a product of a number ($\leqslant d$) of IOS (involutional, orthogonal, and symmetric) matrices of a special type $\bar{R}(h) = \mathbf{1} - 2hh$ which Cartan called a reflection in a plane π_h perpendicular to the vector h. On acount of this theorem it was possible to introduce the phase convention for the spinor representation.

A reasonable extension of Cartan's theorem would be that a unitary matrix may be given by a product of IUH (involutional, unitary, and Hermitian) matrices. The general proof of this statement is lacking at present. However, there exists an interesting example given by the spinor representation $S(R)$ of (4.24) which represents $R\in O(d,\mathbb{R})$ since $\gamma_h = \Sigma\gamma_\nu h_\nu$ is an IUH matrix with the Hermitian set $\{\gamma_\nu\}$. Another interesting example of such factorization of a unitary matrix has been considered in the diagonalizations of the field Hamiltonians in solid state physics.[1]

APPENDIX: THE GROUP PROPERTY OF $G(L_0)$

It has been well established[5] that any element of the proper orthochronous Lorentz group $G(L_0)$ is given by a product of pure Lorentz transformation and a spacial rotation. A simple explicit proof of this statement based on (5.6), however, is worthwhile since it directly leads to Einstein's addition formula for velocities as well as the transformation property of the Lorentz transformation itself. For the proof it is sufficient to prove the following two statements:

1. The spacial rotation of a pure Lorentz transformation in a direction $\hat{\mathbf{p}}$ gives another pure Lorentz transformation in the direction $\hat{\mathbf{p}}'$ given by the spacial rotation of $\hat{\mathbf{p}}$:

$$\exp[\theta\hat{\omega}\cdot\mathbf{I}]\exp[i\chi\hat{\mathbf{p}}\cdot\mathbf{J}]\exp[-\theta\hat{\omega}\cdot\mathbf{I}] = \exp[i\chi\hat{\mathbf{p}}'\cdot\mathbf{J}],$$

$$\hat{\mathbf{p}}' = \exp[\theta\hat{\omega}\cdot\mathbf{I}(3)]\hat{\mathbf{p}}, \quad (\text{A1})$$

where $\mathbf{I}(3)$ is the three-dimensional vector matrix of an infinitesimal rotation. This is a special case of (4.10).

2. Two successive pure Lorentz transformations in the direction of $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_2$ give a pure Lorentz transformation followed by a spacial rotation in the $\hat{\mathbf{p}}_1$-$\hat{\mathbf{p}}_2$ plane:

$$\exp[i\chi_2\hat{\mathbf{p}}_2\cdot\mathbf{J}]\exp[i\chi_1\hat{\mathbf{p}}_1\cdot\mathbf{J}] = \exp[\theta\hat{\omega}\cdot\mathbf{I}]\exp[i\chi\hat{\mathbf{p}}\cdot\mathbf{J}], \quad (\text{A2})$$

where

$$\hat{\mathbf{p}}\sinh\chi = \hat{\mathbf{p}}_2\sinh\chi_2 + \hat{\mathbf{p}}_1\sinh\chi_1 \times[\cosh\chi_2 + \tanh(\chi_1/2)\sinh\chi_2\cos\phi],$$

$$\cosh\chi = \cosh\chi_1\cosh\chi_2 + \sinh\chi_1\sinh\chi_2\cos\phi, \quad (\text{A3})$$

$$\cos\phi = (\hat{\mathbf{p}}_2\cdot\hat{\mathbf{p}}_1),$$

$$\tan(\theta/2) = \tanh(\chi_2/2)\tanh(\chi_1/2)\sin\phi/ \times[1 + \tan(\chi_2/2)\tanh(\chi_1/2)\cosh\phi],$$

$$\hat{\omega} = [\hat{\mathbf{p}}_2\cdot\hat{\mathbf{p}}_1]/\sin\phi. \quad (\text{A4})$$

The proof is simplified if one uses the spinor representation of (A2). Equation (A3) is the Einstein formula. In the limiting case where only the direction of the transformation changes infinitesimally with respect to time, we have $\dot{\theta} = \dot{\phi}(\cosh\chi - 1)$ with $\chi_1 = \chi_2 = \chi$. This is essentially the well-known formula for the Thomas precession.[25]

ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to Professor E.P. Wigner at Princeton University for his comment on the previous work of the author (Ref. 3, 1969) which motivated the author to introduce the phase convention of the spinor representation based on the involutional transformations in the present work. The author is also grateful to Professor A.R. Ree at University of Utah for his encouragement of this work.

[1]Shoon K. Kim, J. Math. Phys. **20**, 2153 (1979), and references therein.
[2]Shoon K. Kim, J. Math. Phys. **20**, 2159 (1979).
[3]Shoon K. Kim, J. Math. Phys. **10**, 1225 (1969); **9**, 1705 (1968), and references therein.
[4]P.A.M. Dirac, *Quantum Mechanics*, 3rd edition Oxford U.P. Oxford, England, 1947).
[5]See, for example, I.M. Gel'fand, R.A. Minlos, Z.Ya. Shapiro, *Representations of the Rotations and Lorentz Groups and Their Applications* (Pergamon, New York, 1963), p. 158.
[6]E. Cartan, Bull. Soc. Math. France, **41**, 53 (1913), and *The Theory of Spinors*, English translation, (MIT Press, Cambridge, Massachusetts, 1966), and references therin.
[7]R. Brauer and H. Weyl, Am. J. Math. **57**, 425 (1935).
[8]Van der Waerden, Nachr. Ges. Wiss. Göttingen, Math-Physk. **100** (1929), and *Group Theory and Quantum Mechanics* (Springer-Verlag, New York, 1974), p. 114.
[9]V. Bargman and E.P. Wigner, Proc. Nat. Acad. Sci., **34**, 211 (1948).
[10]Ref. 6, *The Theory of Spinors*, p. 10.
[11]J.A. Stratton, *Electromagnetic Theory*, International Series in Physics, first edition (1941), p. 78.
[12]Shoon K. Kim, to be published.
[13]W. Pauli, Ann. Inst. Henri Poincaré **6**, 109 (1936).
[14]M.E. Rose, *Relativistic Electron Theory* (Wiley, New York, 1961).
[15]See Ref. 6, *The Theory of Spinors*, p. 14.
[16]See Ref. 14, p. 134, and S.S. Schweber, *Relativistic Quantum Field Theory* (Harper and Row, New York, 1962), pp. 78 and 107. See also J.J. Sakurai, *Advanced Quantum Mechanics* (Addition-Wesley Reading, Massachusetts, 1967), p. 100.
[17]J.M. Jauch and F. Rohrlich, *The Theory of Protons and Electrons* (Addison-Wesley Cambridge, Massachusetts (1955), pp. 87, and 441.
[18]S. Watanabe, Phys. Rev. **84**, 1008 (1951), Appendix.
[19]A. Joseph, Rev. Mod. Phys. **39**, 829 (1967).
[20]F.S. Murnaghan, *The Theory of Group Representations* (Johns Hopkins Press, Baltimore, Maryland, 1949), p. 237.
[21]More general expression for ω in terms of two linearly independent vectors **a** and **b** in the ω plane is given by $\omega = \pm(ab - ba)/(a^2b^2 - (a\cdot b)^2)^{1/2}$ which reduces to (4.4) when $a^2 = b^2 = 1$.
[22]E.P. Wigner, *Group Theory* (Academic, New York, 1959).
[23]V. Heine, Phys. Rev. **107**, 620 (1957), and *Group Theory in Quantum Mechanics* (Pergamon, Oxford, England 1960), p. 361. See also J.S. Lomont, *Applications of Finite Groups* (Academic, New York, 1959), p. 315.
[24]B. Kurşunoğlu, J. Math. Phys. **2**, 22 (1961), and *Modern Quantum Theory* (Freeman, San Francisco, 1962), p. 208.
[25]C. Møller, *The Theory of Relativity*, 2nd edition (Clarendon, Oxford England, 1972), p. 54.

Minimal and centered graded spin-extensions of the SL(3,R) algebra

Y. Ne'eman [a]
Tel Aviv University, Tel Aviv, Israel and University of Texas, Austin, Texas
Dj. Sijacki
Boris Kidric Institute of Nuclear Sciences, Belgrade, Yugoslavia, and Institute for Theoretical Physics, University of Cologne, Cologne, West Germany

(Received 5 December 1979; accepted for publication 14 January 1980)

We construct the minimal-size infinite-dimensional "supergauge" algebras containing sl(3,R) and spinorial generators, with the possibility of having a central term as in the Heisenberg algebra. The results describe the spin-excitations of a membrane and may provide [when extended to sl(4,R)] the framework for an infinitely supersymmetric (renormalizable?) gravity.

1. PHYSICAL MOTIVATION: MEMBRANES, LUMPS, AND THE RENORMALIZATION OF GRAVITY

In two previous publications[1,2] we have explored algebras containing both sl(3,R) and giso(3), the graded extension of the inhomogeneous orthogonal group ISO(3). Since the sl(n,R) have no finite double-valued representations, we found no finite solutions.[1] In Ref. 2 we presented a minimal solution gsl(3,R), involving in addition to sl(3,R) two infinite sets of generators S^j_m and E^j_m behaving under sl(3,R) as the irreducible bandors, $\mathscr{D}(\frac{1}{2};0)$ and $\mathscr{D}(1;\sigma_2)$, respectively, and with

$$[S^j_m, S^{j'}_{m'}]_+ \subset E^{j+j'}_{m+m'}, \quad [S^j_m, E^{j'}_{m'}] = 0, \quad [E^j_m, E^{j'}_{m'}] = 0. \tag{1.1}$$

The interest in such algebra relates to several possible applications. In dual models,[3] although the usual presentation proceeds to "replace" the shears of SL(2,R)/SO(2) by two such infinite sets, it can be checked that the resulting NSR algebra together with the (z,t) shears is just the analogous gsl(2,R), with the infinite sequences behaving as irreducible representations of sl(2,R). This in fact explains their composition, otherwise an *ad hoc* result. Our gsl(3,R) is thus the spin-excitation algebra of a membrane and gsl(4,R) is that of a lump or bag.[4]

In Ref. 2 we showed how gsl(3,R) $\rightarrow$ giso(3) + decoupled operators, under a certain contraction procedure. The lowest levels $S^{\frac{1}{2}}_m$ and E^1_m of the S^j_m and E^j_m can then be identified with the supersymmetry generators and with the translations, respectively. This result opens up another possibility: gauging gsl(4,R) locally. The graviton (or its translation-gauge part) is the gauge field of E^1_m E^j_m. It will thus be accompanied by gauge fields with spin 3, 5, 7, etc. Similarly, the $j=\frac{3}{2}$ gauge field of supergravity[5] is followed by fields with $j=\frac{7}{2},\frac{11}{2},\frac{15}{2}$, etc. Such a modification of gravity may be related to the $\alpha \rightarrow 0$ limit in dual models,[6] where a graviton-like field has been shown to appear.

We have recently studied metric-affine gravity with spinorial matter[7] and a framework for affine gravity.[8] This is motivated by the regularities displayed by hadrons, which correspond to approximate hypermomentum [sl(4,R) currents] conservation and by the existence of (hopefully microscopic) confining solutions[9] perhaps related to quark confinement. In supergravity remarkable progress has been achieved in renormalizing gravity: The theory is finite at the one-loop level and the same result would hold for two-loop diagrams, except that some uncertainty remains owing to anomalies. On the other hand, for the three-loop level and above, examples of counterterms that are not forbidden by supersymmetry have been constructed.[10]

Gauging gsl(4,R) may combine the advantages of both approaches: hypermomentum quasiconservation, confinement, and renormalizability. Moreover, with the single anticommutation of supersymmetry now replaced by an infinite set of such commutators, it may be possible to forbid all counterterms to all orders!

This is an outline for a full research program. In this article we provide a detailed discussion of gsl(3,R), following Refs. 1 and 2. We include a central term in the algebra and also check the possibility of having $\mathscr{D}(0;\sigma_2)$ instead of $\mathscr{D}(1;\sigma_2)$.

2. SL(3,R) ALGEBRA AND UNITARY IRREDUCIBLE REPRESENTATIONS

SL(3,R) is the group of volume-preserving deformations, i.e., the group of linear unimodular transformations in a three-dimensional real vector space. The group is a simple and noncompact Lie group. The group multiplication law is given as a product of transformations, i.e., as matrix multiplication when the group elements are given as 3×3 matrices. The maximal compact subgroup of SL(3,R) is SO(3). In quantum theory one is interested not only in single-valued representations of a symmetry group, but also in projective or ray representations (representations up to a factor of modulus one), i.e.,

$$U(g_1)U(g_2) = e^{i\omega(g_1,g_2)}U(g_1g_2), \tag{2.1}$$

g_1 and g_2 being the group elements and ω a real function. This can be achieved by considering only the single-valued representations of the corresponding covering group. It follows from the work of Bargmann[11] that there is a one-to-one

[a] Research supported in part by the United States Israel Binational Science Foundation.

correspondence between the ray representations of a group and the single-valued representations of its universal covering group. We denote the universal covering group of $SL(3,R)$ by $\overline{SL}(3,R)$. This group has the same Lie algebra as $SL(3,R)$ and its maximal compact subgroup is $SU(2)$, the covering group of $SO(3)$. The center of $\overline{SL}(3,R)$ is a two-element group Z_2, and the factor group of $\overline{SL}(3,R)$ with respect to Z_2 is isomorphic to $SL(3,R)$ i.e.,

$$\overline{SL}(3,R)/Z_2 \simeq SL(3,R). \tag{2.2}$$

In the following we shall consider $\overline{SL}(3,R)$ only. The representations of $SL(3,R)$ form a subset of those of $\overline{SL}(3,R)$, for which angular momentum can take only integer values.

Let $sl(3,R)$ be the Lie algebra of the $SL(3,R)$ and thus also of the $\overline{SL}(3,R)$ group. It is an algebra of real 3×3 traceless matrices. We denote by J_i $(i = 1, 2, 3)$ the $SU(2)$ generators and they constitute the angular momentum part of $sl(3,R)$. The remaining five generators are noncompact and they form with respect to $su(2)$ a second-rank irreducible tensor operator T_{ij}. The $sl(3,R)$ algebra is now given by the following commutation relations:

$$\begin{aligned}
&[J_i,J_j] = i\epsilon_{ijk}J_k \quad (i,j,\cdots = 1,2,3),\\
&[J_i,T_{jk}] = i\epsilon_{ijm}T_{mk} + i\epsilon_{ikn}T_{jn},\\
&[T_{ij},T_{kl}]\\
&\quad = -i(\delta_{ik}\epsilon_{jlm} + \delta_{il}\epsilon_{jkm} + \delta_{jk}\epsilon_{ilm} + \delta_{jl}\epsilon_{ikm})J_m.
\end{aligned} \tag{2.3}$$

It is convenient to write the $\overline{SL}(3,R)$ generators in a spherical basis. The generators are now given (in the matrix form) by

$$J_0 = \begin{pmatrix} 0 & -i & 0\\ i & 0 & 0\\ 0 & 0 & 0\end{pmatrix}, \quad J_\pm = \begin{pmatrix} 0 & 0 & \mp 1\\ 0 & 0 & -i\\ \pm 1 & i & 0\end{pmatrix},$$

$$T_0 = -i\sqrt{2/3}\begin{pmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -2\end{pmatrix},$$

$$T_\pm = \begin{pmatrix} 0 & 0 & \mp i\\ 0 & 0 & 1\\ \mp i & 1 & 0\end{pmatrix}, \quad T_{\pm 2} = \begin{pmatrix} i & \mp 1 & 0\\ \mp 1 & -i & 0\\ 0 & 0 & 0\end{pmatrix}. \tag{2.4}$$

The minimal set of the $sl(3,R)$ commutation relations in the spherical basis is

$$\begin{aligned}
&[J_0,J_\pm] = \pm J_\pm,\\
&[J_+,J_-] = 2J_0,\\
&[J_0,T_\mu] = \mu T_\mu, \quad (\mu = 0, \pm 1, \pm 2),\\
&[J_\pm,T_\mu] = [6-\mu(\mu\pm 1)]^{1/2}T_{\mu\pm 1},
\end{aligned} \tag{2.5}$$

and

$$[T_{+2},T_{-2}] = -4J_0. \tag{2.6}$$

The commutation relation (2.6) is known as "the $sl(3,R)$ condition," while the remaining unspecified commutation relations can be obtained by means of the Jacobi identity.

The $\overline{SL}(3,R)$ group does not have any finite-dimensional spinorial representations, while the finite-dimensional tensorial representations can be easily obtained from the unitary representations of the $SU(3)$ group in the $SO(3)$ basis. Here spinor and/or tensor is defined according to the angular momentum content. There are infinite-dimensional spinorial representations of $\overline{SL}(3,R)$ and all linear continuous spinorial and tensorial unitary irreducible representations (unirreps) have been explicitly constructed.[11,12] The angular momentum content of the simplest spinorial unirrep of $\overline{SL}(3,R)$ features the striking $\Delta j = 2$ rule. This representation belongs to the class of the multiplicity-free or "ladder" representations. They contain in the reduction to the maximal compact subgroup $SU(2)$ each of the corresponding representations at most once. In this work we will confine ourselves to the ladder unirreps, which have been obtained by various methods.[12-14] For the sake of completeness we present here a new and rather short construction of the ladder unirreps of $\overline{SL}(3,R)$ and provide the corresponding matrix elements of the group generators, which are necessary in order to construct explicitly the desired graded spin-extension of the $sl(3,R)$ algebra. The construction is carried out by making use of the decontraction formula[15] in evaluating the form of the $\overline{SL}(3,R)$ generators. This formula, by now rather well-known in physics, describes a deformation which is the inverse of the Wigner–Inönü group contraction.[16] We work in the homogeneous space of functions of the parameters of the maximal compact subgroup $SU(2)$. The Wigner–Inönü contraction of the $sl(3,R)$ algebra with respect to the $su(2)$ subalgebra consists in defining the new set of (noncompact) generators $Q_\mu(\epsilon) = \epsilon T_\mu$ and taking the limit $\lim_{\epsilon\to 0}Q_\mu(\epsilon) = Q_\mu$. The corresponding group, obtained in this way from the $\overline{SL}(3,R)$ group, is $T_5 \otimes\!\!\!\!\times \overline{SU}(2)$, i.e., a semidirect product of the Abelian five-parameter group T_5 and the $SU(2)$ group. The Q_μ operators transform as components of the $SU(2)$ quadrupole tensor operator. The commutation relations of the contracted group are given by (2.5), where T_μ is replaced by Q_μ and Eq. (2.6) now reads

$$[Q_\mu,Q_\nu] = 0, \quad (\mu,\nu = 0, \pm 1, \pm 2). \tag{2.7}$$

The decontraction formula tells us that the following operators

$$T_\mu = \sqrt{2/3}\,(\sigma_2\tilde{Q}_\mu + \tfrac{1}{2}i[J^2,\tilde{Q}_\mu]), \tag{2.8}$$

where

$$\tilde{Q}_\mu = (Q\cdot Q)^{-1/2}Q_\mu,$$

together with J_0 and $J_\pm$, satisfies (2.5) and (2.6), thus generating the $\overline{SL}(3,R)$ group. The parameter σ_2 is an arbitrary real number and the normalization factor $\sqrt{2/3}$ is fixed by the value of the structure constant in the $sl(3,R)$ condition (2.6). We follow the notation of Ref. 12.

The unirreps of the $SU(2)$ group are well known. They are characterized by the label j $(j = 0, \frac{1}{2}, 1, \frac{3}{2},\cdots)$ and in the spherical basis the matrix elements of the compact generators are

$$J_0\left|\begin{matrix} j\\ m\end{matrix}\right\rangle = m\left|\begin{matrix} j\\ m\end{matrix}\right\rangle, \tag{2.9}$$

$$J_{\pm}\left|\begin{matrix} j \\ m \end{matrix}\right\rangle = [(j\mp m)(j\pm m+1)]^{1/2}\left|\begin{matrix} j \\ m\pm 1 \end{matrix}\right\rangle. \quad (2.9)$$

In this basis $J^2 \to j(j+1)$. The vectors of the orthonormal basis $\{|{}^{j}_{m}\rangle\}$ are related to the D functions by

$$\left\langle \beta,\gamma \left| \begin{matrix} j \\ m \end{matrix}\right.\right\rangle = (2j+1)^{1/2} D^{j}_{0m}(\beta,\gamma), \quad (2.10)$$

when β and γ are the Euler angles.

Since Q_μ operators commute mutually, Eq. (2.7), they do not contain derivatives in the Euler angles and they are obviously given, in the spherical basis, by

$$Q_\mu = qD^{2}_{0\mu}(\beta,\gamma), \quad q\in R, \quad \mu = 0, \pm 1, \pm 2. \quad (2.11)$$

The second-order Casimir operator of the $T_5 \otimes SU(2)$ group is

$$Q\cdot Q = \sum_{\mu=-2}^{2} (-)^{\mu} Q_\mu Q_{-\mu} \to q^2. \quad (2.12)$$

The corresponding SL(3,R) generators in the spherical basis now read

$$T_\mu = \sqrt{2/3}\,(\sigma_2 D^{2}_{0\mu} + \tfrac{1}{2}i[J^2, D^{2}_{0\mu}]). \quad (2.13)$$

The unitarity of the group representations, i.e., the hermiticity of the generators, yields $\sigma_2 \in R$. The matrix elements of the noncompact operators T can now be directly read off in the $\{|{}^{j}_{m}\rangle\}$ basis, i.e.,

$$\left\langle \begin{matrix} j' \\ m' \end{matrix}\right| T_\mu \left|\begin{matrix} j \\ m \end{matrix}\right\rangle = (-)^{j'-m'}\begin{pmatrix} j' & 2 & j \\ -m' & \mu & m \end{pmatrix}\langle j'\|T\|j\rangle, \quad (2.14)$$

$$\langle j'\|T\|j\rangle = (-)^{j'}\sqrt{2/3}\,[(2j'+1)(2j+1)]^{1/2}\{\sigma_2 + \tfrac{1}{2}i[j'(j'+1) - j(j+1)]\}\begin{pmatrix} j' & 2 & j \\ 0 & 0 & 0 \end{pmatrix}.$$

The second-order Casimir operator of the $\overline{\mathrm{SL}}(3,R)$ group now takes (for the ladder unirreps) the following value:

$$C_2 = J^2 - \tfrac{1}{2}T\cdot T \to -3 - \tfrac{1}{3}\sigma_2. \quad (2.15)$$

The $3j$ coefficient in $\langle j'\|T\|j\rangle$ is different from zero if $j' = j$, $j\pm 2$, for integer j. Therefore there are two classes of tensorial unirreps of sl(3,R) that are characterized by the minimal j value and an arbitrary real parameter σ_2, and the corresponding j content is given by

$$\mathscr{D}(0;\sigma_2): \quad \{j\} = \{0,2,4,\cdots\},$$
$$\mathscr{D}(1;\sigma_2): \quad \{j\} = \{1,3,5,\cdots\}. \quad (2.16)$$

The nonzero reduced matrix elements are explicitly

$$\langle j\|T\|j\rangle = -\sqrt{2/3}\,\sigma_2\left(\frac{j(j+1)(2j+1)}{(2j-1)(2j+3)}\right)^{1/2},$$
$$\langle j+2\|T\|j\rangle = [(2j+3)(j+1)(j+2)]^{1/2} \times [i + \sigma_2/(2j+3)], \quad (2.17)$$
$$\langle j-2\|T\|\rangle = -[j(j-1)(2j-1)]^{1/2} \times [i - \sigma_2/(2j-1)].$$

We can now consider these matrix elements to be valid in the half-integer case as well and check whether the sl(3,R) condition is satisfied. It turns out, after some straightforward algebra, that the sl(3,R) condition is valid provided that $\sigma_2 = 0$ and the minimal j is $\frac{1}{2}$. Thus there exists a single spinorial unirrep of sl(3,R) with the following j content:

$$\mathscr{D}(\tfrac{1}{2};0): \quad \{j\} = \{\tfrac{1}{2},\tfrac{5}{2},\tfrac{9}{2},\cdots\}, \quad (2.18)$$

and the reduced matrix elements are

$$\langle j\|T\|j\rangle = 0,$$
$$\langle j+2\|T\|j\rangle = i[(2j+3)(j+1)(j+2)]^{1/2}, \quad (2.19)$$
$$\langle j-2\|T\|j\rangle = -i[j(j-1)(2j-1)]^{1/2}.$$

3. GRADED ALGEBRA CONSTRUCTION: CASE A.

We turn now to the explicit construction of a minimal graded sl(3,R) algebra, i.e., gsl(3,R). It is minimal in the sense that, as the Neveu–Schwartz–Ramond algebra, besides the spinor operators it contains only one set of irreducible tensor operators. gsl(3,R) is an infinite-dimensional algebra with generators J_0, $J_\pm$, T_μ ($\mu = 0, \pm 1, \pm 2$), S^{j}_{m} ($j = \frac{1}{2},\frac{5}{2},\frac{9}{2},\cdots$; m takes all half-integer values $|m|\leqslant j$), and E^{j}_{m} ($j = 1,3,5,\cdots$; m takes all integer values $|m|\leqslant j$). The generators J_0, $J_\pm$, and T_μ form sl(3,R) itself, while S^{j}_{m} and E^{j}_{m} transform as components of sl(3,R) irreducible tensor operators corresponding to $\mathscr{D}(\frac{1}{2};0)$ and $\mathscr{D}(1;\sigma_2)$, respectively. The graded Lie brackets are given [we allow the appearance of a central term in Eq. (3.1)] in addition to those of sl(3,R) [Equations (2.5) and (2.6)] by

$$\{S^{j}_{m}, S^{j'}_{m'}\} = A^{jj'}_{mm'} E^{j+j'}_{m+m'} + C^{j}_{m}\delta_{jj'}\delta_{m+m',0},$$
$$[S^{j}_{m}, E^{j}_{m}] = 0, \quad (3.1)$$
$$[E^{j}_{m}, E^{j'}_{m'}] = 0,$$

and by the following commutation relations:

$$[J_0, X^{j}_{m}] = m X^{j}_{m},$$
$$[J_\pm, X^{j}_{m}] = [(j\mp m)(j\pm m+1)]^{1/2} X^{j}_{m\pm 1}, \quad (3.2)$$
$$[T_\mu, X^{j}_{m}] = \sum_{j'}\left\langle \begin{matrix} j \\ m+\mu \end{matrix}\right| T_\mu \left|\begin{matrix} j \\ m \end{matrix}\right\rangle X^{j'}_{m+\mu}.$$

The matrix elements are given by (2.14), i.e., (2.17) or (2.19), with X^{j}_{m} as either E^{j}_{m} or S^{j}_{m}. The structure constants $A^{jj'}_{mm'} = A^{j'j}_{m'm}$ are identically zero if $\sigma_2 \neq 0$, whereas if $\sigma_2 = 0$ they are nontrivial. Using the (graded) Jacobi identities[17] for gsl(3,R), the $A^{jj'}_{mm'}$ can then be expressed in terms of a single parameter $A^{\frac{1}{2}\frac{1}{2}}_{\frac{1}{2}\frac{1}{2}}$ (Equations (3.5) and (3.9) below). $A^{\frac{1}{2}\frac{1}{2}}_{\frac{1}{2}\frac{1}{2}}$ can in its turn be absorbed into the definition of S^{j}_{m}. The constant term $C^{j}_{m}\delta_{jj'}\delta_{m+m',0}$ is analogous to that of the Virasoro–Neveu–Schwarz–Ramond dual string algebras in the quantum case, which plays a crucial role in formulating a physically satisfactory theory. All structure constants C^{j}_{m} can be expressed in terms of $C^{\frac{1}{2}}_{\frac{1}{2}}$, Eq. (3.13), which appears to be a free parameter of the graded algebra.

Let us first take $C^{j}_{m} = 0$. We make use of the (graded) Jacobi identity for $(J_+, S^{j}_{m}, S^{j'}_{m'})$, i.e.,

$$[J_+, \{S^{j}_{m}, S^{j'}_{m'}\}] = \{[J_+, S^{j}_{m}], S^{j'}_{m'}\} + \{S^{j}_{m}, [J_+, S^{j'}_{m'}]\}, \quad (3.3)$$

as well as of (3.1) and (3.2) and obtain

$$[(j+j'-m-m')(j+j'+m+m'+1)]^{\frac{1}{2}} A^{jj'}_{mm'}$$

$$= [(j-m)(j+m+1)]^{\frac{1}{2}} A^{jj'}_{m+1\, m'} + [(j'-m')(j'+m'+1)]^{\frac{1}{2}} A^{jj'}_{m\, m'+1}\,. \tag{3.4}$$

This equation can be recurrently solved for all $A^{jj'}_{m\,m'}$ in terms of $A^{jj'}_{jj'}$. If $m=j$ $(m'=j')$ one can solve (3.4) for all $A^{jj'}_{j\,m'}$ $(A^{jj'}_{m\,j'})$ in terms of $A^{jj'}_{jj'}$. Utilizing this result one can then solve (3.4) for all $A^{jj'}_{j-1\,m'}$ and $A^{jj'}_{m\,j'-1}$, and so forth. After some algebra we arrive at

$$A^{jj'}_{m\,m'} = \{(2j)!(2j')!(j+j'+m+m')!(j+j'-m-m')!/ (2j+2j')!(j+m)!(j-m)!(j'+m')! \times (j'-m')!]^{-1}\}^{1/2} A^{jj'}_{jj'}\,. \tag{3.5}$$

It is obvious from this expression that the $A^{jj'}_{m\,m'}$ coefficients are invariant under the substitution of $(-m,-m')$ for (m,m'), i.e., $A^{jj'}_{-m\,-m'} = A^{jj'}_{m\,m'}$.

The $A^{jj'}_{jj'}$ coefficients are not independent and we will now show that they can all be expressed in terms of a single one, say $A^{\frac{1}{2}\frac{1}{2}}_{\frac{1}{2}\frac{1}{2}}$. Let us make use of the (graded) Jacobi identity for $(T_\mu, S^j_m, S^{j'}_{m'})$, i.e.,

$$[T_\mu,\{S^j_m,S^{j'}_{m'}\}] = \{[T_\mu,S^j_m],S^{j'}_{m'}\} + \{S^j_m,[T_\mu,S^{j'}_{m'}]\}\,. \tag{3.6}$$

Equating the terms in $E^{j+j'}_{m+m'+\mu}$, we extract from (3.6) the following equation:

$$\left\langle {j+j' \atop m+m'+\mu} \middle| T_\mu \middle| {j+j' \atop m+m'} \right\rangle A^{jj'}_{m\,m'} = \left\langle {j \atop m+\mu} \middle| T_\mu \middle| {j \atop m} \right\rangle A^{jj'}_{m+\mu\, m'} + \left\langle {j' \atop m'+\mu} \middle| T_\mu \middle| {j' \atop m'} \right\rangle A^{jj'}_{m\, m'+\mu}\,. \tag{3.7}$$

The matrix elements on the right-hand side of (3.7) correspond to the sl(3,R) unirrep $\mathscr{D}(\frac{1}{2};0)$ and thus vanish [see (2.19)]. Therefore the nontrivial solution for $A^{jj'}_{m\,m'}$ can be achieved provided that the label σ_2 of the $\mathscr{D}(1;\sigma_2)$ sl(3,R) unirrep is taken to be zero, so that Eq. (3.7) is identically satisfied.

Convenient recurrence relations can be obtained from the coefficients in terms of $E^{j+j'-2}_{m+m'+\mu}$ for $m=j$, $m'=j'$, and $\mu=-2$, i.e.,

$$t^{-}_{-2}(j+j')A^{jj'}_{jj'} = t^{-}_{-2}(j)A^{j-2\,j'}_{j-2\,j'} + t^{-}_{-2}(j')A^{j\,j'-2}_{j\,j'-2}\,, \tag{3.8}$$

where

$$t^{-}_{-2}(j) \equiv \left\langle {j-2 \atop j-2} \middle| T_{-2} \middle| {j \atop j} \right\rangle.$$

Fixing $j=\frac{1}{2}$ $(j'=\frac{1}{2})$, one can determine all $A^{\frac{1}{2}j'}_{\frac{1}{2}j'}$ $(A^{j\frac{1}{2}}_{j\frac{1}{2}})$ coefficients in terms of $A^{\frac{1}{2}\frac{1}{2}}_{\frac{1}{2}\frac{1}{2}}$. Having done this, one can then solve (3.8) for $A^{5/2\,j'}_{5/2\,j'}$, i.e., for $A^{j\,5/2}_{j\,5/2}$, and so forth. It is rather straightforward to determine all $A^{jj'}_{jj'}$ in a closed form. We finally find

$$A^{jj'}_{jj'} = \frac{[\frac{1}{2}(j+j'-1)]!}{[\frac{1}{2}(j-\frac{1}{2})]![\frac{1}{2}(j'-\frac{1}{2})]!} \times \frac{\prod\limits_{k=0}^{(\frac{1}{2})(j-5/2)} t^{-}_{-2}(j-2k) \prod\limits_{k'=0}^{(\frac{1}{2})(j'-5/2)} t^{-}_{-2}(j'-2k')}{\prod\limits_{k=0}^{(\frac{1}{2})(j+j'-3)} t^{-}_{-2}(j+j'-2k)} A^{\frac{1}{2}\frac{1}{2}}_{\frac{1}{2}\frac{1}{2}}\,. \tag{3.9}$$

We now explore the possibility of having a constant term C^j_m in (3.1). From the (graded) Jacobi identities for $(J_\pm, S^j_m, S^{j'}_{m'})$ we extract

$$C^j_{m+1} + C^j_m = 0\,. \tag{3.10}$$

The (graded) Jacobi identities for $(T_\mu, S^j_m, S^{j'}_{m'})$ provide us with the following recurrence relations:

$$\left\langle {j' \atop m+\mu} \middle| T_\mu \middle| {j \atop m} \right\rangle C^{j'}_{m+\mu} + \left\langle {j \atop -m} \middle| T_\mu \middle| {j' \atop -m-\mu} \right\rangle C^j_m = 0\,. \tag{3.11}$$

Substituting the explicit expression (2.14) for the matrix elements, by making use of the symmetries of the $3j$ coefficients and taking into account the consistency condition requiring the vanishing of σ_2, we obtain

$$C^{j'}_{m+\mu} = (-)^{\mu+1} C^j_m\,, \tag{3.12}$$

where $j' = j \pm 2$. Finally, we can express an arbitrary C^j_m in terms of $C^{\frac{1}{2}}_{\frac{1}{2}}$ through the formula

$$C^j_m = (-)^{\frac{1}{2}(j-\frac{1}{2})}(-)^{m-\frac{1}{2}} C^{\frac{1}{2}}_{\frac{1}{2}}\,. \tag{3.13}$$

4. GRADED ALGEBRA CONSTRUCTION: CASE B

Besides the simplest minimal graded sl(3,R) Lie algebras of case A, there exists another minimal graded sl(3,R) Lie algebra based on the sl(3,R) unirrep $\mathscr{D}(0;\sigma_2)$. This is an infinite-dimensional algebra with generators J_0, $J_\pm$, and T_μ $(\mu = 0, \pm 1, \pm 2)$, S^j_m $(j=\frac{1}{2},\frac{5}{2},\frac{9}{2},\cdots;|m|\leqslant j)$, and G^j_m $(j=0,2,4,\cdots;|m|\leqslant j)$. The generators J_0, $J_\pm$, and T_μ form sl(3,R) itself, while S^j_m and G^j_m transform as components of the sl(3,R) irreducible tensor operators corresponding to $\mathscr{D}(\frac{1}{2};0)$ and $\mathscr{D}(0;\sigma_2)$, respectively. Graded Lie brackets (including a central term) are given in addition to those of sl(3,R) [Equations (2.5) and (2.6)] by

$$\{S^j_m,S^{j'}_{m'}\} = B^{jj'}_{m\,m'} G^{|j-j'|}_{m+m'} + C^j_m \delta_{jj'}\delta_{m+m',0}\,,$$
$$[S^j_m, G^{j'}_{m'}] = 0\,, \tag{4.1}$$
$$[G^j_m, G^{j'}_{m'}] = 0\,,$$

and by the following commutation relations:

$$[J_0, X^j_m]' = mX^j_m\,,$$
$$[J_\pm, X^j_m] = ((j\mp m)(j\pm m+1))^{\frac{1}{2}} X^j_{m\pm 1}\,, \tag{4.2}$$
$$[T_\mu, X^j_m] = \sum_{j'} \left\langle {j' \atop m+\mu} \middle| T_\mu \middle| {j \atop m} \right\rangle X^{j'}_{m+\mu}\,,$$

where the matrix elements are given by (2.14), i.e., (2.17) or (2.19), and X^j_m now represents G^j_m or S^j_m. The structure constants $B^{jj'}_{m\,m'} = B^{j'j}_{m'\,m}$ are identically zero if $\sigma_2 \neq 0$. For $\sigma_2 = 0$, they are nontrivial; using the (graded) Jacobi identities for gsl(3,R), they can be expressed in terms of a single parameter $A^{5/2\;1/2}_{5/2\;-1/2}$, which can be given unit value through a rescaling of the operators S^j_m. The structure constants C^j_m can be expressed in terms of a single constant $C^{\frac{1}{2}}_{\frac{1}{2}}$ [see Eq. (4.16) below], which is a free parameter of the graded algebra.

Let us take first $C^j_m = 0$, i.e., no center. Without any loss of generality we can always take $j \geqslant j'$, i.e., $j - j' \geqslant 0$, owing to the symmetry $\{S^j_m,S^{j'}_{m'}\} = \{S^{j'}_{m'},S^j_m\}$. Note also that $B^{jj'}_{m\,m'}$ vanishes unless $|m+m'| \leqslant j-j'|$ is satisfied. From (4.1), (4.2), and the (graded) Jacobi identities for $(J_-, S^j_m, S^{j'}_{m'})$ and $(J_+, S^j_m, S^{j'}_{m'})$, respectively, we extract the following recurrence relations:

$$[(j-j'+m+m')(j-j'-m-m'+1)]^{\frac{1}{2}}B^{jj'}_{m\,m'} = [(j+m)(j-m+1)]^{\frac{1}{2}}B^{jj'}_{m-1\,m'} + [(j'+m')(j'-m'+1)]^{\frac{1}{2}}B^{jj'}_{m\,m'-1}\,, \qquad (4.3)$$

$$[(j-j'-m-m')(j-j'+m+m'+1)]^{\frac{1}{2}}B^{jj'}_{m\,m'} = [(j-m)(j+m+1)]^{\frac{1}{2}}B^{jj'}_{m+1\,m'} + [(j'-m')(j'+m'+1)]^{\frac{1}{2}}B^{jj'}_{m\,m'+1}\,. \qquad (4.4)$$

Making use of these two equations, it is possible to express the $B^{jj'}_{m\,m'}$ coefficients in terms of the coefficients $B^{jj'}_{j'\,-j'}$. Let us first fix $m' = -j'$. From (4.3) we arrive at

$$B^{jj'}_{j-k\;-j'} = \left(\frac{(2j-2j')!(2j-k)!}{(2j-2j'-k)!(2j)!}\right)^{\frac{1}{2}} B^{jj'}_{j-j'}\,, \qquad k = 1,2,\ldots,2j-2j'\,. \qquad (4.5)$$

We now turn to Eq. (4.4) and take $m = j-1, j-2,\cdots$; $m' = -j'$, obtaining

$$B^{jj'}_{j-k\;-j'+1} = -\left(\frac{2j'k(2j-2j')-(2j-k)!}{(2j-2j'-k+1)!(2j)!}\right)^{\frac{1}{2}} B^{jj'}_{j-j'}\,, \qquad k = 1,2,\ldots,2j-2j'+1\,. \qquad (4.6)$$

In the same manner we finally find through repeated use of (4.4)

$$B^{jj'}_{m\,m'} = (-)^{j'+m'}\{(2j-2j')!(2j')!(j+m)!(j-m)! \times[(2j)!(j-j'+m+m')!(j-j'-m-m')! \times(j'+m')!(j'-m')!]^{-1}\}^{1/2}B^{jj'}_{j-j'}\,, \qquad (4.7)$$

where $j \geqslant j'$, $m = -j+j'-m',\ldots, j-j'-m'$, and $m' = -j',\ldots,j'$.

Let us consider the special case when $j = j'$, and thus $m+m' = 0$. Equation (4.7) yields $B^{jj}_{m\,-m} = (-)^{j-m}B^{jj}_{j\,-j}$. For $m = -j$ one has $B^{jj}_{-j\,j} = -B^{jj}_{j\,-j}$, since $2j = 1,5,\cdots$, but on the other hand, $\{S^j_{-j},S^j_j\} = \{S^j_j,S^j_{-j}\}$ and therefore $B^{jj}_{j\,-j} = 0$, implying

$$B^{jj}_{m\;\,m} = 0, \quad \forall j,\forall m\,. \qquad (4.8)$$

We will now give a procedure expressing all $B^{jj'}_{j-j'}$ coefficients in terms of a single coefficient $B^{5/2\,\frac{1}{2}}_{5/2\,-\frac{1}{2}}$ (since $B^{\frac{1}{2}\frac{1}{2}}_{\frac{1}{2}\,-\frac{1}{2}} = 0$), which can be absorbed into the definition of S^j_m. The procedure is based on the recurrence relations that are obtained from the (graded) Jacobi identity for $(T_\mu,S^j_m,S^{j'}_{m'})$. Let us first consider the relations obtained from the terms in $G^{j-j'}$, i.e.,

$$\left\langle \begin{matrix} j-j' \\ m+m'+\mu \end{matrix} \middle| T_\mu \middle| \begin{matrix} j-j' \\ m+m' \end{matrix} \right\rangle B^{jj'}_{mm'} = \left\langle \begin{matrix} j \\ m+\mu \end{matrix} \middle| T_\mu \middle| \begin{matrix} j \\ m \end{matrix} \right\rangle B^{jj'}_{m+\mu\,m'} + \left\langle \begin{matrix} j' \\ m'+\mu \end{matrix} \middle| T_\mu \middle| \begin{matrix} j' \\ m' \end{matrix} \right\rangle B^{jj'}_{m\,m'+\mu}\,. \qquad (4.9)$$

Owing to the fact that the matrix elements on the right-hand side vanish, since they correspond to the sl(3,R) unirrep $\mathscr{D}(\frac{1}{2};0)$ [see (2.19)], we are forced, as in case A, to require that the matrix elements on the left-hand side of (4.9) vanish as well, i.e., to take the sl(3,R) unirrep $\mathscr{D}(0;0)$ for the G^j_m operators. This is necessary for the existence of a nontrivial solution for $B^{jj'}_{m\,m'}$. The (graded) Jacobi identity for $(T_\mu,S^j_j,S^{j'}_{-j'})$ provides further relations from which we can extract the following recurrence relations:

$$\left\langle \begin{matrix} j-j'+2 \\ j-j'+\mu \end{matrix} \middle| T_\mu \middle| \begin{matrix} j-j' \\ j-j' \end{matrix} \right\rangle B^{jj'}_{j-j'} = \left\langle \begin{matrix} j-2 \\ j+\mu \end{matrix} \middle| T_\mu \middle| \begin{matrix} j \\ j \end{matrix} \right\rangle B^{j+2\,j'}_{j+\mu\,-j'} + \left\langle \begin{matrix} j'-2 \\ -j'+\mu \end{matrix} \middle| T_\mu \middle| \begin{matrix} j' \\ -j' \end{matrix} \right\rangle B^{jj'-2}_{j\,-j'+\mu}\,, \qquad (4.10)$$

and

$$\left\langle \begin{matrix} j-j'-2 \\ j-j'+\mu \end{matrix} \middle| T_\mu \middle| \begin{matrix} j-j' \\ j-j' \end{matrix} \right\rangle B^{jj'}_{j-j'} = \left\langle \begin{matrix} j-2 \\ j+\mu \end{matrix} \middle| T_\mu \middle| \begin{matrix} j \\ j \end{matrix} \right\rangle B^{j-2\,j'}_{j+\mu\,-j'} + \left\langle \begin{matrix} j'+2 \\ j'+\mu \end{matrix} \middle| T_\mu \middle| \begin{matrix} j' \\ -j' \end{matrix} \right\rangle B^{jj'+2}_{j\,-j'+\mu}\,, \qquad (4.11)$$

which are obtained from the terms in $G^{j-j'+2}_{j-j'+\mu}$ and $G^{j-j'-2}_{j-j'+\mu}$, respectively. It follows from (2.14) and the symmetry of the $3j$ coefficients that

$$\left\langle \begin{matrix} j'\mp 2 \\ -j'\pm 2 \end{matrix} \middle| T_{\pm 2} \middle| \begin{matrix} j' \\ -j' \end{matrix} \right\rangle = \left\langle \begin{matrix} j'\mp 2 \\ j'\mp 2 \end{matrix} \middle| T_{\mp 2} \middle| \begin{matrix} j' \\ m' \end{matrix} \right\rangle,$$

so that the Eqs. (4.10) and (4.11) become respectively

$$t^{+}_{+2}(j-j')B^{jj'}_{j-j'} = t^{+}_{+2}(j)B^{j+2\,j'}_{j+2\,-j'} + t^{-}_{-2}(j')B^{jj'-2}_{j\,-j'+2} \qquad (4.12)$$

and

$$t^{-}_{-2}(j-j')B^{jj'}_{j-j'} = t^{-}_{-2}(j)B^{j-2\,j'}_{j-2\,-j'} + t^{+}_{+2}(j')B^{jj'+2}_{j\,-(j'+2)}\,, \qquad (4.13)$$

where

$$t^{\pm}_{\mu}(j) \equiv \left\langle \begin{matrix} j\pm 2 \\ j+\mu \end{matrix} \middle| T_\mu \middle| \begin{matrix} j \\ j \end{matrix} \right\rangle,$$

and we have set μ to be $+2$ in (4.10) and -2 in (4.11). We first solve recurrently (4.12) for $j' = \frac{1}{2}$ and arrive at

$$B^{j\,\frac{1}{2}}_{j\,-\frac{1}{2}} = \left(\prod_{k=0}^{(\frac{1}{4})(j-5/2)} \frac{t^{+}_{+2}(2+2k)}{t^{+}_{+2}(5/2+2k)}\right) B^{5/2\;1/2}_{5/2\;-1/2}\,, \qquad j = \tfrac{9}{2},\tfrac{13}{2},\cdots\,. \qquad (4.14)$$

We now proceed by making use of (4.13). The expressions for the remaining $B^{jj'}_{j-j'}$ coefficients are not monomials and become more and more cumbersome. We just note that from (4.13), when $j' = \frac{1}{2}$ we find for $j = \frac{9}{2}$:

$$B^{\frac{9}{2}\frac{5}{2}}_{\frac{9}{2}-\frac{5}{2}} = [t^{-}_{-2}(4)t^{+}_{+2}(2) - t^{-}_{-2}(\tfrac{9}{2})t^{+}_{+2}(\tfrac{5}{2})] \times[t^{+}_{+2}(\tfrac{5}{2})t^{+}_{+2}(\tfrac{1}{2})]^{-1}B^{\frac{5}{2}\frac{1}{2}}_{\frac{5}{2}-\frac{1}{2}} \qquad (4.15)$$

and so forth for $j = \frac{13}{2},\frac{17}{2},\cdots$. Then when $j' = \frac{5}{2}$ one can find from (4.13) all $B^{j\frac{5}{2}}_{j-\frac{5}{2}}$, $j = \frac{13}{2},\frac{17}{2},\cdots$, and so forth.

We consider now the nontrivial structure constant C^j_m in (4.1). From the (graded) Jacobi identities for $(J_\pm,S^j_m,S^{j'}_{m'})$ and $(T_\mu,S^j_m,S^{j'}_{m'})$ we extract the same set of recurrence relations for C^j_m as in case A above and thus have

$$C^j_m = (-)^{\frac{1}{2}(j-\frac{1}{2})}(-)^{m-\frac{1}{2}}C^{\frac{1}{2}}_{\frac{1}{2}}\,. \qquad (4.16)$$

5. GRADING, DILATIONS, AND THE CENTER

We now extend the sl(3,R) algebra to the algebra of the GL(3,R) group, i.e., gl(3,R), by adjoining the dilation operator D, i.e.,

$$[D,J_0] = [D,J_\pm] = [D,T_\mu] = 0, \quad \mu = 0, \pm 1, \pm 2\,. \tag{5.1}$$

Owing to these commutation relations, we obtain the following commutation relations for the graded case of ggl(3,R):

$$[D,S^{\,j}_{\,m}] = d_s S^{\,j}_{\,m}, \quad [D,E^{\,j}_{\,m}] = d_\epsilon E^{\,j}_{\,m},$$
$$[D,G^{\,j}_{\,m}] = d_G G^{\,j}_{\,m}\,. \tag{5.2}$$

Applying the (graded) Jacobi identity for $(D,S^{\,j}_{\,m},S^{\,j'}_{\,m'})$, we find $d_\epsilon = 2d_s$ and $d_G = 2d_s$ for cases A and B, respectively. The constant d_s can be absorbed in the definition of D.

Dilations thus provide us with a grading. Taking $d_s = 1$, we get a Z grading with trivial L_i subalgebras for $i<0$, $i>2$:

$$\begin{aligned} i=0:&\quad D,J_0,T_\mu,e\,,\\ i=1:&\quad S^{\,j}_{\,m}\,,\\ i=2:&\quad E^{\,j}_{\,m}\quad (G^{\,j}_{\,m})\,,\end{aligned} \tag{5.3}$$

where e is the identity, i.e., the central term multiplying $C^{\,j}_{\,m}$ in (3.1) and (4.1).

The need for a central term e in dual models[3] relates primarily to covariance, i.e., to the preservation of the commutation relations of the Poincare algebra

$$\begin{aligned}[J_{(L)}^-, J_{(R)}^-] &= \frac{2}{P+2}\sum_{n=1}\Big\{ n(1-\tfrac{1}{24}(d_V-2)\\ &\quad + \frac{1}{n}\big[\tfrac{1}{24}(d_V-2)-\alpha(0)\big]\Big\}\\ &\quad \times \text{operators} \to 0\,,\end{aligned}$$

where $J_{(L)}$ and $J_{(R)}$ correspond to the $SU(2)_L \times SU(2)_R$ decomposition of SO(4). This commutator has to vanish, which fixes $\alpha(0)=1$ and $d_V=26$.

In our work the Poincaré group commutation relations [here the iso(3)] are correct by construction, since we start from an algebraic ansatz rather than from canonical variables. We could thus dispense with e here. However, the construction of a dynamical theory may involve further complications and we have thus chosen to include in this study the possibility of adjoining a central term.

ACKNOWLEDGMENTS

One of us (Dj.S.) would like to acknowledge the hospitality of Y.N. and the Wolfson Chair Extraordinary in Theoretical Physics at Tel Aviv University.

[1]Y. Ne'eman and T.N. Sherry, Nucl. Phys. B **138**, 31 (1978).
[2]Y. Ne'eman and T.N. Sherry, Phys. Lett. B **76**, 413 (1978).
[3]See for example J. Scherk, Rev. Mod. Phys. **47**, 123 (1975).
[4]Dj. Sijacki (unpublished).
[5]D.Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, Phys. Rev. D **13**, 3214 (1976); S. Deser and B. Zumino, Phys. Lett. B **62**, 335 (1976).
[6]J. Scherk and J.H. Schwarz, Nucl. Phys. B **81**, 118 (1974); T. Yoneya, Prog. Theor. Phys. **51**, 1907 (1974).
[7]F.W. Hehl, E.A. Lord, and Y. Ne'eman, Phys. Lett. B **71**, 432 (1977); Phys. Rev. D **17**, 428 (1978).
[8]Y. Ne'eman and Dj. Sijacki, Proc. Nat. Acad. Sci. USA **76**, 561 (1979); Ann. Phys. (N.Y.) **120**, 292 (1979).
[9]F.W. Hehl, Y. Ne'eman, J. Nitsch, and P. von der Heyde, Phys. Lett B **78**, 102 (1978).
[10]S. Deser, J. Kay, and K. Stelle, Phys. Rev. Lett. **38**, 527 (1977).
[11]V. Bargmann, Ann. Math. **59**, 1 (1954).
[12]Dj. Sijacki, J. Math. Phys. **16**, 298 (1975).
[13]E. Angelopoulos, J. Math. Phys. **19**, 2108 (1978).
[14]Papers in Ref. 12 and V.I. Ogievetskii and E. Sokachev, Teor. Math. Phys. (USSR) **23**, 214 (1975); J.W.B. Hughes, J. Phys. A **7**, 779 (1974).
[15]Y. Dothan and Y. Ne'eman, in *Symmetry Groups in Nuclear and Particle Physics*, edited by F.J. Dyson (Benjamin, New York, 1966).
[16]E. Inönü and E.P. Wigner, Proc. Nat. Acad. Sci. (US) **39**, 510 (1953); I.E. Segal, Duke Math. J. **18**, 221 (1951).
[17]L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).

A symmetry approach to exactly solvable evolution equations [a)]

A. S. Fokas
Department of Applied Mathematics, California Institute of Technology, Pasadena, California 91125

(Received 27 December 1979; accepted for publication 7 March 1980)

A method is developed for establishing the exact solvability of nonlinear evolution equations in one space dimension which are linear with constant coefficient in the highest-order derivative. The method, based on the symmetry structure of the equations, is applied to second-order equations and then to third-order equations which do not contain a second-order derivative. In those cases the most general exactly solvable nonlinear equations turn out to be the Burgers equation and a new third-order evolution equation which contains the Korteweg-de Vries (KdV) equation and the modified KdV equation as particular cases.

INTRODUCTION

This paper is concerned with the problem of determining whether a given nonlinear evolution equation is exactly solvable, and also with the problem of finding all such equations of a given order. An equation is called exactly solvable if it admits a Lax formulation,[1] that is, if there exist differential or integral operators L and A such that the given equation can be written in the form $L_t = [A, L]$.

Our approach to the above problem is based on the consideration of the symmetry structure of the given equation. If an equation is exactly solvable it possesses infinitely many generalized (as opposed to Lie-point) symmetries. The existence of a generalized symmetry manifests itself by the existence of an admissible generalized or Lie-Bäcklund (LB) operator.[2] The existence of infinitely many symmetries is expressed by the existence of a recursion operator Δ [3] (see also Sec. 1) which generates a new admissible LB operator from a given one. In almost[4] all known cases the admissible operator expressing invariance of equation under t-translation is generated from that expressing invariance under x-translation. We call equations possessing this property exactly solvable equations of normal type. Another interesting fact regarding the symmetry structure of evolution equations is that in all known cases the existence of one generalized symmetry implies the existence of infinitely many. (However, this has not been proved in general.) With the above in mind we now formulate our criterion, which is elaborated in Sec. 2.

A. Proposition

A necessary condition for a nonlinear evolution equation of nth order to be exactly solvable of normal type is that it admit an LB operator with a generator of order $2n-1$.

Having obtained this generator, it is usually possible by inspection to obtain the recursion operator Δ, the existence of which provides a sufficient condition for the exact solvability of this equation, since Δ and the Fréchet derivative of the t-independent part of the equation form a Lax pair (see Sec. 1).

The above criterion is quite practical since it is algorithmically very straightforward to find out if a given equation admits a generator of a given order. Further, it is also algorithmically possible to determine which equations of a certain order admit a generator of a given order. This is illustrated in Sec. 2, where we find all second-order equations and all third-order equations (not involving a second-order derivative) which are of normal type. Within equivalence (see Sec. 2.1), the most general nonlinear second-order equation with the above property is the Burgers equation. The most general third-order equations turn out to be: (i) A generalization of the Korteweg–de Vries (KdV) equation, see (2.18) which, in particular, contains any linear combination of the KdV and of the modified KdV (MKdV) as a special case, (ii) A linear combination of the potential KdV (PKdV) and of the potential MKdV (PMKdV). The potential KdV (or the potential version of the KdV) is the equation obtained from the KdV after replacing the dependent variable u by the "potential" w, $u = w_x$, and integrating once.

B. Outline of the paper

In Sec. 1 we first define admissible LB operators[2] as restricted to evolution equations as well as their commutators[5] and prove a lemma expressing the admissibility of an LB operator in a convenient form. We further recall the definition of a recursion operator[3] and then prove that the recursion operator together with the Fréchet derivative of the time-independent part of a given evolution equation form a Lax pair, and also give a convenient characterization of a recursion operator as well as of its main property (see Lemmas 2 and 3). Finally, for completeness of the presentation, the definition of a hereditary operator[6,7] is recalled. In Sec. 2 we first outline our method and then present some concrete results which also illustrate the general theory. In Sec. 2A we find all second-order equations possessing a third-order symmetry and the corresponding admissible generators. We further show that all these equations can be linearized, and give explicitly the corresponding linearizing Bäcklund transformations (BT). The recursion operators possessed by the above equations are also explicitly given. In Sec. 2.2 we find all third-order equations (not involving a second-order derivative) possessing a fifth-order symmetry. We also give the corresponding admissible generators and recursion operators. Finally, in Sec. 3 we compare our method with other existing ones and indicate some open questions.

[a)]Research supported in part by NSF grant MES 78-0306 and by the Saul Kaplun Memorial Fund.

1. MATHEMATICAL PRELIMINARIES

A. Admissible LB operators

In what follows we shall consider evolution equations of the form

$$\Omega \equiv u_t + K(x,u,u_1,\ldots,u_n) = 0, \tag{1.1}$$

where

$$u_j \equiv \left(\frac{\partial}{\partial_x}\right)^j u, \quad j = 0,1,\ldots,n. \tag{1.2}$$

The most general LB operator associated with (1.1) is given by

$$X(\eta) \equiv \eta \frac{\partial}{\partial u} + (D_t\eta)\frac{\partial}{\partial u_t} + \sum_{j=1}^{\infty} (D^j\eta)\frac{\partial}{\partial u_j}, \tag{1.3}$$

where $\eta = \eta(x,t,u,u_1,\ldots,u_N)$, N arbitrary, is called the generator of the above LB operator, D is the total derivative with respect to x

$$D \equiv D_x \equiv \frac{\partial}{\partial x} + u_1 \frac{\partial}{\partial u} + u_{t1}\frac{\partial}{\partial u_t} + u_2 \frac{\partial}{\partial u_1} + \cdots,$$

and D_t is defined analogously. Without loss of generality, we assume that η does not depend on t-derivatives since for admissible operators they can always be eliminated using equation (1.1).

The LB operator $X(\eta)$ is an *admissible* LB operator of (1.1) iff $X(\eta)\Omega = \sigma$, where $\sigma = 0$ when Eq. (1.1) and its differential consequences are assumed. The above is denoted by

$$X(\eta)\Omega\,|_{\Omega=0} = 0 \tag{1.4}$$

Equation (1.4) provides an algorithm for finding η as the solution of a system of liner overdetermined equations.

An important special class of LB operators is the class of Lie (point) operators. The most general such operator associated with (1.1) is given by

$$Z = \xi\frac{\partial}{\partial x} + \tau\frac{\partial}{\partial t} + \nu\frac{\partial}{\partial u}, \tag{1.5}$$

where ξ, τ, ν are functions of x, t, and u only. The operator Z can be written in the form (1.3) by the equivalence[8]

$$Z \Leftrightarrow X(\nu - \xi u_1 - \tau u_t), \tag{1.6}$$

or, using Eq. (1.1), $X(\nu - \xi u_1 + \tau K)$. The above equivalence means that Eq. (1.1) admits Z iff it admits the corresponding X.

The commutator of two LB operators is an LB operator whose generator is expressed by a simple formula,[5]

$$[X(\eta_1), X(\eta_2)] = X(\eta_3),$$

$$\text{where } \eta_3 = X(\eta_1)\eta_2 - X(\eta_2)\eta_1. \tag{1.7}$$

Obviously, the admissible LB operators of a given equation form a Lie algebra.

In considering the symmetries of an equation, it is convenient to use an operator formulation.[9] We define the Fréchet derivative of a function $N(u) \equiv N(u,u_t,u_1,\ldots,u_n)$ by $N'(u)$, where

$$N'(u)[v] \equiv \left.\frac{\partial N(u+\epsilon v)}{\partial\epsilon}\right|_{\epsilon=0}. \tag{1.8}$$

Clearly, the right-hand side of Eq. (1.8) is linear in v and therefore $N'(u)$ is an operator acting linearly on v. Actually,

$$N'(u) = \frac{\partial N}{\partial u} + \frac{\partial N}{\partial u_t}D_t + \sum_{j=1}^{n}\frac{\partial N}{\partial u_j}D^j. \tag{1.9}$$

Comparing Eq. (1.3) and (1.9) we obtain

$$X(\eta)\Omega = \Omega'[\eta]. \tag{1.10}$$

Therefore Eq. (1.4) takes the form

$$\Omega'[\eta]\,|_{\Omega=0} = 0. \tag{1.11}$$

In the case of evolution equations, Eq. (1.11) [or, equivalently, (1.4)] can be further simplified

Lemma 1: The evolution equation (1.1) admits the LB operator $X(\eta)$ generated by $\eta = \eta(x,t,u,u_1,\ldots,u_n)$ iff

$$\eta_t + X(\eta)K - X(K)\eta = 0 \tag{1.12a}$$

or, equivalently,

$$\eta_t + K'[\eta] - \eta'[K] = 0. \tag{1.12b}$$

Proof: Writing out Eq. (1.4), we obtain

$$(D_t\eta) + X(\eta)K = 0, \text{ when (1.1) holds,}$$

or

$$\eta_t + \frac{\partial\eta}{\partial u}u_t + \sum_{j=1}^{N}\frac{\partial\eta}{\partial u_j}u_{jt} + X(\eta)K = 0, \text{ when (1.1) holds,}$$

or

$$\eta_t - \frac{\partial\eta}{\partial u}K - \sum_{j=1}^{N}\frac{\partial\eta}{\partial u_j}D^jK + X(\eta)K = 0. \qquad \text{Q.E.D.}$$

The advantage of (1.12) in either form as compared with (1.4) [or (1.11)] is that the validity of (1.1) has already been assumed. Therefore the admissibility of $X(\eta)$ by (1.1) is expressed as a relation between η and K with no further assumptions to be made. Further, using Eq. (1.7) we obtain

Corollary 1: If η does not depend on t explicitly, Eq. (1.1) admits $X(\eta)$ iff $X(\eta)$ and $X(K)$ commute.

Recursion Operators: The operator $\Delta(u)$ is a recursion operator for Eq. (1.1) iff[3]

$$[\Omega',\Delta]_{\Omega=0} = 0. \tag{1.13}$$

It follows from the above definition that if $X(\eta)$ is an admissible LB operator of (1.1) and Δ is a recursion operator of (1.1), then the infinitely many LB operators $X(\Delta^j\eta)$, $j = 0, 1, 2,\cdots$, are also admissible operators for the equation (1.1).[3] (See also Lemma 3 and its corollary.)

We define the Fréchet derivative of an operator-valued function $\Delta(u)$ by

$$\Delta'(u)[v]w \equiv \left.\frac{\partial[\Delta(u+\epsilon v)w]}{\partial\epsilon}\right|_{\epsilon=0}, \tag{1.14}$$

and say that $\Delta'(u)[v]w$ is the derivative of $\Delta(u)$ evaluated at v and then applied to w. For example, the recursion operator of the KdV equation is given by $\Delta(u) = D^2 + 2/3u + 1/3u_1D^{-1}$, where D^{-1} is the inverse total derivative $D^{-1}(w)(x) = \int_\alpha^x w(\xi)d\xi$. Therefore $\Delta'(u)[v] = 2/3v + 1/3\,v_1D^{-1}$.

Lemma 2: The operator $\Delta(u)$ is a recursion operator for

Eq. (1.1) iff the operators $\Delta(u)$ and $K'(u)$ form a Lax pair for Eq. (1.1). This is a consequence of the following equivalence

$$[\Omega',\Delta]_{\Omega=0} = (\Delta_t + [K',\Delta])_{\Omega=0} = -\Delta'[K] + [K',\Delta]. \quad (1.15)$$

(Here Δ_t actually means $D_t\Delta$ whereas, for instance, in (1.12) η_t means $\partial\eta/\partial t$. The reason for this regrettable inconsistency of notations is that in discussing a Lax pair $D_t L$ is customarily denoted by L_t.)

Proof:

$$[\Omega',\Delta]\eta = [D_t + K',\Delta]\eta$$
$$= [D_t,\Delta]\eta + [K',\Delta]\eta = \Delta_t\eta + [K',\Delta]\eta,$$

since

$$[D_t,\Delta]\eta = D_t(\Delta\eta) - \Delta(D_t\eta) = \Delta_t\eta.$$

The second equivalence in (1.15) follows from the above using the chain rule of differentiation and Eq. (1.1), since $\Delta_t = \Delta'[u_t] = -\Delta'[K]$. Q.E.D.

The most convenient characterization of a recursion operator follows from the equation

$$\Delta'[K] = [K',\Delta], \quad (1.16)$$

since Eq. (1.1) has now been eliminated.

The following lemma expresses a useful property of a recursion operator.

Lemma 3: The operator Δ is a recursion operator of Eq. (1.1) iff

$$K'[\Delta\zeta] - (\Delta\zeta)'[K] = \Delta(K'[\zeta] - \zeta'[K]), \quad (1.17)$$

where $\zeta(x,t,u,u_1,\ldots,u_n)$ is an arbitrary function of the arguments indicated.

Proof: Using Leibnitz's rule, we obtain

$$(\Delta\zeta)'[K] = \Delta'[K]\zeta + \Delta(\zeta'[K]).$$

Then using Eq. (1.16) we obtain that

$$(\Delta\zeta)'[K] = K'[\Delta\zeta] - \Delta(K'[\zeta]) + \Delta(\zeta'[K])$$

iff Δ is a recursion operator of Eq. (1.1). Q.E.D.

From the above lemma and Eq. (1.12b) one finds

Corollary 2: If $\Delta(u)$ is a recursion operator [of Eq. (1.1)] not depending explicitly on t and $X(\eta)$, $\eta(x,t,u,u_1,\ldots,u_N)$ is admitted by Eq. (1.1), then the LB operators $X(\Delta^j\eta)$, $j = 1,2,\cdots$, are also admitted by (1.1).

Hereditary operators: Assume that Eq. (1.1) possesses a recursion operator Δ. We call hierarchy 1 the hierarchy of admissible operators

$$X(\Delta^j u_1), \quad j = 0,1,2,\cdots, \quad (1.18)$$

which are generated from the x-translation operator $X(u_1)$. It is obvious that the operators $X(u_t + \Delta^j u_1)$ are also admissible. Equating to zero the generators of these admissible LB operators we obtain

$$u_t + \Delta^j u_1 = 0, \quad j = 1,2,\cdots, \quad (1.19)$$

which is the Lax hierarchy of equations associated with Eq. (1.1).[1] In Ref. 6 it is shown that the operator Δ is a recursion operator of the whole hierarchy (1.19) if Δ satisfies

$$[\Delta,\Delta'][v]w = [\Delta,\Delta'][w]v, \quad (1.20)$$

where

$$[\Delta,\Delta'][v]w = \Delta(\Delta'[v]w) - \Delta'[\Delta v]w, \quad (1.21)$$

and v, w are arbitrary functions of $u, u_1,\ldots,u_N$. An operator Δ satisfying the above property is called a hereditary operator.[6] It is clear that any operator Δ is a recursion operator for the equation $u_t + u_1 = 0$, since $\Delta[u_1] = [D,\Delta]$. Therefore, any hereditary operator Δ is a recursion operator for the whole hierarchy $u_t + \Delta^j u_1 = 0$ generated by this operator. Therefore, an alternative way to find a recursion operator Δ of Eq. (1.1) is to look for a Δ such that (i) Δ is hereditary and (ii) $\Delta u_1 = K$. The above property of Δ was first introduced in Ref. (7), where it was used to prove that such a Δ generates the exactly solvable equations $u_t + C(\Delta)u_1 = 0$, where $C(Z)$ is an arbitrary function of Z, regular, except possibly at $|Z|\to\infty$ and some points Z_c $(Z_c < \infty)$.

2. A METHOD FOR FINDING OUT IF A GIVEN EQUATION IS EXACTLY SOLVABLE

If an evolution equation admits a Lax formulation it also admits infinitely many symmetries. Actually, every member of the Lax hierarchy [see Eq. (1.19)] associated with a given solvable equation is a generator of a generalized symmetry admitted by this equation. Therefore, in order to establish that an equation is exactly solvable we must prove that it possesses infinitely many symmetries. Although there exists an algorithmic way of finding out if a function of the general form $\eta = \eta(u,u_1,\ldots,u_N)$ is an admissible generator, this does not lead to a very practical method for establishing the existence of infinitely many symmetries. However, in all known cases the existence of one generalized symmetry seems to be sufficient for the existence of infinitely many. Further, having obtained one generalized symmetry it is usually possible, almost by inspection, to find a recursion operator Δ which generates infinitely many symmetries. Therefore, the problem of finding out if an equation is exactly solvable reduces to finding an LB symmetry.

In order to find an LB symmetry we must assume the order of the highest derivative in $\eta(u,u_1,\ldots,u_N)$. But how can we know N *a priori*? It is at this point that we use the *existence* of Δ. The only assumption we make is that this Δ generates the t-translation symmetry of the equation from the x-translation symmetry. Let us be more specific. Suppose we are given an evolution equation of the form

$$u_t + u_n + \tilde{K}(u,u_1,\ldots,u_{n-1}) = 0. \quad (2.1)$$

This equation possesses two Lie-point generators, $\eta_1 = u_1$ and $\eta_2 = u_n + \tilde{K}(u)$. If there exists a Δ which generates η_2 from η_1, then $\Delta = D^{n-1} + \cdots$. Therefore, the first LB generator is of the form $\eta_3 = u_{2n-1} + g(u,u_1,\ldots,u_{2n-2})$. That is, $N = 2n-1$ and, furthermore, u_N appears linearly. The above discussion justifies, in our opinion, the proposition made in the introduction.

A. Finding all second-order equations which possess a third order symmetry

In this subsection we first determine all equations of the form

$$u_t + u_2 + A(u,u_1) = 0 \quad (2.2)$$

possessing an admissible generator of the form

$$\eta = u_3 + B(u,u_1,u_2). \tag{2.3}$$

It turns out that all equations having this property also possess infinitely many symmetries and further, all can be linearized.

The following equations and corresponding generators are obtained (for details see Appendix A)

(i)

$$u_t + u_2 + \frac{b''(u)}{b'(u)} u_1^2 + \alpha b(u) u_1 = 0, \tag{2.4a}$$

$$\eta = u_3 + \frac{b''}{b'} u_1^3 + \frac{3b''}{b'} u_1 u_2 + \tfrac{3}{2}\alpha\left(b' + \frac{bb''}{b'}\right) u_1^2 + \tfrac{3}{2}\alpha b u_2 + \tfrac{3}{4}\alpha^2 b^2 u_1, \tag{2.4b}$$

where $b(u)$ is an arbitrary function of u, $b'(u) = db/du$, and α is an arbitrary parameter. (Everywhere in this paper greek lower-case letters stand for constant parameters.)

(ii)

$$u_t + u_2 + \left[\frac{\gamma - c'(u)}{c(u)}\right] u_1^2 + \alpha c(u) = 0, \tag{2.5a}$$

$$\eta = u_3 + \left[\left(\frac{\gamma - c'}{c}\right)^2 + \left(\frac{\gamma - c'}{c}\right)'\right] u_1^3 + 3\left(\frac{\gamma - c'}{c}\right) u_1 u_2, \tag{2.5b}$$

where $c(u)$ is an arbitrary function of u. For the discussion to follow it is convenient to let $c \equiv d/d'$, $d \equiv d(u)$. Then (2.5a) becomes

$$u_t + u_2 + [d''/d' + (\gamma - 1)d'/d]u_1^2 + \alpha d/d' = 0. \tag{2.5c}$$

We can add a constant multiple of u_1 to the left-hand side of (2.4a) and (2.5a) without altering the above results. This has been omitted for economy of writing.

We define two equations to be equivalent if one can be obtained from the other by a transformation involving only the dependent variable. Then it is clear that Eq. (2.4a) is equivalent to

$$u_t + u_2 + \alpha u u_1 = 0 \tag{2.6}$$

under the transformation $u \to b(u)$, while Eq. (2.5c) is equivalent to

$$u_t + u_2 + \alpha\gamma u = 0 \tag{2.7}$$

under the transformation $u \to [d(u)]^\gamma$. Therefore, within equivalence the only nonlinear equation of the form (2.2) admitting a generalized symmetry of the form (2.3) is the Burgers equation. Note that under an equivalence transformation Eqs. (2.4a) and (2.5a) remain exactly solvable and must hence retain the same form.

1. Linearizing Bäcklund Transformations

It turns out that all the above equations can be linearized. Also, if they are the only second-order equations exactly solvable, then they are the only equations of the general form (2.2) which can be linearized. The following results are obtained in Ref. 10, Sec. 5.4.1:

(i) The only equation of the form (2.2) mapped to $v_t + v_2 = 0$ under a BT of the general form $v_1 - f(u,v) = 0$ is given by (2.4a). This BT takes the form

$$b(u) = 2v_1/(\alpha v + \lambda). \tag{2.8}$$

(ii) The only equation of the form (2.2) mapped to $v_t + v_2 = 0$ under a BT of the form $u_1 - f(u,v) = 0$ is given by (2.5a) with $\gamma \equiv 0$. This BT takes the form

$$v = (u_1/\alpha c(u)). \tag{2.9}$$

(iii) The only equation of the form (2.2) mapped to $v_t + v_2 + \alpha\gamma v = 0$ under a map of the form $u = f(v)$ is given by (2.5c). This map is

$$d(u) = v^{1/\gamma}. \tag{2.10}$$

Every linear equation possesses infinitely many symmetries. Therefore, every nonlinear equation which can be linearized also possesses infinitely many symmetries. However, the reverse is not true; that is, not every equation possessing infinitely many symmetries can be linearized (for example, the KdV). In the case of second-order equations, however, we see that the class of equations possessing a generalized symmetry (and actually infinitely many, see below) coincides with the class of second-order equations which can be linearized.

2. Recursion Operators

Equations (2.4a) and (2.5a) possess, respectively, the following recursion operators

$$\Delta = D + \frac{b''}{b'} u_1 + \tfrac{1}{2}\alpha b + \tfrac{1}{2}\alpha u_1 D^{-1}(b' .), \tag{2.11}$$

$$\Delta = D + \left(\frac{\gamma - c'}{c}\right) u_1. \tag{2.12}$$

The operator (2.11) reduces to

$$\Delta = D + \tfrac{1}{2}\alpha u + \tfrac{1}{2}\alpha u_1 D^{-1} \tag{2.13}$$

when $b = u$, which is known to be the recursion operator of the Burgers equation.[3]

It is easy to check that the operator Δ defined by (2.11) is a hereditary operator. Further, $\Delta^2(u_1)$ is the generator (2.4b). Consider, now, the operator (2.12). It can be shown easily that $\Delta = D + a(u)u_1$, where $a(u)$ is an arbitrary function of u, is a hereditary operator. Therefore, Δ is a recursion operator for the equation $u_t + \Delta u_1 = 0$ or $u_t + u_2 + a u_1^2 = 0$. Further, it is clear that the above operator will also be a recursion operator of the equation $u_t + u_2 + a u_1^2 + c = 0$, where $c(u)$ is an arbitrary function of u, iff it is a recursion operator of the equation $u_t + c = 0$, that is iff [see (1.16)]

$$\Delta'[c] = [c', \Delta].$$

This implies $c'' + (ac)' = 0$ or $a = (\gamma - c')/c$. Therefore, Δ is a recursion operator of

$$u_t + u_2 + a u_1^2 + c(u) = 0, \tag{2.14a}$$

iff

$$a = (\gamma - c')/c. \tag{2.14b}$$

Note that since Δ is a recursion operator of (2.14a), where a is given by (2.14b), $\Delta(u_1) = u_2 + a u_1^2$ is a symmetry generator of (2.14a) and, since the whole right-hand side of (2.14a) is also a symmetry, it follows that $c(u)$ is also. This can be trivially checked directly. Also $\Delta c = \gamma u_1$ and therefore the generator $c(u)$ does not produce a new hierarchy of symmetries.

In Ref. 11 (which is an excellent exposition of the Estabrook–Wahlquist method) Kaup asks which equations of the general form $u_t + u_2 + u_1^2 + f(u) = 0$ possess a nontrivial prologation structure. He then finds that $f(u) = \beta e^{-u} + \gamma$ and also develops a method of solving the above equation. Note that if $(\gamma - c')/c = 1$ in Eq. (2.5a), $c(u) = \beta e^{-u} + \gamma$ and, further, this equation is equivalent to a linear one; therefore it is trivially solved.

B. Finding all third-order equations, not involving second-order derivatives, which possess a fifth-order symmetry

In this section we determine all equations of the form

$$u_t + u_3 + A(u,u_1) = 0 \tag{2.15}$$

possessing an admissible generator of the form

$$\eta = u_5 + B(u,u_1,u_2,u_3,u_4). \tag{2.16}$$

The following equations and corresponding generators are obtained (for details see Appendix B).

(i)

$$u_t + u_3 + \alpha u_1^3 + \beta u_1^2 + \gamma u_1 = 0, \tag{2.17a}$$

$$\eta = u_5 + 5\alpha u_1^2 u_3 + \tfrac{10}{3}\beta u_1 u_3 + 5\alpha u_1 u_2^2 + \tfrac{5}{3}\beta u_2^2 + \tfrac{3}{2}\alpha^2 u_1^5 + \tfrac{10}{9}\beta^2 u_1^3 + \tfrac{5}{2}\alpha\beta u_1^4. \tag{2.17b}$$

(ii)

$$u_t + u_3 + \alpha u_1^3 + b(u)u_1 = 0, \tag{2.18a}$$

where $b(u)$ solves

$$b''' + 8\alpha b' = 0, \tag{2.18b}$$

$$\eta = u_5 + 5\alpha u_1^2 u_3 + \tfrac{5}{3}bu_3 + 5\alpha u_1 u_2^2 + \tfrac{10}{3}b'u_1u_2 + \tfrac{3}{2}\alpha^2 u_1^5 + \tfrac{5}{3}\alpha b u_1^3 + \tfrac{5}{6}b''u_1^3 + \tfrac{5}{6}b^2 u_1. \tag{2.18c}$$

It is clear that Eq. (2.17a) is the potential version of the special case of (2.18a) where $\alpha = 0$. In this sense, the most general equation of the form (2.15) admitting a symmetry generator of the form (2.16) is given by (2.18).

Particular Cases:

(i)

$\alpha = 0$ in (2.17a) (PKdV),

$$\eta = u_5 + \tfrac{10}{3}\beta u_1 u_3 + \tfrac{5}{3}\beta u_2^2 + \tfrac{10}{9}\beta^2 u_1^3. \tag{2.19}$$

(ii)

$\beta = 0$ in (2.17a) (PMKdV),

$$\eta = u_5 + 5\alpha u_1^2 u_3 + 5\alpha u_1 u_2^2 + \tfrac{3}{2}\alpha^2 u_1^5. \tag{2.20}$$

(iii)

$b = 0$ in (2.18a) (PMKdV),

$$\eta = u_5 + 5\alpha u_1^2 u_3 + 5\alpha u_1 u_2^2 + \tfrac{3}{2}\alpha^2 u_1^5. \tag{2.21}$$

(iv)

$\alpha = 0, b = u$ in (2.18a) (KdV),

$$\eta = u_5 + \tfrac{5}{3}uu_3 + \tfrac{10}{3}u_1u_2 + \tfrac{5}{6}u^2u_1. \tag{2.22}$$

(v)

$\alpha = 0, b = u^2$ in (2.18a) (MKdV),

$$\eta = u_5 + \tfrac{5}{3}u^2u_3 + \tfrac{20}{3}uu_1u_2 + \tfrac{5}{3}u_1^3 + \tfrac{5}{6}u^4u_1. \tag{2.23}$$

1. Recursion Operators

Equations (2.17a) and (2.18a) possess, respectively, the following recursion operators

$$\Delta = D^2 + \gamma + 2\alpha u_1^2 + \tfrac{4}{3}\beta u_1 - 2\alpha u_1 D^{-1}(u_2\cdot) - \tfrac{2}{3}\beta D^{-1}(u_2\cdot), \tag{2.24}$$

$$\Delta = D^2 + 2\alpha u_1^2 + \tfrac{2}{3}b - 2\alpha u_1 D^{-1}(u_2\cdot) + \tfrac{1}{3}u_1 D^{-1}(b'\cdot). \tag{2.25}$$

Letting α or β equal zero in (2.24), we obtain the recursion operator of the PKdV or of the PMKdV, respectively. Letting $b = 0$ in (2.25), we obtain the recursion operator of the PMKdV. Letting $\alpha = 0$ in (2.25), we obtain the recursion operator of the linear combination of the KdV[3] and of the MKdV.[3] It is easily checked that both (2.24) and (2.25) are hereditary operators. Further, it is interesting that if we start with (2.25) and require that it is a hereditary operator we find out that this is the case iff b satisfies (2.18b). Equation (2.18b) also appears when applying Δ to $u_3 + \alpha u_1^3 + b(u)u_1$ in order to obtain the generator (2.18c). Let us consider only the terms involving integration

$$\tfrac{1}{3}u_1 D^{-1}(b'u_3 + \alpha b'u_1^3 + bb'u_1) - 2\alpha u_1 D^{-1}(u_2u_3 + \alpha u_2 u_1^2 + bu_2u_1).$$

The terms involving $bb'u_1$, $\alpha u_2 u_1^2$, and u_2u_3 integrate exactly and so we are left with

$$\tfrac{1}{3}u_1 D^{-1}(b'u_3 + \alpha b'u_1^3) - 2\alpha u_1 D^{-1}(bu_2u_1).$$

Integrating the first term by parts and ignoring the part integrated exactly, we are left with

$$-\tfrac{1}{3}u_1 D^{-1}[u_2u_1(b'' + 8\alpha b)],$$

which is exactly integrable iff Eq. (2.18b) holds.

2. A Bäcklund Transformation

It is well known tht the KdV equation is related to the MKdV equation through the Miura transformation. It is interesting that Eq. (2.18) is also related to the MKdV equation [trivially generalized, see (2.28) below]. Taking into consideration (2.18b), Eq. (2.18a) becomes

$$u_t + u_3 + \alpha u_1^3 + (\tau_1 e^{2\sqrt{-2\alpha}u} + \tau_2 e^{-2\sqrt{2\alpha u}} + \tau_3)u_1 = 0, \tag{2.26}$$

where τ_1, τ_2, τ_3 are constant parameters. The Bäcklund transformation

$$u_1 + \kappa v + \lambda + \left(\frac{\tau_1}{3\alpha}\right)^{1/2} e^{u\sqrt{2\alpha}} + \left(\frac{\tau_2}{3\alpha}\right)^{1/2} e^{-u\sqrt{2\alpha}} = 0, \tag{2.27}$$

where κ, λ are constant parameters, maps Eq. (2.26) to

$$v_t + v_3 + \left[3\alpha(\kappa v + \lambda)^2 + \tau_3 - \frac{1}{3\alpha}(\tau_1\tau_2)^{1/2}\right]v_1 = 0. \tag{2.28}$$

Note that if Eq. (2.27) is viewed as an ordinary differential equation with x and u as the independent and dependent variables, respectively, (t is regarded as a parameter) then it is of the Riccati type. If we put $u = (1/\sqrt{-2\alpha})\ln w$, (2.27) becomes

$$w_x + (-\tfrac{2}{3}\tau_2)^{1/2} + (-2\alpha)^{1/2}(\kappa v + \lambda)w + (-\tfrac{2}{3}\tau_1)^{1/2}w^2 = 0.$$

3. COMPARISON WITH OTHER METHODS AND OPEN QUESTIONS

The most obvious approach to establishing the exact solvability of a given equation is to guess operators L and A such that the given equation can be expressed in the form

$$L_t = [A,L]. \tag{3.1}$$

However, this approach is the least practical, since both operators A and L must be guessed. A way out is to assume the form of L and then find all equations that correspond to it. In this respect there exist two basic approaches; (i) Gel'fand and Dikii[12] assume L and then, by solving Eq. (3.1) algebraically, find all equations that correspond to it. (ii) AKNS[13] (see also Ref. 14) assume a given L, but rather than using Eq. (3.1) directly, they determine all equations corresponding to this L by requiring that the evolution of the scattering data takes a simple form. This method has been extended by Newell.[15] The above approach has the advantage that it also paves the way for the actual solution of the evolution equation involved, but has the weakness that it starts with a given eigenvalue problem and finds all equations that correspond to it, rather than starting directly with a given equation.

The most widely used direct method for finding whether a given equation is exactly solvable has been developed by Estabrook and Wahlquist.[16] This method consists, briefly, of the following (for consistence of presentation we do not use the language of differential forms employed in Ref. 16): Find functions $A(u, Q)$ and $B(u, u_1,...,u_{n-1}, Q)$ (the assumption made about the dependence of A and B is based on experience) such that the equations $Q_x = A$, $Q_t = B$ are compatible when u satisfies the given nth-order equation. This easily leads to $A = \Sigma_j a_j(u)\xi_j(Q)$, $B = \Sigma_j b_j(u,u_1,...,u_{n-1})\xi_j(Q)$, where the functions a_j and b_j are completely determined and the ξ_j satisfy some given commutator relations. The main problem now is to find a closed algebra of ξ_j and then a representation of this algebra in terms of Q. Also, sometimes it is necessary to allow Q to be a vector.

Another direct approach is introduced in Ref. 17, where A in Eq. (3.1) is taken to be the adjoint of $K'(u)$, and L is a recursion operator connecting polynomial solutions of the equation $\psi_t + A\psi = 0$. These solutions are simply related to the conservation laws of the given equation by a theorem due to Lax.[18] The authors of Ref. 17 employ a perturbative scheme to find conservation laws and then L. A weakness of this method is that it is applicable only to exactly solvable equations with infinitely many conservation laws. However, there exist exactly solvable equations with a finite number of conservation laws (for example, the Burgers equation).

The formalism of taking as a Lax pair $K'(u)$ and the operator Δ (which recursively relates solutions of the equation $\psi_t + K'(u)\psi = 0$) is developed in Ref. 7. However, this formalism was not directly related to the symmetry structure of the given equation. This is done (apparently independently) in Ref. 6; see also Sec. 1. The advantage of this approach is that Eq. (3.1) has to be solved only for L, since $A = K'(u)$ is explicitly known.

In this paper we emphasize that the knowledge of one generalized symmetry makes it possible to obtain Δ almost by inspection. That is why we concentrate on finding such a generalized symmetry; the relevant algorithm employed is very straightforward. Furthermore, demanding that an equation of a certain order admit a generator of a given order, we obtain in a straightforward way (the algorithm involved is linear) all such equations. Our method has the weakness that it is applicable to equations of normal type only. Furthermore, having obtained Δ, we must solve the equations

$$\Delta\psi = \lambda\psi, \quad \psi_t + K'(u)\psi = 0. \tag{3.2}$$

However, these equations are not in a very convenient form, since the first equation involves an integral operator. A proper transformation makes it possible to transform the above equations to differential ones (for example, in the case of the KdV this is achieved by taking $\psi = \phi^2$). A general method for doing this as well as an investigation of Eq. (2.18) will be presented in a future publication.

Apparently, there exists an intimate connection between our method and Estabrook-Wahlquist's one. By asking two different questions (namely, when a given equation has a nontrivial prologation structure and when it admits a generalized symmetry) we obtain similar answers[19] (see also Sec. 2.1). The problem of relating these two methods is under investigation. The problem of extending the results obtained here to equations of less restricted form is also under investigation. For example, results have been obtained for nonlinear heat equations.

We hope that the results presented here together with those of Refs. 20 and 21 (where the group-theoretical nature of BT and of the constants of motion admitted by evolution equations is established) as well as those of Refs. 22–26, indicate the importance played by symmetries in understanding and solving the problems appearing in the analysis of nonlinear evolution equations.

ACKNOWLEDGMENTS

The author thanks P. A. Lagerstrom for his comments on an early draft of this paper and for many interesting discussions. He also thanks A. C. Newell for suggesting to the author that he investigate whether (2.18) is related to the KdV or the modified KdV by means of a Bäcklund transformation.

APPENDIX A

In this appendix we indicate briefly how Eq. (2.4) and (2.5) are obtained. Equation (2.2) admits the LB operator associated with the generator (2.3) iff (see Lemma 1)

$$[X(u_2 + A(u,u_1)), X(u_3 + B(u,u_1,u_2))] = 0, \tag{A1}$$

or

$$\Big(D^3 + \sum_{j=0}^{2} B_{j+1} D^j\Big)(u_2 + A) = (D^2 + A_1 + A_2 D)(u_3 + B), \tag{A2}$$

where

$B_{j+1} = \partial B/\partial u_j$, $A_{j+1} = \partial A/\partial u_j$, $j = 0,1,2,$

or

$$D^2B + \sum_{j=0}^{1} A_{j+1} + D^jB + \sum_{1}^{2} A_j u_{j+2}$$
$$= D^3A + \sum_{j=0}^{2} B_{j+1} D^jA + \sum_{1}^{3} B_j u_{j+1}. \quad \text{(A3)}$$

Writing out $D^jA, D^jB, (j = 1, 2, 3)$ explicitly in Eq. (A3) and then equating to zero the coefficients of u_3^2 and u_3, we obtain

$$B_{33} = 0, \quad 2DB_3 = 3DA_2, \quad \text{(A4)}$$

or

$$B = \tfrac{3}{2}A_2u_2 + \alpha u_2 + F(u,u_1). \quad \text{(A5)}$$

The parameter α generates the t-translation group so we set it equal to zero. Now substituting (A5) in (A3) and equating to zero the coefficients of $u_2{}^j$ $(j = 3, 2, 1, 0)$, we obtain

$$A_{222} = 0, \quad F_{22} = \tfrac{3}{2}A_2A_{22} + 3A_{12},$$
$$2F_{12}u_1 = \tfrac{3}{2}AA_{12} + \tfrac{3}{2}A_1A_{22}u_1 + \tfrac{3}{2}A_2A_{12}u_1 + \tfrac{3}{2}A_{112}u_1^2 + 3A_{11}u_1,$$
$$A_1F + A_2F_1u_1 + F_{11}u_1^2 + AF_1 + A_1F_2u_1 + \tfrac{3}{2}A_2A_{11}u_1^2 + A_{111}u_1^3 = 0. \quad \text{(A6)}$$

Solving Eqs. (A6) and taking into consideration (A5), we obtain

(i)

$$A = a(u)u_1^2, \quad \text{(A7a)}$$

where a is an arbitrary function of u.

(ii)

$$A = \frac{b''}{b'}u_1^2 + bu_1, \quad \text{(A7b)}$$

where $b(u)$ is an arbitrary function of u and $b' = db/du$. Equations (A7) can be combined into one by letting $b \to \beta b$. Then

$$A = \frac{b''(u)}{b'(u)}u_1^2 + \beta b(u)u_1 \quad \text{(A8a)}$$

[and $\beta = 0$ gives (A7a)]. To the above A there corresponds

$$B = \frac{b'''}{b'}u_1^3 + \frac{3b''}{b'}u_1u_2 + \tfrac{3}{2}\beta\left(b' + \frac{bb''}{b'}\right)u_1^3 + \tfrac{3}{2}\beta bu_2 + \tfrac{3}{4}\beta^2b^2u_1. \quad \text{(A8b)}$$

(iii)

$$A = (\gamma - c')/cu_1^2 + c, \quad \text{(A9)}$$

where c is an arbitrary function of u, B is given by (2.5b).

APPENDIX B

In this appendix we indicate briefly how Eqs. (2.17) and (2.18) are obtained. Eq. (2.15) admits the LB operator associated with the generator (2.16) iff

$$[X((u_3 + A(u,u_1)), X(u_5 + B(u,...,u_4))] = 0, \quad \text{(B1)}$$

or

$$(D^3 + A_1 + A_2D)(u_5 + B)$$
$$= (D^5 + B_1 + \sum_{j=2}^{4} B_jD^{j-1})(u_3 + A) = 0. \quad \text{(B2)}$$

Writing out $D^jA, D^jB, (j = 1, 2, 3, 4)$ and equating the coefficients of u_6 and u_5 in (B2) to zero, we obtain

$$\partial B/\partial u_4 = 0, \quad 3DB_4 = 5DA_2, \quad \text{(B3)}$$

or

$$B = \tfrac{5}{3}A_2u_3 + F(u,u_1,u_2). \quad \text{(B4)}$$

Replacing B in (B2) by (B4) and equating the coefficients of u_4 to zero, we obtain

$$F = \tfrac{5}{6}A_{22}u_2^2 + \tfrac{5}{3}A_{12}u_1u_2 + \tfrac{5}{3}A_1u_2 + g(u,u_1). \quad \text{(B5)}$$

Replacing B in (B2) by (B4), where F is given by (B5) and equating the coefficients of $u_3u_2^j$, $(j = 3, 2, 1, 0)$ to zero, we obtain, respectively,

$$A_{222} = 0, \quad A_{2221} = 0,$$
$$3g_{22} = 5A_2A_{22} + 10A_{112}u_1 + 5A_{11} + 5A_{1122}u_1^2, \quad \text{(B6)}$$
$$3g_{12}u_1 = \tfrac{5}{3}AA_{12} + \tfrac{5}{3}A_1A_{22}u_1 + \tfrac{10}{3}A_2A_{12}u_1 + \tfrac{10}{3}A_{1112}u_1^3 + 5A_{111}u_1^2.$$

Equating the coefficients of u_2^3, u_2^2 in (B2) to zero, we obtain

$$A_{2211} = 0, \quad (A_2A_{221} - A_1A_{222})u_1 + A_1A_{22} - AA_{221} = 0. \quad \text{(B7)}$$

Solving Eqs. (B6a), (B6b), and (B7a), we obtain

$$A = \alpha u_1^3 + \gamma uu_1^2 + \beta u_1^2 + b(u)u_1 + c(u). \quad \text{(B8)}$$

Eq. (B7b) gives $\gamma = 0, \beta b' = 0, \beta c' = 0$. Therefore,

$$A = \alpha u_1^3 + \beta u_1^2 + b(u)u_1 + c(u), \quad \text{(B9)}$$
$$B = \tfrac{5}{3}A_2u_3 + \tfrac{5}{6}A_{22}u_2^2 + \tfrac{5}{3}A_{12}u_1u_2 + \tfrac{5}{3}A_1u_2 + g(u,u_1), \quad \text{(B10)}$$

where

$$\beta b' = 0, \quad \beta c' = 0, \quad \text{(B11)}$$

and we have still to satisfy the compatibility equation of (B6c) and (B6d)

(i) $\beta \neq 0$. Then Eqs. (B9) and (B10) indicate that

$$A = \alpha u_1^3 + \beta u_1^2 + \gamma u_1 + \delta. \quad \text{(B12)}$$

The compatibility equation of (B6c) and (B6d) is then satisfied and, by integrating them, we obtain

$$g = \tfrac{5}{6}\int (A_2)^2 du_1. \quad \text{(B13)}$$

Replacing g in (B10) by (B13), where A is given by (B12), we obtain (2.17b).

(ii) $\beta = 0$. Then Eqs. (B9) and (B10) indicate that

$$A = \alpha u_1^3 + bu_1 + c. \quad \text{(B14)}$$

The compatibility equation of (B6c) and (B6d) gives

$$b''' + 8\alpha b' = 0, \quad \alpha c' = cb' = 0. \quad \text{(B15)}$$

If $c \neq 0$ we obtain trivial results, therefore

$$A = \alpha u_1^3 + b(u)u_1, \quad \text{(B16)}$$

where b satisfies (2.18b). Integrating (B6c), (B6d) we obtain g and then, using (B10), we obtain the generator (2.18c).

[1]P. D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968).

[2]N. H. Ibragimov and R. L. Anderson, J. Math. Anal. Appl. **59**, 145 (1977).

[3]P. J. Olver, J. Math. Phys. **18**, 1212 (1977).

[4]The author knows of only one exception, which is a fifth-order equation considered in P. J. Caudrey, R. K. Dodd, J. D. Gibbon, Proc. R. Soc. (London) Ser. A **351**, 407 (1976).
[5]A. S. Fokas and P. A. Lagerstrom, "Quadratic and Cubic Invariants in Classical Mechanics," to appear in J. Math. Anal. Appl. (1980).
[6]B. Fuchssteiner, Nonlinear Anal. **3**, 849 (1979).
[7]N. Asano and Y. Kato, Prog. Theor. Phys. **58**, 161 (1977).
[8]A. S. Fokas and P. A. Lagerstrom, "On the Use of LB Operators in Quantum Mechanics," to appear in J. Math. Anal. Appl. (1980).
[9]F. Magri, Nuovo Cimento B **34**, 334 (1976).
[10]A. S. Fokas, Ph.D. Thesis, California Institute of Technology, 1979.
[11]D. J. Kaup, "Lectures on the Estabrook–Wahlquist Method, with Examples of Application," Plasma Research Report, PRR 79/6 (1979).
[12]I. M. Gel'fand and L. A. Dikii, Funct, Anal. Appl. **10**, 13 (Russian), 259 (English) (1976); Funct. Anal. and its Appl. **11**, 11 (Russian), 93 (English) (1977).
[13]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **31**, 125 (1973); Stud. Appl. Math. **53**, 249 (1974).
[14]H. Flaschka and A. C. Newell, "Integrable Systems of Nonlinear Evolution Equations," in *Dynamical Systems, Theory and Applications*, edited by J. Moser (Springer, Berlin, 1975).
[15]A. C. Newell, Proc. R. Soc. (London) A **365**, 283 (1979).
[16]H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **16**, 1 (1975); J. Math. Phys. **17**, (1976); see also the contribution of the above authors in the *Proceedings of the International Symposium on Nonlinear Evolution Equations Solvable via the IST*, Academia Naxionale dei Liucei, Roma, June 15–18, 1977, edited by F. Calogevo, for a rather complete bibliography.
[17]H. H. Chen, Y. C. Lee, and C. S. Liu, "Integrability of Nonlinear Hamiltonian Systems by the ISM" (preprint).
[18]P. D. Lax, Commun. Pure Appl. Math. **28**, 141 (1975).
[19]F. A. E. Pirani (private communication).
[20]A. S. Fokas, and R. L. Anderson, Lett. Math. Phys. **3**, 17 (1979).
[21]A. S. Fokas, Lett. Math. Phys. **3**, 467 (1979).
[22]P. J. Olver, "Symmetry Groups and Conservation Laws in the Formal Variational Calculus" (preprint).
[23]S. Kumei, J. Math. Phys. **18**, 256 (1977).
[24]B. Fuchssteiner, "Application of Spectral-Gradient Methods to Nonlinear Soliton Equations" (preprint).
[25]Y. Kosmann-Schwarzbach, "Generalized Symmetries of Nonlinear Partial Differential Equations," to appear in Lett. Math. Phys.
[26]F. Magri, J. Math. Phys. **19**, 1156 (1978).

Evolution of a stable profile for a class of nonlinear diffusion equations. III. Slow diffusion on the line

James G. Berryman
Bell Laboratories, Whippany, New Jersey 07981

(Received 27 August 1979; accepted for publication 16 November 1979)

We show that the nonlinear diffusion equation $\partial n/\partial t = \partial^2(n^{1+\delta})/\partial x^2$ with compact initial data on $-\infty < x < \infty$ can be transformed into another nonlinear diffusion equation $\alpha(\partial\theta/\partial t) = \theta^{1+\alpha} \times \partial^2\theta/\partial\hat{x}^2$ on a fixed finite interval of the $\hat{x}$ axis. Thus, the original moving boundary problem is transformed into a fixed boundary problem. The new form of the equation has advantages both analytically and computationally as the examples illustrate. Linear stability analysis for the transformed equation is straightforward whereas the moving boundaries of the original problem complicate the analysis for that case. The advantages of the resulting computational algorithm for solving the moving boundary problem are also discussed. A nonlinear Rayleigh–Ritz quotient and a Lyapunov functional are shown to be bounded, monotonically decreasing functions of time. Both functionals give added insight into the mathematical character of the diffusion process.

I. INTRODUCTION

Consider the nonlinear diffusion equation

$$\frac{\partial n}{\partial t} = \frac{\partial}{\partial x}\left[D(n)\frac{\partial n}{\partial x}\right], \quad \text{for} \quad -\infty < x < \infty, \qquad (1)$$

where n is a particle density, x is the spatial variable, and t is the time. The diffusion coefficient D is a function of the density which we will take to be

$$D(n) = (1+\delta)n^\delta \qquad (2)$$

for simplicity. The boundary conditions are $n \to 0$ as $|x| \to \infty$ for all t, plus the physical requirement that the total number of particles be conserved. In this third paper in our series,[1,2] we will restrict the exponent to be $0 < \delta < \infty$ and we will assume that the initial data $n(x, 0) \geqslant 0$ have compact support. The conserved integrals and the self-similar solution of Eq. (1) are discussed in Appendix A.

The resulting mathematical problem arises in many contexts. When $\delta = 1$, Eq. (1) is a model of classical diffusion in a plasma. For porous media,[3] Eq. (1) is a model of the motion of a polytropic gas whose pressure and density satisfy $p = \text{const.}\, n^\delta$, where δ is the polytropic exponent. If the flow is isothermal, $\delta = 1$; if the flow is adiabatic, then $\delta > 1$. For thermal waves,[4] the variable n is reinterpreted as a temperature. Then Eq. (1) is a model of radiation heat conduction with $\delta = 13/2$ for a fully ionized gas or $\delta \simeq 4.5$–5.5 in regions of multiply ionized gas. Similarly, electron heat conduction in a plasma can be modeled by Eq. (1) when $\delta = 5/2$. Equation (1) may also be viewed as a nonlinear one-phase generalized Stefan problem.[5,6]

Equation (1) has been studied extensively. Zel'dovich and Kompaneets[4,7,8] introduced the nonlinear problem as an alternative to linear heat conduction since it is known that when $\delta = 0$ the effect of any perturbation is propagated instantaneously through all space while the propagation of the heat is retarded when $\delta > 0$. Kalashnikov[9] gave the first rigorous proof that if the initial data of Eq. (1) with $\delta > 0$ have compact support, then the solution will have compact support at any finite time. Zel'dovich and Barenblatt[4,10] discussed the importance of the self-similar solution as the asymptotic solution for arbitrary initial data and also provided a linear stability analysis. Kamin (Kamenomostskaya)[11,12] gave the first rigorous proof that the self-similar solution represents the asymptotic solution by using a scaling argument. Aronson[3,13,14] has studied the regularity properties of the solution and, in particular,[13] the properties of the interface curves (i.e., moving boundaries) of the region of compact support. In a recent article, Knerr[15] provides a brief review of past rigorous results for Eq. (1) and also extends Aronson's results on the interface. Existence and uniqueness results for weak solutions of Eq. (1) were established by Oleinik, Kalashnikov, and Yui-Lin.[16,17]

As mentioned previously, for $\delta = 0$ (linear diffusion) the influence of a point source of particles at $x = t = 0$ is felt at every point x for any finite $t > 0$, i.e., the propagation speed is infinite. By contrast, the propagation speed for Eq. (1) is finite when $\delta > 0$. Thus, in comparison to a linear diffusion process, Eq. (1) implies a relatively slow diffusion process for all $\delta > 0$; hence, the subtitle of this paper is "Slow diffusion on the line." However, it should be emphasized that the nonlinear diffusion process may not be slow compared to other processes in a given physical environment. For example, Zel'dovich and Raizer[4] note that, when radiation heat conduction is important (at temperatures greater than or $\sim 10^4$–$10^{5\circ}$K), the nonlinear "conduction mechanism can transfer energy at a speed much faster than the speed of sound in the medium... because the speed of light at nonrelativistic temperatures is very much greater than the speed of sound." *Slow diffusion* may be very *fast* in such circumstances but the propagation speed is still finite and therefore much slower than that for linear diffusion in similar circumstances.

Our main task in the present paper is to show that Eq. (1), with its implicit moving boundaries, may be transformed into another type of nonlinear diffusion equation with *fixed boundaries*. The transformation we use is a generalization of one recently used by Rosen[18] in a different context. Although the boundaries are stationary, this new class of nonlinear diffusion equations does not belong to the same class we studied previously.[1,2] Nevertheless, it turns out that the same methods of analysis may be applied. In Sec. II, we

 0022-2488/80/061326-06$1.00

present the mathematical transformation and then provide a physical interpretation. The transformation is illustrated in Sec. III by application to the self-similar solution. The perturbation analysis is performed in Sec. IV and then used to provide estimates of the boundary motion in the original problem. Section V discusses the computational advantages of the fixed boundary problem. Appendix B discusses two bounded auxiliary functionals which decrease monotonically in time and which may be used to gain insight into the mathematical character of the evolving diffusion process.

II. TRANSFORMATION AND INTERPRETATION

With the definitions $\theta = n^{1+\delta}$, $q = (2+\delta)/(1+\delta)$, and $\alpha = q-1$, Eq. (1) becomes

$$\partial(\theta^{\alpha})/\partial t = \partial^2\theta/\partial x^2, \quad \text{for } -\infty < x < \infty, \tag{3}$$

which turns out to be a more convenient form of the equation both analytically and computationally. Note that with $0<\delta<\infty$, we have $1<q<2$ and $0<\alpha<1$.

Following Rosen,[18] we now introduce new space and time coordinates $\hat{x}$ and $\hat{t}$ defined by

$$\partial\hat{x}/\partial x = \theta^{\alpha}, \quad \partial\hat{x}/\partial t = \partial\theta/\partial x, \quad \hat{t} = t. \tag{4}$$

We see immediately that Eq. (3) is just an integrability condition on $\hat{x}$, i.e., $\partial(\partial\hat{x}/\partial x)/\partial t = \partial(\partial\hat{x}/\partial t)/\partial x$. The line integral which determines $\hat{x}(x,t)$ is just

$$\hat{x} = \int(\theta^{\alpha}\,dx + \partial\theta/\partial x\,dt). \tag{5}$$

The inverse transform to Eq. (4) is given by

$$\partial x/\partial\hat{x} = \hat{\theta}^{-\alpha}, \quad \partial x/\partial\hat{t} = -\partial\hat{\theta}/\partial\hat{x}, \quad t = \hat{t}, \tag{6}$$

where $\hat{\theta}$ is related to θ by

$$\hat{\theta}(\hat{x},\hat{t}) = \theta(x,t). \tag{7}$$

The inegrability condition for x is

$$\partial(\hat{\theta}^{-\alpha})/\partial\hat{t} = -\partial^2\hat{\theta}/\partial\hat{x}^2,$$

or equivalently (8)

$$\alpha(\partial\hat{\theta}/\partial\hat{t}) = \hat{\theta}^{1+\alpha}(\partial^2\hat{\theta}/\partial\hat{x}^2).$$

The line integral of $x(\hat{x},\hat{t})$ is

$$x = \int(\hat{\theta}^{-\alpha}\,d\hat{x} - \partial\hat{\theta}/\partial\hat{x}\,d\hat{t}). \tag{9}$$

Equations (3)–(9) include all the formulas needed to transform back and forth between the nonlinear diffusion equations (3) and (8). However, what is the advantage of doing so since both equations are nonlinear and therefore difficult to solve analytically? To understand this, consider Eq. (3) when the initial data have compact support. Then we may choose to compute the line integral in Eq. (5) by starting at $x=-\infty$, $t=0$, integrating along the t axis from 0 to t, and then integrating along the x axis from $-\infty$ to x. The resulting expression for $\hat{x}$ is

$$\hat{x}(x,t) = \int_{-\infty}^{x}\theta^{\alpha}(y,t)\,dy = \int_{-\infty}^{x}n(y,t)\,dy \tag{10}$$

since $\partial\theta/\partial x$ clearly vanishes at $x=-\infty$ for all finite t and since $\theta^{\alpha}=n$ follows from the definition of θ. If $x=a(t)$ and $x=b(t)$ are, respectively, the left and right boundaries of the region of finite density, we find that

$$\hat{x}(a,t) = 0, \tag{11}$$

and

$$\hat{x}(b,t) = \int_a^b n(y,t)\,dy = N, \tag{12}$$

where N is the total number of particles and therefore independent of time (see Appendix A). It follows that $\hat{x}(x,t)$ may be interpreted physically as the total number of particles to the left of x at time t. Thus, the moving boundaries $x=a(t)$ and $b(t)$ transform into fixed boundaries for Eq. (8) in an intuitively appealing way. The following sections demonstrate the advantages of the fixed boundary problem.

A physical interpretation may also be given to the definition

$$\partial x/\partial\hat{t} = -\partial\hat{\theta}/\partial\hat{x} = -\frac{\partial\theta}{\partial x}\frac{\partial x}{\partial\hat{x}} = -\sigma^{-1}(\partial(\theta^{\sigma})/\partial x). \tag{13}$$

Following Aronson,[13] we define $v = n^{\delta} = \theta^{2-q} = \theta^{\sigma}$ so Eq. (13) becomes

$$\partial x/\partial\hat{t} = -(1+\delta^{-1})\partial v/\partial x. \tag{14}$$

When Eq. (14) is evaluated at $x=a(t)$ or $b(t)$, the formula determines the speed of the boundaries. Aronson's result for the speed of the boundaries is the same as Eq. (14) except when $v_x = 0$ at a boundary. Aronson's theorem leaves the interface speed undefined when v_x vanishes. From Eq. (14), we expect the boundary speed to vanish when v_x vanishes and this expectation agrees with Knerr's result.[15]

In the next section, we illustrate these transformations by studying the self-similar solution.

III. SELF-SIMILAR SOLUTION

The self-similar solution of Eq. (3) is exhibited in Appendix A. In this section, we show that the self-similar solution for Eq. (3) transforms into a separable solution for Eq. (8). Then we study the separable solution's shape function.

Evaluating Eq. (10) for the self-similar solution, we find easily that

$$\hat{x}(x,t) = \int_{-z_0}^{z} S^{\alpha}(z)\,dx, \tag{15}$$

where $z = xT^{\alpha}(t)$. The integral in Eq. (15) can be expressed in terms of the incomplete beta function. Note that $\hat{x}$ depends on x and t only through the similarity variable z. This means that

$$\theta(x,t) = S(z)T(t) = \hat{S}(\hat{x})T(t) = \hat{\theta}(\hat{x},\hat{t}), \tag{16}$$

or

$$\hat{S}(\hat{x}) = S(z), \tag{17}$$

with z related to $\hat{x}$ implicitly by Eq. (15).

Note that $\sigma = 2-q$ so $-\alpha = \sigma-1$. Also note that the change of scale $\hat{x}\to\hat{x}/N$ and $\hat{t}\to\hat{t}/N^2$ leaves Eq. (8) invariant so that, without loss of generality, we may set $N=1$. Then Eq. (8) becomes

$$\partial(\hat{\theta}^{\sigma-1})/\partial\hat{t} = \partial^2\hat{\theta}/\partial\hat{x}^2, \quad \text{for } 0\leqslant\hat{x}\leqslant 1, \tag{18}$$

with the boundary conditions $\hat{\theta}(0,t) = \hat{\theta}(1,t) = 0$. Equation (18) is identical in form to the nonlinear diffusion equations studied in Refs. 1 and 2. The only difference is that the exponent σ was restricted to the range $1 < \sigma < \infty$ in those papers whereas $0 < \sigma < 1$ here. In fact, all the formulas derived[1] for the shape function may be carried over without change since the restrictions were placed on this parameter for physical, not for mathematical, reasons. Listing the relevant equations from Ref. 1, we have

$$\hat{S}'' + \lambda \hat{S}^{\sigma-1} = 0, \tag{19}$$

which is solved by

$$I(\hat{S}) = \int_0^{\hat{S}} \frac{dy}{(1-y^\sigma)^{1/2}} = \rho \hat{x}, \quad \text{for } 0 \leqslant \hat{x} \leqslant 1/2, \tag{20}$$

where

$$\rho = \frac{2}{\sigma} \frac{\Gamma(1/2)\Gamma(1/\sigma)}{\Gamma((1/2)+(1/\sigma))}, \tag{21}$$

and

$$I(\hat{S}) = \frac{1}{\sigma} B_{\hat{S}^\sigma}\left(\frac{1}{\sigma}, \frac{1}{2}\right) = \hat{S}\,{}_2F_1\left(\frac{1}{\sigma}, \frac{1}{2}; 1 + \frac{1}{\sigma}; \hat{S}^\sigma\right). \tag{22}$$

The special functions which appear in Eqs. (21) and (22) are the gamma function Γ, the incomplete beta function B, and the hypergeometric function ${}_2F_1$. The eigenvalue λ is given by

$$\lambda = \tfrac{1}{2}\sigma\rho^2. \tag{23}$$

[The eigenvalue in Appendix A was chosen to be equal to the λ in Eq. (23).] Two relevant integrals are given by

$$\gamma = \int_0^1 \hat{S}^{\sigma-1}(\hat{x})\, d\hat{x} = 4/\sigma\rho \tag{24}$$

and

$$c = \int_0^1 \hat{S}^\sigma(\hat{x})\, d\hat{x} = 2/(2+\sigma). \tag{25}$$

The relationship between σ and δ is given by

$$\sigma = \delta/(1+\delta). \tag{26}$$

Table I illustrates the values of several of these constants as δ and σ are varied.

The formula (20) determines $\hat{S}(\hat{x})$ implicitly. Since the analytical form of the self-similar solution for Eq. (3) is known, it is clear that no advantage has been gained by transforming to Eq. (8) or (18) in this particular example. However, we will show that important gains are made for the perturbation analysis and also for numerical computations in the following two sections.

IV. PERTURBATION ANALYSIS

Repeating the argument of Ref. 1, consider

$$\hat{\theta}(\hat{x}, \hat{t}) = \hat{S}(\hat{x})T(\hat{t}) + u_l(\hat{x})v_l(\hat{t}), \tag{27}$$

where $u_l v_l$ is a small (separable) perturbation to the separable solution of Eq. (18). The analysis may be carried through exactly as in Ref. 1 simply by replacing q everywhere by σ. We find that

$$u_l'' + \kappa_l \hat{S}^{\sigma-2} u_l = 0, \tag{28}$$

and

$$v_l(\hat{t}) = T^{p_l}(\hat{t}), \tag{29}$$

where

$$p_l = 2 - \sigma + \kappa_l/\lambda, \quad \text{for } l \geqslant 1. \tag{30}$$

The solution of Eq. (28) is given in Ref. 2. Here we will quote only the eigenvalues

$$\kappa_l = \tfrac{1}{2}(l+1)(\sigma l + 2)\lambda, \quad \text{for all } l = 0, 1, 2, \cdots. \tag{31}$$

As before, $u_0(\hat{x}) = \hat{S}(\hat{x})$ and $\kappa_0 = \lambda$. Once again note that, for $l = 1$, we have $\kappa_1 = (\sigma+2)\lambda$ and therefore $p_1 = 4$. For this case, $u_1 \propto \hat{S}\hat{S}'$. Thus, the slowest decaying (infinitesimal) perturbation to the separable solution decays as the fourth power of the separable solution's time factor.

In light of our previous work, the present perturbation analysis turns out to be extremely simple. By contrast, the perturbation analysis performed for Eq. (3) by Zel'dovich and Barenblatt[10] suffers from the serious conceptual limitation that the analysis dealt with a region whose boundaries are determined by the boundaries of the self-similar solution. However, as these authors have noted,[10] the perturbations themselves affect the boundary location so that results derived in this way are of uncertain value. These difficulties do not arise in our approach because we have rigorously transformed the moving boundary problem into a fixed boundary problem.

Now let us consider briefly how the perturbations alter the boundaries of the original problem (1) or (3). To transform back, we must use Eq. (9). In this case, we choose to begin the line integration at $\hat{x} = \frac{1}{2}$, $\hat{t} = 0$, integrating first along the $\hat{x}$ axis and then along the $\hat{t}$ axis. The resulting formula is

$$x(\hat{x}, \hat{t}) = \int_{1/2}^{\hat{x}} \hat{\theta}^{\sigma-1}(y, 0)\, dy - \int_0^{\hat{t}} \frac{\partial \hat{\theta}}{\partial \hat{x}}(\hat{x}, \tau)\, d\tau. \tag{32}$$

The first integral depends only on the initial data and may be replaced by $x(\hat{x}, 0)$ which is determined implicitly when Eq. (10) is evaluated at $t = 0$. Within perturbation theory, the second integral in Eq. (32) may be evaluated by noting that the u_l's form a complete set of expansion functions so that

$$\hat{\theta}(\hat{x}, \hat{t}) = \sum_{l=0}^{\infty} a_l(\hat{t}) u_l(\hat{x}), \tag{33}$$

where the a_l's satisfy an infinite set of coupled ordinary

TABLE I. Values of σ, ρ, λ, and γ for selected values of δ. The defining equations in the text are Eqs. (26), (21), (23), and (24), respectively. The slope of the separable solution shape function is $\pm\rho$ near the boundaries. The separation constant is λ and from Eq. (48) $\gamma = 2z_0$.

δ	σ	ρ	λ	γ
1/3	1/4	256/35	8192/35	35/16
1/2	1/3	32/5	512/15	15/8
1	1/2	16/3	64/9	3/2
2	2/3	$3\pi/2$	$3\pi^2/4$	$4/\pi$
5/2	5/7	4.5784	7.4864	1.2231
3	3/4	4.4870	7.5499	1.1886
∞	1	4	8	1

(though nonlinear) differential equations.[1] If the perturbations are sufficiently small, we may write

$$a_0(\hat{t}) \simeq T(\hat{t}),$$

and

$$a_l(\hat{t}) \simeq \epsilon_l T^{p_l}(\hat{t}), \quad \text{for } l \geqslant 1, \tag{34}$$

where the ϵ_l's are small constants dependent upon the initial data. Then we find

$$\int_0^{\hat{t}} \frac{\partial\hat{\theta}}{\partial\hat{x}}(\hat{x},\tau)\,d\tau \simeq \lambda^{-1}[T^{\sigma-1}(\hat{t}) - 1]\hat{S}'(\hat{x}) + (1-\sigma)\sum_{l=1}^{\infty}\frac{\epsilon_l}{\kappa_l}[1 - T^{\kappa_l/\lambda}(\hat{t})]u_l'(\hat{x}). \tag{35}$$

The time-dependent factor in the first term is monotonically increasing while all the others are monotonically decreasing. Because $T^{\kappa_l/\lambda}$ decays at least as fast as T for all $l \geqslant 1$ and $1 < q < 2$ (since $\kappa_1/\lambda q = 4/q - 1 > 1$), we may may neglect these time-dependent terms asymptotically. Therefore, we have

$$x(\hat{x},\hat{t}) \underset{\hat{t}\to\infty}{\longrightarrow} x(\hat{x},0) - (1-\sigma)\sum_{l=1}^{\infty}\frac{\epsilon_l}{\kappa_l}u_l'(\hat{x}) - \lambda^{-1}[(\hat{t}/t_0)^{1-1/q} - 1]\hat{S}'(\hat{x}). \tag{36}$$

Evaluating Eq. (36) at either $\hat{x} = 0$ or $\hat{x} = 1$ gives the asymptotic perturbation formula for the motion of the boundary. Equation (36) agrees asymptotically with Eq. (42) evaluated at $\pm z_0$ since $z_0 = \lambda^{-1}\rho = \gamma/2$ follows from Eq. (48), where $S_0 = 1$.

V. COMPUTATIONAL ALGORITHM

Besides the previously demonstrated analytical advantages obtained with the transformation to fixed boundaries, computational advantages are also gained.

With fixed boundaries, Eq. (8) can be solved numerically with the same three-level method of Lees[19,20] which was used in Refs. 1 and 2. The only additional complications involve the transformations between x and $\hat{x}$ [Eqs. (10) and (32)]. The transformation (10) needs to be employed only once (at $t = 0$)

$$\hat{x}(x,0) = \int_{-\infty}^{x} \theta^{\sigma-1}(y,0)\,dy, \tag{37}$$

thereby transforming into the domain of Eq. (8). Thereafter, $\hat{\theta}(\hat{x},\hat{t})$ evolves according to Eq. (8). Whenever we wish to determine $\theta(x,t)$ from $\hat{\theta}(\hat{x},\hat{t})$, we must evaluate

$$x(\hat{x},\hat{t}) = x(\hat{x},0) - \int_0^{\hat{t}} \frac{\partial\hat{\theta}}{\partial\hat{x}}(\hat{x},\tau)\,d\tau, \tag{38}$$

where $x(\hat{x},0)$ is determined implicitly by Eq. (37). It follows that we must compute $\partial\hat{\theta}/\partial\hat{x}$ as well as $\hat{\theta}$ at each grid point in $\hat{x}$ and $\hat{t}$. These are very modest complications compared to the computational difficulties inherent in treating the singular moving boundaries of the original problem (3).

ACKNOWLEDGMENT

I benefited greatly from discussions and correspondence with Professor Charles J. Holland during an early phase of this work.

APPENDIX A. CONSERVED INTEGRALS AND THE SELF-SIMILAR SOLUTION

Equation (1) has two conserved integrals. In the absence of sources and sinks, the total number of particles must be conserved. We see that

$$\frac{d}{dt}\int_{-\infty}^{\infty} n(x,t)\,dx = \int(\partial^2(n^{1+\delta})/\partial x^2)\,dx = \partial(n^{1+\delta})/\partial x\,\Big|_{a(t)}^{b(t)} = 0, \tag{39}$$

where a and b are the boundaries of the region of compact support. Thus, the flux $\partial(n^{1+\delta})/\partial x$ must vanish at the boundaries as expected physically. The second conserved integral is the center of mass (or density) which follows from

$$\frac{d}{dt}\int_{-\infty}^{\infty} xn\,dx = \int x(\partial^2(n^{1+\delta})/\partial x^2)\,dx = [x\partial(n^{1+\delta})/\partial x]\,\Big|_{a(t)}^{b(t)} - n^{1+\delta}\Big|_{a(t)}^{b(t)} = 0. \tag{40}$$

In order for Eq. (1) to have a classical solution satisfying these two physical conservation laws, both the flux and the density must vanish at the boundaries. If the initial density distribution is sufficiently smooth, these conditions will be satisfied for all $t > 0$.

Since the center of mass is stationary, we are free to pick its location due to the translational invariance of Eq. (1). We assume throughout this paper that the center of mass is located at $x = 0$.

Equation (3) has a self-similar solution of the form (also see Refs. 4, 7, 8, 10, and 21)

$$\theta(x,t) = S(z)T(t), \tag{41}$$

with

$$z = x/R(t). \tag{42}$$

The relation

$$R(t) = T^{1-q}(t) \tag{43}$$

follows from the requirement of particle conservation (39). Substituting Eqs. (41)–(43) into Eq. (3) yields the equations

$$S'' + \lambda[S^{q-1} + (q-1)zS^{q-2}S'] = 0,$$

or equivalently

$$S' + \lambda z S^{q-1} = 0, \tag{44}$$

and

$$(q-1)T^{-1-q}\dot{T} = -\lambda, \tag{45}$$

which have the solutions

$$S(z) = \begin{cases} S_0(1 - z^2/z_0^2)^{1/\sigma}, & \text{for } z^2 \leqslant z_0^2, \\ 0, & \text{otherwise,} \end{cases} \tag{46}$$

and

$$T(t) = (1 + t/t_0)^{-1/q}. \tag{47}$$

The constants appearing in Eqs. (46) and (47) are given by

$$\sigma = 2 - q, \quad z_0^2 = 2S_0^{\sigma}(\sigma\lambda)^{-1}, \quad \text{and} \quad t_0 = \left(1 - \frac{1}{q}\right)\lambda^{-1}. \tag{48}$$

The factor S_0 is an arbitrary constant while λ is the separation constant in Eqs. (44) and (45). To determine these constants, we consider the properties of the solution.

From the conservation of particles, we have

$$N=\int_{-\infty}^{\infty} n(x,t)\,dx = \int_{-z_0}^{z_0} S^{q-1}(z)\,dz = (2\pi S_0^q/\sigma\lambda)^{1/2}\frac{\Gamma(1/\sigma)}{\Gamma((1/2)+(1/\sigma))}. \quad (49)$$

So we find that

$$S_0^q = \frac{\sigma\lambda}{2\pi}\left[N\frac{\Gamma((1/2)+(1/\sigma))}{\Gamma(1/\sigma)}\right]^2, \quad (50)$$

showing that S_0^q is directly proportional to λ, which is still a free parameter.

From Eq. (48), we see that λ is inversely proportional to t_0. The meaning of t_0 is simple: When $t=-t_0$, Eqs. (46) and (47) are singular, indicating that if a point source of particles of strength N were placed at $x=0$, $t=-t_0$, the self-similar solutions (46) and (47) would evolve.[21] For later convenience [see Eq. (23)], we will choose t_0 so that

$$\lambda = \frac{2\pi}{\sigma}\left[\frac{\Gamma(1/\sigma)}{\Gamma((1/2)+(1/\sigma))}\right]^2, \quad (51)$$

and

$$S_0 = N^{2/q}. \quad (52)$$

APPENDIX B. MONOTONE AUXILIARY FUNCTIONALS

Two bounded, monotonically decreasing functionals are known for Eq. (18). These functionals are inherently interesting since their behavior gives additional insight into the mathematical properties of the nonlinear diffusion process. For simplicity of notation, we will drop the carets from $\hat{\theta}$, $\hat{S}$, $\hat{x}$, and $\hat{t}$ in this Appendix.

1. Nonlinear Rayleigh–Ritz quotient

Following Ref. 2, we define

$$Q(t) = c^{-1}\int_0^1 \theta^\sigma(x,t)\,dx \quad (53)$$

and

$$R(t) = c^{-1}\int_0^1 \theta_x^2(x,t)\,dx. \quad (54)$$

Then

$$\frac{d}{dt}Q(t) = -\frac{\sigma}{1-\sigma}R(t)\leqslant 0 \quad (55)$$

and

$$\frac{d}{dt}R(t) = -\frac{2c^{-1}}{1-\sigma}\int \theta_{xx}^2\theta^{2-\sigma}\,dx \leqslant 0. \quad (56)$$

Furthermore, Schwarz's inequality implies that

$$R^2(t)\leqslant Q(t)c^{-1}\int \theta_{xx}^2\theta^{2-\sigma}\,dx. \quad (57)$$

It follows from Eqs. (55)–(57) that the nonlinear Rayleigh–Ritz quotient

$$\mathscr{R}(\theta) = R(t)/Q^{2/\sigma}(t) \quad (58)$$

is a monotonically decreasing function of time, i.e.,

$$\frac{d}{dt}\mathscr{R}(\theta)\leqslant 0. \quad (59)$$

Equality is achieved in Eq. (59) only when θ is the separable solution.

Using a variational argument, it is also straightforward to show that

$$\mathscr{R}(\theta)\geqslant\lambda, \quad (60)$$

where λ is the eigenvalue in Eq. (19). Thus, the nonlinear diffusion equation (18) causes $\theta(x,t)$ to evolve so $\mathscr{R}(\theta)$ monotonically approaches its minimum value λ and therefore θ asymptotically approaches the separable solution.

II. Lyapunov functional

Following Ref. 22, define a new time variable τ such that

$$\tau = -\lambda^{-1}\ln U(t), \quad (61)$$

where

$$U(t) = (\zeta_0 + t/t_0)^{-1/q} \quad (62)$$

and ζ_0 is a constant which will be specified later. Also define a new dependent variable such that

$$W(x,\tau) = \theta e^{\lambda\tau}. \quad (63)$$

Then from Eq. (8) it follows easily that

$$W_\tau = W^q W_{xx} + \lambda W. \quad (64)$$

Next, define the functional

$$I(W) = \frac{1}{2}\int W_x^2\,dx - \frac{\lambda}{\sigma}\int W^\sigma\,dx, \quad (65)$$

which satisfies

$$\frac{d}{d\tau}I(W) = -\int W^{-q}W_\tau^2\,dx\leqslant 0. \quad (66)$$

Equality occurs in Eq. (66) only when $W_\tau\equiv 0$ or $W_{xx}+\lambda W^{\sigma-1}=0$, which is identical to Eq. (19). Thus, the functional I is monotonically decreasing in τ unless $W\equiv S$.

To show that $I(W)$ is bounded below, we must show that $\int W^\sigma dx$ is bounded above. To demonstrate this, reconsider Eqs. (60) and (55), which together imply

$$\frac{d}{dt}Q^{-q/\sigma}\geqslant t_0^{-1}. \quad (67)$$

Upon integration, we find

$$Q(t)\leqslant[Q^{-q/\sigma}(0)+t/t_0]^{-\sigma/q}. \quad (68)$$

If we now choose $\zeta_0 = Q^{-q/\sigma}(0)$ in Eq. (62), Eq. (68) implies that

$$\int_0^1 W^\sigma(x,\tau)\,dx\leqslant c = \int_0^1 S^\sigma(x)\,dx. \quad (69)$$

This fact proves that $I(W)$ is bounded below for all τ.

Again we have found a bounded functional monotonically decreasing in t unless θ is the separable solution. The existence of these functionals is heuristic evidence that the separable solution represents the asymptotic solution for arbitrary initial data.

[1]J.G. Berryman, J. Math. Phys. **18**, 2108 (1977).
[2]J.G. Berryman and C.J. Holland, J. Math. Phys. **19**, 2476 (1978).
[3]D.G. Aronson, SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **17**, 461 (1969).

[4]Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Academic, New York, 1967), Vol. 2, Chap. 10.
[5]A.B. Taylor, in *Moving Boundary Problems in Heat Flow and Diffusion*, edited by J.R. Ockendon and W.R. Hodgkins (Clarendon, Oxford, 1975), p. 120.
[6]H.S. Carslaw and J.C. Jaeger, *Conduction of Heat in Solids* (Clarendon, Oxford, 1959), Chap. 11.
[7]Ya. B. Zel'dovich and A.S. Kompaneets, in *Lectures Dedicated on the* 70th *Anniversary of A.F. Ioffe*, Akademii Nauk SSSR, 1950 (Russian).
[8]L.D. Landau and E.M. Lifshitz, *Fluid Mechanics* (Pergamon, London, 1959), p. 195.
[9]A.S. Kalashnikov, Zh. Vychisl. Mat. Mat. Fiz. **7**, 440 (1967) (Russian).
[10]Ya. B. Zel'dovich and G.I. Barenblatt, Sov. Phys. Dokl. **3**, 44 (1958).
[11]S. Kamenomostskaya (S. Kamin), Israel J. Math. **14**, 76 (1973).
[12]S. Kamin, Arch. Rational Mech. Anal. **60**, 171 (1976).
[13]D.G. Aronson, Arch. Rational Mech. Anal. **37**, 1 (1970).
[14]D.G. Aronson, SIAM (Soc. Ind. Math. Appl.) J. Appl. Math. **19**, 299 (1970).
[15]B.F. Knerr, Trans. Am. Math. Soc. **234**, 381 (1977).
[16]O.A. Oleinik, A.S. Kalashnikov, and Yui-Lin Czou, Izv. Akad. Nauk SSSR Ser. Mat. **22**, 667 (1958) (Russian).
[17]O.A. Oleinik, *Seminari* 1962–1963 *di Analisi, Algebra, Geometria e Topologia*, (Ist. Naz. Alta Mat., Ediz. Cremonese, Roma, 1965) Vol. 1, p. 355.
[18]G. Rosen, Phys. Rev. B **19**, 2398 (1979).
[19]M. Lees, Math. Comput. **20**, 516 (1966).
[20]A.R. Mitchell, *Computational Methods in Partial Differential Equations* (Wiley, New York, 1969), p. 96.
[21]R.E. Pattle, Q. J. Mech. Appl. Math. **12**, 407 (1959).
[22]J.G. Berryman and C.J. Holland, Phys. Rev. Lett. **40**, 1720 (1978); C.J. Holland and J.G. Berryman, in *Applied Nonlinear Analysis*, edited by V. Lakshmikantham, Proceedings of an International Conference on Applied Nonlinear Analysis, University of Texas at Arlington, 20–22 April 1978 (Academic, New York, 1979), pp. 61–66.

A Hilbert–Padé method for multipole approximations. Application to the Gaussian function

Pablo Martín
Universidad Simón Bolívar, Departamento de Física, Apartado 80659-Caracas, Venezuela
Jorge Zamudio-Cristi
Universidad Metropolitana, Departamento de Física, Apartado 76819-Caracas, Venezuela
Guillermo Donoso
Universidad Simón Bolívar, Departamento de Física, Apartado 80659-Caracas, Venezuela

(Received 30 October 1979; accepted 24 January 1980)

A method is developed where a Hilbert transform is combined with an asymptotic Padé method in order to obtain good multipole approximations for functions whose power series have a large radius of convergence. This method has been used to find two- to eight-pole approximations for the Gaussian function.

I. INTRODUCTION

In previous papers[1] we discuss a modified procedure to the Padé method[2] where both the asymptotic expansion and the power series for the same function are considered. In this way we obtain a fractional approximated function (multipole approximation) which gives better agreement to the exact function for small and large values of the independent variable than the usual Padé method. However, this procedure fails when applied to some functions as the Gaussian, where the radius of convergence of the power series is very large (infinite in this case) and an asymptotical series cannot be obtained. For those cases we have developed a method of obtaining a good multipole approximation to those functions by using a Hilbert transform in combination with the asymptotic Padé's method in the line of the Borel–Padé method.[3] Our procedure can be used for several kinds of functions; however, the method is described here in relation to the Maxwellian distribution, where a four-pole approximation has been proposed recently.[4] In that paper the authors have shown that the function $1/(v^2+B)^2+C^2$ appears to approximate the Maxwellian reasonably well and they correlate this approximation to the two-pole approximation for the plasma dispersion function $Z(s)$ given by Fried, *et al.*[5] Here we discuss multipole approximations for the normalized Maxwellian distribution function $f(v) = \exp(-v^2)/\sqrt{\pi}$ and we correlate it with the two-pole and multipole approximations to the $Z(s)$. Each pole in the $Z(s)$ approximation generates two poles in the $f(v)$ approximation. A new Lorentzian approximation for $f(v)$ is obtained by a one-pole approximation for $Z(s)$, using the asymptotic Padé method.[1]

First we consider the general case of approximating $f(v)$ by a sum of simple rational fractions, and, by the symmetry and reality conditions, we correlate the poles of the different fractions. In this way we find a general form in terms of the independent poles. Except for very small values of v, our method gives better approximation than a straightforward Padé method.

In Sec. II the theoretical foundations of the method is presented and the computation of the $2n$-approximation for $f_n(v)$ is obtained by using the corresponding main approximation for $Z(s)$ with n poles. In Sec. III the results are presented in a graphical form and the analysis and discussion are carried out. The conclusions are presented in the last Section.

II. THEORETICAL ANALYSIS

We look for a fractional approximation for a given function $f(v)$ of the form

$$f_{\text{approx}}(v) = \frac{1}{\sqrt{\pi}} \sum_i \frac{c_i}{v-a_i} \cong f(v), \tag{1}$$

where the coefficients c_i and the poles a_i are, in general, complex numbers.

The general procedure will be to obtain the Hilbert transform $\tilde{f}(s)$ of $f(v)$ and to determine the power series and the asymptotic expansion of this transformed function. Then we obtain a fractional approximation $\tilde{f}_{\text{approx}}(s)$ to this transformed function by using the asymptotic Padé method previously discussed.[1] Finally, by applying the inverse Hilbert transform to the approximated transformed function, an approximated fractional function $f_{\text{approx}}(v)$ for the original $f(v)$ function is obtained.

The method is better described in relation to the Gaussian function, as we do here. However in this case, before applying the general procedure, it is convenient to reduce the number of independent coefficients by means of the symmetry and reality conditions of this function.

A first consideration is that the poles a_i of the approximated Guassian function can not be real because $f(v)$ does not show any singularity for v real. In this case the reality condition of $f(v)$ implies that each pole a_i must have a corresponding complex conjugate, and the same applies to the coefficients c_i. Thus,

$$f_{\text{approx}}(v) = \frac{1}{\sqrt{\pi}} \sum_{i=1}^{n} \left(\frac{c_i}{v-a_i} + \frac{c_i^*}{v-a_i^*} \right). \tag{2}$$

Besides, the symmetry condition $f(v) = f(-v)$ imposes a further restriction on the number of independent poles. If

 0022-2488/80/061332-04$1.00

a_i is pure imaginary, then c_i must be pure imaginary. If a_i is complex, then each pole a_i must have a corresponding pole of opposite sign, and similarly for the c_i's. Therefore, it is convenient to separate the complex poles a_i and coefficients c_i from the pure imaginary poles and coefficients which will be designated by α_k and γ_k respectively. Then

$$f_{\text{approx}} = \frac{1}{\sqrt{\pi}} \sum_{j=1}^{p} \left(\frac{c_j}{v - a_j} + \frac{c_j^*}{v - a_j^*} - \frac{c_j}{v + a_j} - \frac{c_j^*}{v + a_j^*} \right) + \frac{1}{\sqrt{\pi}} \sum_{k=1}^{r} \left(\frac{\gamma_k}{v - \alpha_k} + \frac{\gamma_k^*}{v - \alpha_k^*} \right). \quad (3)$$

The total number of poles $2n$ is related to r and p by the equation

$$2p + r = n. \quad (4)$$

Since in Eq. (3) the a_j are not real, we can choose all the poles to be in the lower half plane without losing generality. It is useful to designate the terms of the sum in Eq. (3) by g_j and h_k. Then

$$g_j = \frac{1}{\sqrt{\pi}} \left(\frac{c_j}{v - a_j} + \frac{c_j^*}{v - a_j^*} - \frac{c_j}{v + a_j} - \frac{c_j^*}{v + a_j^*} \right) = \frac{1}{\sqrt{\pi}} \frac{E_j v^2 + F_j}{(v^2 + B_j)^2 + C_j^2}, \quad (5)$$

$$h_k = \frac{1}{\sqrt{\pi}} \left(\frac{\gamma_k}{v - \alpha_k} + \frac{\gamma_k^*}{v - \alpha_k^*} \right) = \frac{1}{\sqrt{\pi}} \frac{H_k}{v^2 + G_k}. \quad (6)$$

where E_j, F_j, B_j, C_j, H_k, and G_k are real numbers defined as

$$\begin{aligned} E_j &= 4\,\mathrm{Re}(a_j c_j), \\ F_j &= -4\,|a_j|^2\,\mathrm{Re}(a_j c_j^*), \\ B_j &= -\mathrm{Re}(a_j^2), \\ C_j &= \mathrm{Im}(a_j^2), \\ H_k &= 2\alpha_k \gamma_k = -2\,|\alpha_k|\,|\gamma_k|, \\ G_k &= -\alpha_k^2 = |\alpha_k|^2. \end{aligned} \quad (7)$$

The value h_k is obtained from $\frac{1}{2} g_j$ when a_j and c_j are pure imaginaries.

Following now the general procedure, we apply the Hilbert transform to the approximated and exact functions. Thus we obtain for $\mathrm{Im} s > 0$

$$Z_{\text{approx}}(s) = \int_{-\infty}^{\infty} \frac{f_{\text{approx}}(t)}{t - s}\, dt = \frac{1}{\sqrt{\pi}} \sum_{j=1}^{p} \left(\frac{2\pi i c_j}{s - a_j} - \frac{2\pi i c_j^*}{s + a_j^*} \right) + \frac{1}{\sqrt{\pi}} \sum_{k=1}^{r} \frac{2\pi i \gamma_k}{s - \alpha_k}. \quad (8)$$

The two-pole approximation of Fried *et al.*[5] is equivalent to taking $p = 1$ and $r = 0$ in Eq. (8). Our values a_1 and c_1 are given in terms of their parameters, $a^{-1} = \xi + i\eta$, by

$$a_1 = a = |a|^2(\xi - i\eta) = (0.55^2 + \pi/4)^{-1}(0.55 - i\sqrt{\pi}/2), \quad (9)$$

$$c_1 = \frac{\sqrt{\pi}}{2\pi i} \frac{-1}{2\xi a} = -\frac{1}{8\times 0.55} + \frac{i}{4\sqrt{\pi}}. \quad (10)$$

The g_1 defined in Eq. (5) is the four-pole approximation given by Ward,[4] denoted here by f_W. Since a_1 and c_1 come out pure imaginary, from Eq. (7), E_1 becomes zero. Therefore,

$$g_1 = \frac{1}{\sqrt{\pi}} \frac{|a|^4}{[v^2 + |a|^4(\pi/4 - \xi^2)]^2 + \pi |a|^8 \xi^2} = -\frac{1}{\sqrt{\pi}} \frac{0.845}{(v^2 + 0.408)^2 + 0.6787}. \quad (11)$$

The coefficients B_1 and C_1 are now identified with B and C as defined by Ward, and F_1 come out equal to $B^2 + C^2$.

In relation to the general multipole approximation for the Z function (one to four poles), our coefficients c_j and γ_k are obtained from the corresponding b-pole residues of $Z_{lm}(s)$ by the equations

$$c_j = -\frac{b_j}{2\sqrt{\pi}}, \quad \gamma_k = -\frac{i b_k}{2\sqrt{\pi}}. \quad (12)$$

Here $Z_{lm}(s)$ is the n-pole approximation for $Z(s)$,

$$Z_{lm}(s) = \sum_{j=1}^{p} \left(\frac{b_j}{s - a_j} + \frac{b_j^*}{s + a_j} \right) + \sum_{k=1}^{r} \frac{b_k}{s - \alpha_k}, \quad (13)$$

obtained by means of l-terms of the power series and m-terms of the asymptotic expansion.

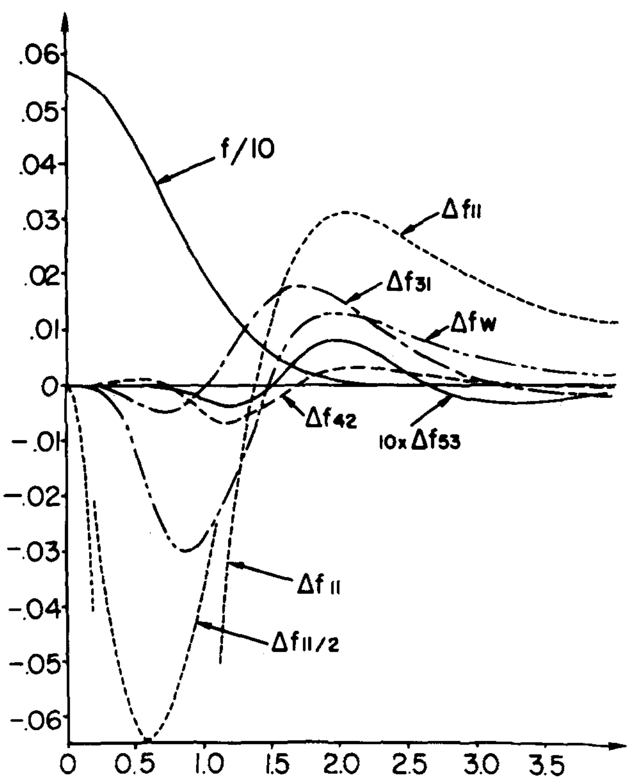

FIG. 1. Deviations Δf_{lm} from the exact value, for the best n-pole approximations. The deviation of Ward approximation is shown for comparison. The values of the exact function (divided by 10) are also shown. To be able to show the different deviations in the same graph, part of Δf_{11} is shown divided by 2, and Δf_{53} is shown multiplies by 10.

TABLE I. Poles (a_j, α_k), coefficients (c_j, γ_k) and parameters of Eq. (7) of the best approximations f_{lm} for different number of poles.

	$n=1$	
f_{11}	$\alpha_1 = -\dfrac{1}{\sqrt{\pi}} i$	$\gamma_1 = \frac{1}{2} i$
	$= -0.56i$	$= 0.28i$
	$G_1 = \dfrac{1}{\sqrt{\pi}}$	$H_1 = \dfrac{1}{\pi}$
	$= 0.318$	$= 0.318$
	$n=2$	
f_W	$a_1 = 0.505 - 0.815i$	$c_1 = -0.227 + 0.141i$
	$B_1 = 0.408$	$C_1^2 = 0.6787$
	$E_1 = 0$	$F_1 = 0.845$
f_{31}	$a_1 = \dfrac{(23\pi - 32 - 4\pi^2)^{1/2}}{2(4-\pi)} - \dfrac{\sqrt{\pi}}{2(4-\pi)} i$	$c_1 = \dfrac{1}{(23\pi - 32 - 4\pi^2)^{1/2}} + \dfrac{1}{4\sqrt{\pi}} i$
	$= 0.5138 - 1.0324i$	$= -0.3636 + 0.1410i$
	$B_1 = \dfrac{2\pi^2 + 16 - 11\pi}{2(4-\pi)^2} = 0.8018$	$C_1^2 = \dfrac{(23\pi - 32 - 4\pi^2)}{4(4-\pi)^4} = 1.1257$
	$E_1 = -\dfrac{\pi - 3}{4 - \pi} = -0.1649$	$F_1 = \dfrac{(\pi-2)^2}{(4-\pi)^2} = 1.7686$
	$n=3$	
f_{42}	$a_1 = 0.9050 - 1.1317i$	$c_1 = -0.2088 - 0.1639i$
	$\alpha_1 = -1.2278i$	$\gamma_1 = 0.6098i$
	$B_1 = 0.4617$	$C_1^2 = 4.1959$
	$E_1 = -1.4978$	$F_1 = 0.0292$
	$G_1 = 1.5075$	$H_1 = 1.4974$
	$n=4$	
f_{53}	$a_1 = 1.2359 - 1.2150i$	$c_1 = 0.0105 - 0.1542i$
	$a_2 = 0.3786 - 1.3509i$	$c_2 = -0.5929 + 0.2953i$
	$B_1 = -0.0512$	$C_1^2 = 9.0192$
	$E_1 = -0.6975$	$F_1 = -2.4069$
	$B_2 = 1.6816$	$C_2^2 = 1.0463$
	$E_2 = 0.6978$	$F_2 = 4.9081$

We will denote by $f_{lm}(v)$ the function whose Hilbert transform is $Z_{lm}(s)$. This notation is now more adequated than the previously used $f_{\text{approx}}(v)$ in Eqs. (1)–(3). The poles a_j, α_k and the pole-residues b_j, b_k for $Z_{lm}(s)$ have been already determined.[1] Hence the coefficients c_j and γ_k can be computed.

For a given number of poles we have computed only the coefficients corresponding to the best approximation to the exact function, which are shown in Table I, together with the values for E_j, F_j, B_j, C_j, H_k, and G_k.

III. ANALYSIS AND DISCUSSION

In Fig. 1, differences between the approximated functions $f_{lm}(v)$ and the exact function $f(v)$ are shown for v in the interval (0,4). In the same figure the differences for the best of the four-pole approximation f_W given by Ward (Eq. 15 of Ref. 4) is also shown.

All the new approximations, except the two pole approximation, give better agreement than f_W.

In Table II, the maximum absolute deviations Δf_{lm} are also indicated. The relative deviations are irrelevant since $f(v)$ goes to zero very quickly for large v. The approximations improve with the number of poles.

TABLE II. Maximum absolute deviation Δf_{lm} and the values of v and f where this deviation is found for different number of poles.

	$(\Delta f_{lm})_{\max}$	$v_{\text{for max}}$	$f(v)_{\text{for max}}$
f_{11}	-0.129	0.60	0.394
f_W	-0.030	0.90	0.251
f_{31}	0.018	1.70	0.031
f_{42}	-0.007	1.20	0.134
f_{53}	0.0008	2.00	0.010

Improvements of one order of magnitude are obtained by increasing the number of poles from one to two and from three to four. Also, by increasing the number of poles, the maximum deviation is obtained by larger values of v.

In connection with the moments, it must be pointed out that the exact function has finite values for moments of any order. For the approximated functions, from some order on, all the even moments are infinite. However the mean square velocities are finite except for the Lorentzian approximations (two-pole).

It is covenient to point out that fractional approximations to the Maxwellian distribution can obviously be obtained by a straightforward Padé method. However, for a given number of poles the Padé method results in very poor approximations compared with ours. Only in the case of very small values of v can the Padé method be compared favorably with ours.

IV. SUMMARY AND CONCLUSIONS

A method has been devised to find good multipole approximations for a given function, by means of the Hilbert transform and the asymptotic Padé method. The application of this method is adequate for functions where the radius of convergence of the power series is very large.

In the case of the Gaussian function, we have used the reality and symmetry conditions in order to reduce the number of independent poles. The procedure requires finding previously a multipole approximation for the plasma dispersion function Z. Using the known approximation for Z, we have determined approximations to Maxwellian distributions with two, four, six, and eight poles. All of our approximations with four or more poles give better agreement with the exact function than those found by Ward.

ACKNOWLEDGMENT

Work supported in part by the Venezuelan Government Agency CONICIT, Project No. 31.26.S1.0606.

[1]Pablo Martín and Miguel Angel González, Phys. Fluids **22**, 1413 (1979); P. Martín, G. Donoso, and J. Zamudio-Cristi, J. Math. Phys. **21**, 280 (1980).
[2]G.A. Baker Jr., *Essentials of Padé's Approximants* (Academic, New York, 1975).
[3]S. Graffi, V. Grecchi, and G. Simon, Phys. Lett. B **32**, 631 (1970); G.J. Ruggieri and D.J. Thoules, J. Phys. F. **6**, 2063 (1976).
[4]Michael A.V. Ward, Phys. Fluids **20**, 1372 (1977).
[5]B.D. Fried, C.L. Hedrick, and J. McCune, Phys. Fluids **11**, 249 (1978).

Existence and completeness of the Möller wave operators for radial potentials satisfying $\int_0^1 r|v(r)|dr + \int_1^\infty |v(r)|\,dr < \infty$

J. D. Dollard and C. N. Friedman
Department of Mathematics, University of Texas, Austin, Texas 78712

(Received 14 May 1979; accepted for publication 25 June 1979)

We give an elementary proof of the existence of the (three-dimensional) Möller wave operators and the unitarity of the S operator (weak asymptotic completeness) for radial Kato potentials $v(r)$ satisfying $\int_0^1 r|v(r)|dr + \int_1^\infty|v(r)|dr < \infty$.

1. INTRODUCTION

In the present article, we shall give an elementary proof of the existence of the three-dimensional Möller wave operators and the unitarity of the S operator (weak asymptotic completeness) for radial Kato potentials $v(r)$ which satisfy the condition

$$\int_0^1 r|v(r)|\,dr + \int_1^\infty |v(r)|\,dr < \infty. \tag{1.1}$$

This result has been proved by Kuroda[1], who combined the results of Green and Lanford[2] with some trace ideal methods. However, our proof is so elementary that we feel it may add considerable insight to the result, and the method of proof may be of some independent interest. Basically, our method consists of analyzing the asymptotic form of solutions of the one-dimensional radial Schrödinger equation for fixed angular momentum, and a consistent use of the spreading of wave packets under time evolution. This method is quite similar to (although somewhat simpler than) that of a previous article[3] by the present authors. In analyzing solutions of the radial Schrödinger equation we shall make use of product integration methods which are summarized[3] for the case of continuous integrands. In fact, in the present context we shall need to product integrate functions which are Lebesgue integrable rather than continuous; in this setting the construction of the product integral is slightly different, and the differential equation satisfied by the product integral has to be interpreted as holding almost everywhere or may be replaced by the corresponding integral equation, but essentially nothing else is changed. (See Refs. 4, 5, or 6 for detailed study of the product integral.)

2. SOLUTIONS OF THE RADIAL SCHRÖDINGER EQUATION AND CONSTRUCTION OF WAVE PACKETS

Suppose that $v(r)$ is a radial function defined almost everywhere on $\mathbf{R}^3$ and satisfying

$$v \in L^2(\mathbf{R}^3,dx) + L^\infty(\mathbf{R}^3,dx) \tag{2.1}$$

[where $x = (x_1,x_2,x_3)$, dx = Lebesgue measure on $\mathbf{R}^3$, and $r = |x| = (x_1^2 + x_2^2 + x_3^2)^{1/2}$], and (1.1). Consider the partial differential operators

$$H_0 = -\Delta = -\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}\right),$$
$$H = -\Delta + v(r). \tag{2.2}$$

H_0 and H act on locally integrable functions $\psi(x_1,x_2,x_3)$ in the sense of distributions and define self-adjoint operators in $L^2(\mathbf{R}^3,dx)$ with domains

$$\mathscr{D}(H_0) = \{\psi \in L^2 : H_0\psi \in L^2\},$$
$$\mathscr{D}(H) = \{\psi \in L^2 : H\psi \in L^2\}, \tag{2.3}$$

and $\mathscr{D}(H_0) = \mathscr{D}(H)$.

Let S_{lm} be subspaces of $L^2(\mathbf{R}^3,dx)$ given by

$$S_{lm} = \{\psi \in L^2 : \psi(x) = R(r)Y_{lm}(\theta,\varphi)\}, \tag{2.4}$$

where Y_{lm} is a spherical harmonic; then S_{lm} reduces H_0 and H, and if we put

$$rR = \varphi, \tag{2.5}$$

then on S_{lm}, H_0 and H are unitarily equivalent to operators h_0 and h on $L^2((0,\infty),dr)$ given by

$$h_0\varphi = \left(-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2}\right)\varphi,$$
$$h\varphi = [h_0 + v(r)]\varphi. \tag{2.6}$$

Since $\mathscr{D}(H_0) = \mathscr{D}(H)$, we have $\mathscr{D}(h_0) = \mathscr{D}(h)$. In Ref. 3 it was shown that $\mathscr{D}(h_0)$ consists of those φ in $L^2((0,\infty),dr)$ such that $(-d^2/dr^2 + l(l+1)/r^2)\varphi$ computed in the sense of distributions on the open set $(0,\infty)$ is given by a function in $L^2((0,\infty),dr)$ *and* $\varphi(r) = 0(r)$ as $r \to 0$.

We now analyze the asymptotic behavior as $r \to \infty$ of solutions of the equation

$$h\varphi(r) = k^2\varphi(r), \quad k > 0. \tag{2.7}$$

First, rewrite (2.7) as

$$\begin{pmatrix}\varphi\\ \varphi'\end{pmatrix}' = \begin{pmatrix}0 & 1\\ w(r) - E & 0\end{pmatrix}\begin{pmatrix}\varphi\\ \varphi'\end{pmatrix},$$
$$w(r) = v(r) + \frac{l(l+1)}{r^2}. \tag{2.8}$$

We now proceed as in Ref. 3, Eqs. (1.21)–(1.28).

Setting

$$\Phi(r) = \begin{pmatrix} \varphi(r) \\ \varphi'(r) \end{pmatrix}, \tag{2.9}$$

$$M = \begin{pmatrix} 1 & 1 \\ ik & -ik \end{pmatrix}, \tag{2.10}$$

$$P(r) = \begin{pmatrix} e^{ik(r-r_0)} & 0 \\ 0 & e^{-ik(r-r_0)} \end{pmatrix}, \tag{2.11}$$

$$A_k(s) = \frac{w(s)}{2ik}\begin{pmatrix} 1 & e^{-2ik(s-r_0)} \\ -e^{2ik(s-r_0)} & -1 \end{pmatrix}, \tag{2.12}$$

we have

$$\Phi(r) = MP(r)\prod_{r_0}^{r} e^{A_k(s)\,ds} M^{-1}\cdot\Phi(r_0), \quad r_0 > 0. \tag{2.13}$$

[We remark that $\Pi_{r_0}^{r} e^{A_k(s)\,ds}$ is equal to the time ordered exponential $T\exp(\int_{r_0}^{r} A_k(s)\,ds)$.] Since $A_k(s)\in L^1((r_0,\infty),ds)$ by (1.1), $\lim_{r\to\infty} \Pi_{r_0}^{r} e^{A_k(s)\,ds} = \Pi_{r_0}^{\infty} e^{A_k(s)\,ds}$ exists by Theorem 1.12 of Ref. 3, and we have

$$\begin{aligned}\Phi(r) &= MP(r)\prod_{r_0}^{\infty} e^{A_k(s)\,ds} M^{-1}\cdot\Phi(r_0) \\ &\quad + MP(r)\left(\prod_{r_0}^{r} e^{A_k(s)\,ds} - \prod_{r_0}^{\infty} e^{A_k(s)\,ds}\right) M^{-1}\cdot\Phi(r_0) \\ &= MP(r)\prod_{r_0}^{\infty} e^{A_k(s)\,ds} M^{-1}\cdot\Phi(r_0) \\ &\quad - MP(r)\int_r^{\infty} ds\, A_k(s)\prod_{r_0}^{s} e^{A_k(\xi)\,d\xi} M^{-1}\cdot\Phi(r_0) \\ &\equiv MP(r)[C_k + R_k(r)]\cdot\Phi(r_0),\end{aligned} \tag{2.14}$$

where

$$C_k = \prod_{r_0}^{\infty} e^{A_k(s)\,ds} M^{-1}, \tag{2.15}$$

$$R_k(r) = \int_r^{\infty} ds\, A_k(s)\prod_{r_0}^{s} e^{A_k(\xi)\,d\xi} M^{-1}. \tag{2.16}$$

Since $R_k(r)\to 0$ as $r\to\infty$, it follows from (2.14) that every solution of (2.7) is of the form

$$\varphi(r) = C_+ e^{ikr} + C_- e^{-ikr} + o(1) \quad \text{as } r\to\infty. \tag{2.17}$$

In Ref. 3 it was shown that there exists a solution of (2.7) which is $O(r^{l+1})$ uniformly on bounded k intervals as $r\to 0$. [Since $v(r)$ is not necessarily continuous, "solution" means solution in the L^2 or distributional sense or solution of the integral equation equivalent to (2.7)]. Putting these facts together, we have that for each $k>0$ there is a solution $\varphi(l,v,k,r)$ of (2.7) satisfying

$$\begin{aligned}\varphi(l,v,k,r) &= O(r^{l+1}) && \text{as } r\to 0, \\ \varphi(l,v,k,r) &= \sin\left(kr - \frac{l\pi}{2} + \delta_l(k)\right) + o(1) && \text{as } r\to\infty.\end{aligned} \tag{2.18}$$

[The phase term $-l\pi/2$ is chosen for convenience since, as is well-known, for $v=0$ (2.18) holds with $\delta_l(k)=0$.]

We now construct wave packets from $\varphi(l,v,k,r)$.

Definition: For $f(k)$, a C^∞ function with compact support in $(0,\infty)$ we define

$$\varphi(l,v,f,r) = \int_0^{\infty} dk\,\varphi(l,v,k,r)f(k). \tag{2.19}$$

Theorem: (i) $\varphi(l,v,f,r)\in L^2((0,\infty),dr)$.

(ii) $\varphi(l,v,f,r)\in\mathscr{D}(h)$ and $h\varphi(l,v,f,r) = \varphi(l,v,k^2 f,r)$.

(iii) $e^{ith}\varphi(l,v,f,r) = \varphi(l,v,e^{itk^2}f,r)$.

Proof: (i) $\int_0^\infty dk\,\sin(kr - l\pi/2 + \delta_l(k))f(k) \in L^2((0,\infty),dr)$ by the Plancherel theorem. So we need only show that $\int_0^\infty dk\,P(r)R_k(r)f(k)$ has entries in $L^2((1,\infty),dr)$ [where $P(r)$ and $R_k(r)$ are as in (2.11), (2.12), and (2.16)]. [Note that by the first equation in (2.18), $\varphi(l,v,f,r) = O(r^{l+1})$ as $r\to 0$, so $\varphi(l,v,f,r)$ is in L^2 of any bounded r interval.]

To see this, write, for $1\leqslant r\leqslant\infty$,

$$\begin{aligned}&\int_0^{\infty} dk\,P(r)R_k(r)f(k) \\ &= \int_1^{\infty} ds\,w(s)\chi_{[r,\infty)}(s)\int_0^{\infty}(dk/2ik)f(k)P(r) \\ &\quad\times\begin{pmatrix} 1 & e^{-2ik}(s-r_0) \\ -e^{2ik(s-r_0)} & -1\end{pmatrix}\prod_{r_0}^{s} e^{A_k(\xi)\,d\xi}M^{-1} \\ &\equiv \int_1^{\infty} ds\,w(s)F(r,s).\end{aligned} \tag{2.20}$$

By the Plancherel theorem, the entries of $F(r,s)$ are for fixed s in $L^2((0,\infty)dr)$ and have $L^2((0,\infty)dr)$ norm bounded uniformly in s for $0\leqslant s<\infty$. [Note that $\|\Pi_{r_0}^{s} e^{A_k(\xi)\,d\xi}\|\leqslant\exp(1/k \times\int_{r_0}^{\infty}|w(\xi)|\,d\xi)$ for all s,k.] Since $w(s)\in L^1((1,\infty),ds)$, the result follows.

(ii) follows from the previous description of $\mathscr{D}(h)$ and then (iii) is easily proved by writing e^{ith} as a Taylor series to n terms with integral remainder and using (ii).

3. EXISTENCE AND COMPLETENESS OF THE WAVE OPERATORS

In this section we shall prove that

$$\Omega^{\pm} = \operatorname*{strong\,lim}_{t\to\pm\infty} e^{itH}e^{-itH_0} \tag{3.1}$$

exist on $L^2(\mathbf{R}^3,dx)$ where H_0,H are as in (2.2), and that

$$\text{range}(\Omega^+) = \text{range}(\Omega^-). \tag{3.2}$$

It follows that

$$S = (\Omega^+)^*\Omega^- \tag{3.3}$$

is unitary.

To prove (3.1) and (3.2) it suffices to prove that for each fixed l,

$$\omega^{\pm} = \operatorname*{strong\,lim}_{t\to\pm\infty} e^{ith}e^{-ith_0} \tag{3.4}$$

exist on $L^2((0,\infty),dr)$ and that

$$\text{range}(\omega^+) = \text{range}(\omega^-). \tag{3.5}$$

We consider (3.4) first, and we shall only consider the case $t\to+\infty$ since the case $t\to-\infty$ is similar. Since the $e^{ith}e^{-ith_0}$ are unitary, it suffices to prove strong convergence on the dense set of functions of the form $\varphi(l,0,f,r)$. [$\varphi(l,0,f,r)$ is the wave packet of (2.19) for $v=0$.] We have

$$\begin{aligned}&e^{ith}e^{-ith_0}\varphi(l,0,f,r) \\ &= e^{ith}\varphi(l,0,e^{-itk^2}f,r) \\ &= e^{ith}\varphi(l,v,e^{-i\delta_l(k)}e^{-itk^2}f,r)\end{aligned}$$

$$+ e^{ith}[\varphi(l,0,e^{-itk^2}f,r) - \varphi(l,v,e^{-i\delta_l(k)}e^{-itk^2}f,r)]$$
$$= \varphi(l,v,e^{-i\delta_l(k)}f,r) + e^{ith}\int_0^\infty dk f(k)e^{-itk^2}$$
$$\times[\varphi(l,0,k,r) - e^{-i\delta_l(k)}\varphi(l,v,k,r)]$$
$$= \varphi(l,v,e^{-i\delta_l(k)}f,r) + e^{ith}\int_0^\infty dk f(k)e^{-itk^2}$$
$$\times[\sin(kr - l\pi/2) - e^{-i\delta_l(k)}$$
$$\times\sin(kr - l\pi/2 + \delta_l(k))]$$
$$+ e^{ith}\int_0^\infty dk f(k)e^{-itk^2}E_k(r), \tag{3.6}$$

where

$$E_k(r) = \varphi(l,0,k,r) - \sin(kr - l\pi/2) + e^{-i\delta_l(k)}$$
$$\times[\varphi(l,v,k,r) - \sin(kr - l\pi/2 + \delta_l(k))]$$
$$\equiv E_{k,0}(r) + E_{k,v}(r). \tag{3.7}$$

It was proved in Ref. 3 that the term

$$e^{ith}\int_0^\infty dk f(k)e^{-itk^2}[\sin(kr - l\pi/2) - e^{-i\delta_l(k)}$$
$$\times\sin(kr - l\pi/2 + \delta_l(k))] \to 0$$
$$\text{in } L^2((0,\infty),dr) \text{ as } t\to+\infty. \tag{3.8}$$

We shall now show that

$$e^{ith}\int_0^\infty dk f(k)e^{-itk^2}E_{k,v}(r)\to 0$$
$$\text{in } L^2((0,\infty),dr) \text{ as } t\to+\infty. \tag{3.9}$$

Our proof will, of course, apply also to the term involving $E_{k,0}(r)$; this term can also be handled more easily because $E_{k,0}(r)$ is dominated by a fixed $L^2((0,\infty),dr)$ function for k in a compact interval and one can apply the Riemann–Lebesgue lemma and the dominated convergence theorem as in Ref. 3. (3.9) is more subtle because $E_{k,v}(r)$ is *not* generally dominated by a fixed $L^2((0,\infty),dr)$ function; e.g., if $v(r) = 0(r^{-1-\epsilon})$ as $r\to\infty$, then $E_{k,v}(r) = 0(r^{-\epsilon})$ as $r\to\infty$.]

We remark that in proving (3.9) we may ignore the e^{ith} since this is a unitary operator on $L^2((0,\infty),dr)$. Furthermore, since $E_{k,v}(r)$ is bounded near $r = 0$ (uniformly on bounded k intervals) by the first equation in (2.18), it suffices to prove that

$$\int_0^\infty dk f(k)e^{-itk^2}E_{k,v}(r)\to 0 \quad \text{in } L^2((1,\infty),dr)$$
$$\text{as } t\to+\infty, \tag{3.10}$$

since the interval $0\leqslant r\leqslant 1$ is easily handled with the Riemann–Lebesgue lemma and the dominated convergence theorem. Now $E_{k,v}(r)$ is a linear combination of entries of $P(r)R_k(r)$ [see (2.14)] with coefficients which are bounded functions of k. Hence, to prove (3.10) it suffices to prove that if $g(k)$ is a bounded function of k with compact support in $(0,\infty)$, then the entries of

$$\int_0^\infty dk\, g(k)e^{-itk^2}P(r)R_k(r)\to 0 \quad \text{in } L^2((1,\infty),dr)$$
$$\text{as } t\to+\infty. \tag{3.11}$$

Recalling (2.21) we see that it suffices to prove that the entries of

$$\int_1^\infty ds\, w(s)\chi_{[r,\infty)(s)}\int_0^\infty dk\, g(k)e^{-itk^2}e^{\pm ikr}$$
$$\times\prod_{r_0}^{s} e^{A_k(\xi)\,d\xi}\to 0 \quad \text{in } L^2((1,\infty),dr), \tag{3.12}$$

when $g(k)$ is a bounded function with compact support in $(0,\infty)$. Now the entries of $\Pi_{r_0}^{s}\exp[A_k(\xi)\,d\xi]$ are bounded functions of k uniformly in s for k bounded away from 0. Let $\alpha_k(s)$ denote any of these entries. Then

$$\left\|\int_1^\infty ds\, w(s)\chi_{[r,\infty)}(s)\int_0^\infty ds\, g(k)e^{-itk^2}e^{\pm ikr}\alpha_k(s)\right\|_{L^2((1,\infty),dr)}$$
$$\leqslant\int_1^\infty ds\,|w(s)|\,\|\chi_{[r,\infty)}(s)$$
$$\times\int_0^\infty dk\, g(k)e^{-itk^2}e^{\pm ikr}\alpha_k(s)\|_{L^2((1,\infty),dr)}. \tag{3.13}$$

Now $\|\chi_{[r,\infty)}(s)\int_0^\infty dk\, g(k)e^{-itk^2}e^{\pm ikr}\alpha_k(s)\|_{L^2((1,\infty),dr)}$ is bounded by a fixed constant independent of s and t by the Plancherel theorem, so by the dominated convergence theorem we need only show that this norm tends to zero as $t\to\infty$ for each fixed s. This follows from the more general fact:

Lemma: Suppose $g(k)\in L^2((-\infty,\infty),dk)$ and $f(\xi)$ is a bounded function on $-\infty<\xi<\infty$ which tends to zero as $\xi\to\pm\infty$. Then if we put

$$h_t(\xi) = f(\xi)\int_{-\infty}^\infty dk\, g(k)e^{-itk^2}e^{ik\xi}, \tag{3.14}$$

we have

$$\lim_{t\to\infty}\|h_t(\xi)\|_{L^2((-\infty,\infty),d\xi)} = 0. \tag{3.15}$$

Proof (essentially contained in Ref. 7): We have

$$h_t(\xi) = f(\xi)e^{-it\mathscr{H}_0}\check{g}, \tag{3.16}$$

where $\mathscr{H}_0$ is the self-adjoint operator

$$\mathscr{H}_0 = -d^2/d\xi^2 \quad \text{on } L^2((-\infty,\infty),d\xi), \tag{3.17}$$

and $\check{g}$ is the inverse Fourier transform of g. By the results of Ref. 7 (actually discussed for the three-dimensional case but applicable with essentially no change to the one-dimensional case), it suffices to show that

$$\lim_{t\to\infty}\|f(\xi)(C_t\check{g})(\xi)\|_{L^2((-\infty,\infty),d\xi)} = 0,$$

where

$$(C_t\check{g})(\xi) = (1/2it)^{1/2}e^{i\xi^2/4t}g(\xi/2t). \tag{3.18}$$

But then

$$\|f(\xi)(C_t\check{g})(\xi)\|_{L^2}^2 = \int d\xi\,|f(\xi)|^2 1/2t\,|g(\xi/2t)|^2$$
$$= \int d\eta\,|f(2t\eta)|^2|g(\eta)|^2\to 0 \quad \text{as } t\to\infty$$
$$(\eta = \xi/2t), \tag{3.19}$$

by the dominated convergence theorem.

We have now proved [see (3.6)] that $\omega^+ = \text{strong}\lim_{t\to+\infty} e^{ith}e^{-ith_0}$ exists and is determined by

$$\omega^{+}\varphi\,(l,0,f,r) = \varphi\,(l,v,e^{-i\delta_l}f,r). \tag{3.20}$$

An analogous calculation shows that ω^{-} exists and

$$\omega^{-}\varphi\,(l,0,f,r) = \varphi\,(l,v,e^{+i\delta_l}f,r). \tag{3.21}$$

Hence

$$\begin{aligned}\mathrm{range}(\omega^{+}) &= \mathrm{range}(\omega^{-})\\ &= \text{closure of the functions } \varphi(l,v,f,r)\\ &\quad \text{in } L^2((0,\infty),dr).\end{aligned} \tag{3.22}$$

The operator $s = (\omega^{+})^{*}(\omega^{-})$ is unitary and is determined by

$$s\varphi\,(l,0,f,r) = \varphi\,(l,0,e^{2i\delta_l}f,r). \tag{3.23}$$

Hence $\Omega^{\pm}$ exist and $S = (\Omega^{+})^{*}(\Omega^{-})$ is unitary.

[1]S.T. Kuroda, J. Math. Phys. **3**, 933 (1962).
[2]T.A. Green and O.E. Lanford, J. Math. Phys. **1**, 139 (1960).
[3]John D. Dollard and Charles N. Friedman, Ann. Phys. **3**, 251 (1978).
[4]John D. Dollard and Charles N. Friedman, J. Funct. Anal. **28**, 309 (1978).
[5]John D. Dollard and Charles N. Friedman, J. Diff. Eqns. **31**, 418 (1979).
[6]John D. Dollard and Charles N. Friedman, "Product integration," to appears as volume in *Encyclopedia of Applied Mathematics* (Addison-Wesley, Reading, Mass.).
[7]John D. Dollard, Rocky Mtn. J. Math. **1**, 5 (1971).

Jet bundles and path structures[a)]

R. A. Coleman

Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7
and Mathematics and Physics Department, Ryerson Polytechnical Institute, 50 Gould St., Toronto, Ontario, Canada M5B 1E8

H. Korte

Department of Philosophy, University of Western Ontario, London, Ontario, Canada
and Department of Philosophy, University of Regina, Regina, Saskatchewan, Canada

(Received 27 March 1979; accepted for publication 25 June 1979)

The analysis of path structures is formulated in terms of jet bundles with particular emphasis on the transformation laws and symmetry properties of geodesic path structures. The role played by geodesic path structures in the constructive axioms of Ehlers, Pirani, and Schild for GRT is discussed and it is shown that these axioms are decidable.

1. INTRODUCTION

Ehlers, Pirani, and Schild[1] (EPS) proposed a set of constructive axioms for general relativity theory based on the local behavior of arbitrary massive particles, freely falling massive particles, and light propagation. The analysis of the aspect of spacetime structure revealed by the paths followed by freely falling massive particles leads to the study of path structures on manifolds. Ehlers and Köhler[2] have presented an analysis of path structures and their symmetries using the standard formalism of the first and second order tangent bundles, $T(M)$ and $T(T(M))$, of the spacetime manifold M. However, the simplest and most natural description of structures of higher order contact is in terms of the jets and jet bundles of Ehresmann.[3]

In the present paper, the analysis of curve and path structures is developed using jets. A great simplification, both conceptual and technical, results. Conceptually, the elements of the second order jet bundle $J^2(\mathbb{R}_0, M)$ have a direct interpretation as second degree Taylor approximations to curves through a given point of M. The derivation of the coordinate, parameter, and active transformation laws is a straightforward exercise in the application of the chain law. In contrast, the elements of $T(T(M))$ which in jet language is $J^1(\mathbb{R}_0, J^1(\mathbb{R}_0, M))$, are more complicated. The desired elements of the sub-bundle $J^2(\mathbb{R}_0, M)$ of $J^1(\mathbb{R}_0, J^1(\mathbb{R}_0, M))$, must in the standard approach be selected by imposing the spray condition on the elements of $J^1(\mathbb{R}_0, J^1(\mathbb{R}_0, M))$. Of course, the interpretation of the elements of the subbundle and the discussion of their coordinate, parameter, and active transformation laws is considerably obscured by the use of this indirect approach. The definition and analysis of sprays, called acceleration fields in this paper, are also much simplified by using jets. Moreover, the relationship between these fields and second order differential equations becomes transparent. Finally, the discussion of the corresponding structures for paths is very difficult if the standard approach is used. The discussion in terms of jets is easy in comparison. In the case of geodesic acceleration fields and the projective analog, geodesic directing fields, the description in terms of jets is so much simpler that such fields are readily obtained as cross sections of appropriate fiber bundles. The bundle of geodesic directing fields, $\mathscr{G}\Xi(M)$, (4.22), provides an elegant coordinate free formulation[4] of the second projective axiom of EPS; namely, the directing field which governs the motion of freely falling particles is a cross section of $\mathscr{G}\Xi(M)$.

The definitions and notations for jets and jet bundles are established in Sec. 2. In Sec. 3, curve structures, acceleration fields, and the one to one relationship between them are discussed. Also, geodesic acceleration fields are defined and it is shown how these may be obtained as cross sections of a fiber bundle. The analogous discussion for path structures and directing fields is presented in Sec. 4.

The definitions of active transformations and symmetries of curve and path structures are given in Sec. 5. The discussion is presented for the three customary levels of analysis, global, local, and micro (infinitesimal neighborhood of a point p of M). The formulas for the microtransformations and microsymmetry conditions are particularly relevant for this paper and are presented in detail. These results are then used in Sec. 6 to prove some theorems concerning geodesic curve and path structures and their microsymmetry groups. Theorem 4 states that a curve structure is geodesic if and only if its microsymmetry group is isomorphic to $\mathrm{GL}^1(n)$. The maximal microsymmetry group of a geodesic path structure is derived in Theorem 5. In comparison with the standard treatment of this projective group, the jet bundle language offers a marked improvement in conceptual clarity. Theorems 6 and 7 correspond to Theorems 2 and 3 of Ehlers and Köhler.[2] The first of these theorems states that a path structure which admits a microsymmetry transformation at every point whose first order part is a dilatation other than the identity is geodesic. The proof given by Ehlers and Köhler is reproduced for completeness. The second theorem states that a path structure which is maximally isotropic to first order in the sense that it admits, at every point of the manifold, a microsymmetry group whose first order part acts transitively on the space of one-directions $\mathbb{D}^1_p(M)$ is geodesic and conversely. Ehlers and Köhler present the proof of this theorem only for analytic path structures and for manifold dimension $n = 2$. The proof presented below does not require analyticity (only C^6), and the organization of the proof is sufficiently improved so that it can be written down

[a)]Supported in part by the Natural Sciences and Engineering Research Council Canada.

 0022-2488/80/061340-12$1.00

in reasonably concise form for the case of arbitrary manifold dimension n.

The geodesic method of EPS has recently been criticized.[5-8] It has been argued that the geodesic method is beset with logical and derivatively with epistemological circularity. Specifically, criteria that determine which bodies are suitable as freely falling test bodies and permit their identification presuppose metrical considerations, thereby leading to circularity. A particle which has a gravitational multipole structure will not in general travel along a timelike geodesic even if no forces act on it. Without already knowing the spacetime structure, how are we to know which particles are gravitational monopoles and which are not?

In Sec. 7, these criticisms are briefly analyzed. It is shown that they rest on a serious misunderstanding of the nature of inertial laws and the geodesic method.

Morever, using radar coordinates and the concept of a directing field, it is shown that the criticisms are without any substance; that is, it is shown that the truth of the projective axioms concerning free fall motion is epistemically decidable in a noncircular way.

2. JETS AND JET BUNDLES

Let M and N be C^∞ differentiable manifolds of dimensions m and n, respectively. Let $(U,x)_p$ and $(V,y)_q$ be charts for neighborhoods $p\in M$ and $q\in N$. The k-jet $j^k_p(f)$ of a C^k map $f:M\rightarrow N$ with source $p\in M$ and target $q=f(p)\in N$ is the equivalence class of such maps which agree at the point $p\in M$ and for which the derivatives of the maps $y\circ f\circ x^{-1}$ agree at $x(p)$ up to and including order k. That the equivalence is not dependent on the choice of coordinate charts follows from the chain rule. The set of such k-jets is denoted by $J^k(M_p,N_q)$. If the source, target or both are unrestricted, the sets of k-jets are denoted by $J^k(M,N_q)$, $J^k(M_p, N)$, and $J^k(M,N)$, respectively. These four sets of k-jets are differentiable manifolds, and the coefficients of the kth order Taylor expansion of $y\circ f\circ x^{-1}$ may be used as local coordinates of the k-jet $j^k_p(f)$. Moreover, the source and target maps σ: $J^k(M,N)\rightarrow M$ and $\tau: J^k(M,N)\rightarrow N$ defined by

$$\sigma(j^k_p(f)) = p, \qquad \tau(j^k_p(f)) = f(p) \tag{2.1}$$

are differentiable.

If $m=n$, denote by $D(M_p,N_q)$ the set of diffeomorphisms $f:M\rightarrow N$ such that $f(p)=q$, and by $J^kD(M_p, N_q)$ the set of k-jets $j^k_p(f)$. The Lie group $\mathrm{GL}^k(n)$ is defined to be the set of k-jets $J^kD(R^n_0, R^n_0)$ with the group product defined by k-jet composition

$$j^k_0(f_1)\circ j^k_0(f_2) = j^k_0(f_1\circ f_2). \tag{2.2}$$

This group acts on $J^k(\mathbb{R}_0, \mathbb{R}^n_0)$, the set of k-jets of curves through $0\in\mathbb{R}^n$, according to

$$j^k_0(f)\circ j^k_0(\gamma) = j^k_0(f\circ\gamma), \tag{2.3}$$

where $\gamma:\mathbb{R}\rightarrow\mathbb{R}^n$ and $\gamma(0)=\mathbf{0}$.

A local grid for $p\in M$ is a diffeomorphism $\zeta:\mathbb{R}^n \rightarrow\zeta_{\vdash}(\mathbb{R}^n)\subset M$ such that $\zeta(0)=p$. A k-grid is a k-jet $j^k_0(\zeta)$ of a grid. Let $Z^k(M)$ denote the set of k-grids for all $p\in M$. $Z^k(M)$ is the total space of a principal fiber bundle (PFB)

$$\mathscr{Z}^k(M) = \langle Z^k(M), \pi_{Zk}, M, \mathrm{GL}^k(n)\rangle, \tag{2.4}$$

where the differentiable projection map $\pi_{Zk}: Z^k(M)\rightarrow M$ is defined by

$$\pi_{Zk}(j^k_0(\zeta)) = \zeta(0). \tag{2.5}$$

In view of the action (2.3) of $\mathrm{GL}^k(n)$ on $J^k(\mathbb{R}_0, \mathbb{R}^n_0)$, one may construct the associated fiber bundle (AFB)[9] of k-arcs on M,

$$\mathscr{A}^k(M) = \langle J^k(\mathbb{R}_0,M), \pi_k, M, J^k(\mathbb{R}_0,\mathbb{R}^n_0), \mathscr{Z}^k(M)\rangle, \tag{2.6}$$

with typical fiber $J^k(\mathbb{R}_0,\mathbb{R}^n_0)$. As the notation indicates, the elements of the total space $J^k(\mathbb{R}_0,M)$ may be more directly obtained as the k-jets of curves $\gamma:\mathbb{R}\rightarrow M$ in M. The projection map $\pi_k: J^k(\mathbb{R}_0,M)\rightarrow M$ is defined by

$$\pi_k(j^k_0(\gamma)) = \gamma(0). \tag{2.7}$$

There is a sequence of natural, differentiable projection maps $\pi^l_k: J^k(\mathbb{R}_0, M)\rightarrow J^l(\mathbb{R}_0, M)$ for $1\leqslant l<k$ defined by

$$\pi^l_k(j^k_0(\gamma)) = j^l_0(\gamma). \tag{2.8}$$

In many cases of physical interest, the parameter of a curve is either arbitrary or not specified in advance; for example, in general relativity the world line of a freely falling test particle is determined by a point on it and its direction (a nonzero multiple of its tangent vector) at that point. Since the tangent vectors of physical particles are everywhere nonzero, curves such as $\gamma:\mathbb{R}\rightarrow\mathbb{R}^2$ with

$$\gamma(t) = (t^2, t^3) \tag{2.9}$$

need not be considered in the definition of a path ("parameter free curve") for the purposes of this paper.

A parameter transformation is an element of $D(\mathbb{R},\mathbb{R})$. The k-jets $j^k_0(\mu)\in J^kD(\mathbb{R}_0,\mathbb{R}_0)$ for $\mu\in D(\mathbb{R}_0,\mathbb{R}_0)$ form a group P^k where the group product is k-jet composition.

Define an equivalence relation in the set of curves with nowhere vanishing tangent vectors by $\hat{\gamma}\sim\gamma$ iff $\exists\mu\in D(\mathbb{R}_0, \mathbb{R}_0)$, $\hat{\gamma}=\gamma\circ\mu$. Then a path is an equivalence class of such curves.

There is an action of the group P^k on $J^k(\mathbb{R}_0,\mathbb{R}^n_0)$ and on $J^k(\mathbb{R}_0, M)$ which will be denoted by R_k in both cases. It is given by

$$R_k(j^k_0(\gamma), j^k_0(\mu)) = j^k_0(\gamma)\circ j^k_0(\mu). \tag{2.10}$$

This right action is compatible with the structure of the bundle $\mathscr{A}^k(M)$; that is, $R_k\circ j^k(f) = j^k(f)\circ R_k$ for $f:M\rightarrow M$ and $\pi_k\circ R_k=\pi_k$ and $\pi^l_k\circ R_k = R_l\circ\pi^l_k$. Denote by $\mathbb{D}^k$ and $\mathbb{D}^k(M)$ the sets of equivalence classes of elements of $J^k(\mathbb{R}_0,\mathbb{R}^n_0)$ and $J^k(\mathbb{R}_0, M)$ defined by R_k. These equivalence classes will be called k-directions (or simply directions for $k=1$). Note that 2-directions are called special directions in Ehlers and Köhler.[2]

For $k>1$, the manifold of k-directions $\mathbb{D}^k(M)$ is the total space of an AFB with typical fiber $\mathbb{D}^k$ and PFB $\mathscr{Z}^k(M)$

$$\mathscr{D}^k(M) = \langle \mathbb{D}^k(M), \pi_k, M, \mathbb{D}^k, \mathscr{Z}^k(M)\rangle. \tag{2.11}$$

For $k=1$, the structure group of the bundle is $\mathrm{PG}(n)$, the projective group in n dimensions. $\mathrm{PG}(n)$ is the factor group of $\mathrm{GL}(n)$ with respect to the invariant subgroup of elements of the form $(\lambda\delta^i_j)$ with $\lambda\neq 0$ called dilatations. The appropriate PFB is the bundle of projective 1-grids.

$$\mathscr{P}\mathscr{Z}^1(M) = \langle PZ^1(M), \pi_{Z1}, M, \mathrm{PG}(n)\rangle , \qquad (2.12)$$

where the elements of the fiber $PZ^1_p(M)$ at $p \in M$ are equivalence classes of 1-grids in $Z^1_p(M)$ related by a dilatation. The AFB of 1-directions is then

$$\mathscr{D}^1(M) = \langle \mathrm{D}^1(M), \pi_1, M, \mathrm{D}^1, \mathscr{P}\mathscr{Z}^1(M)\rangle . \qquad (2.13)$$

The above considerations do not require C^∞ manifolds. A differentiability class C^r for some finite r would be sufficient. In the following, it is assumed that mappings are sufficiently differentiable that any derivative maps which occur are at least C^1. It is also assumed that the base manifold M has dimension $n \geqslant 2$.

3. CURVE STRUCTURES

Following Ehlers and Köhler,[2] we restrict the concept of a curve in M, $\gamma: I \to M$, where I is an open interval of $\mathbb{R}$ by requiring: For every $s_1, s_2 \in I$ such that $\gamma(s_1) = \gamma(s_2)$ and $\dot\gamma(s_1) = \dot\gamma(s_2)$, there exist open intervals $I_1 \ni s_1$ and $I_2 \ni s_2$ and a smooth invertible map $\mu: I_1 \to I_2$ such that $\mu(s_1) I s_2$ and $\gamma | I_1 = (\gamma \circ \mu) | I_1$.

A curve which retraces itself periodically such as $\gamma: \mathbb{R} \to \mathbb{R}^2$

$$\gamma(s) = (\cos s, \sin s) \qquad (3.1)$$

is not excluded, nor is a curve which touches itself or retraces a portion of its track in the opposite sense $[\dot\gamma(s_1) = -\dot\gamma(s_2)]$. However, a curve which touches itself or retraces part of its track in the same sense $[\dot\gamma(s_1) = \dot\gamma(s_2)]$ is excluded for the condition is not satisfied at the point (s) where the curve bifurcates. For such a curve, at the point of bifurcation, information of higher order [say $\ddot\gamma(s)$] would be required to determine which branch to follow; consequently, the curve could not satisfy everywhere a differential equation of second order. Note that the condition excludes bifurcation for both increasing and decreasing values of the parameter.

One may also consider curves $\gamma^1: I \to J^1(\mathbb{R}_0, M)$ and $\gamma^2: I \to J^2(\mathbb{R}_0, M)$ given in terms of local coordinates by

$$\begin{aligned} \gamma^1(s) &= (\gamma^{1i}(s), \gamma_1^{1i}(s)) , \\ \gamma^2(s) &= (\gamma^{2i}(s), \gamma_1^{2i}(s), \gamma_2^{2i}(s)) . \end{aligned} \qquad (3.2)$$

For those special curves for which

$$\begin{aligned} \gamma_1^{1i}(s) &= \dot\gamma^{1i}(s) , \\ \gamma_1^{2i}(s) &= \dot\gamma^{2i}(s) , \\ \gamma_2^{2i}(s) &= \dot\gamma_1^{2i}(s) , \end{aligned} \qquad (3.3)$$

the γ^1 and γ^2 are called the first and second lifts of the curves $\pi_1 \circ \gamma^1: \mathbb{R} \to M$ and $\pi_2 \circ \gamma^2: \mathbb{R} \to M$, respectively. If $\gamma = \pi_1 \circ \gamma^1$: and $\gamma = \pi_2 \circ \gamma^2$:, then one writes

$$\begin{aligned} j^1(\gamma) &= \gamma^1 , \\ j^2(\gamma) &= \gamma^2 . \end{aligned} \qquad (3.4)$$

The relations (3.3) do not hold in general since the coordinates γ_1^i, γ_2^i are defined as derivatives only at a point.

Definition: A curve structure (CS) $\mathscr{C}$, on M is a set of curves in M such that for every element $\gamma^1 \in J^1(\mathbb{R}_0, M)$, there exists exactly one maximal curve $\gamma \in \mathscr{C}$ such that $j^1(\gamma)$ passes through γ^1.

Definition: An acceleration field on M is a map $A: J^1(\mathbb{R}_0, M) \to J^2(\mathbb{R}_0, M)$ such that $\pi^1_2 \circ A = \mathrm{id}$.

Lemma: Every curve structure on M defines a unique acceleration field on M and conversely.

Given a curve structure $\mathscr{C}$ and $\gamma^1 \in J^1(\mathbb{R}_0, M)$, let $\gamma \in \mathscr{C}$ be the unique curve such that $j^1(\gamma)$ passes through γ^1. Then define

$$A(\gamma^1) = j^2_s(\gamma) , \qquad (3.5)$$

where $s \in I$ is uniquely defined by $j^1(\gamma)(s) = \gamma^1$. Because of the restriction on curves in M stated above, $j^1(\gamma)$ does not self-intersect.

Conversely, in a given coordinate system an acceleration field A is given by

$$\begin{aligned} A(\gamma^{1i}, \gamma_1^{1i}) &= (A^i(\gamma^{1i}, \gamma_1^{1i}), A^i_1(\gamma^{1i}, \gamma_1^{1i}), A^i_2(\gamma^{1i}, \gamma_1^{1i})) \\ &= ((\gamma^{1i}, \gamma_1^{1i}, A^i_2(\gamma^{1i}, \gamma_1^{1i})) . \end{aligned} \qquad (3.6)$$

Then, the initial conditions $\gamma^i(0) = \gamma^{1i}$ and $\dot\gamma^i(0) = \gamma_1^{1i}$ and the differential equation

$$\ddot\gamma^i(s) = A^i_2(\gamma^i(s), \dot\gamma^i(s)) \qquad (3.7)$$

determine a unique curve γ up to a translation in parameter space such that $j^1(\gamma)$ passes through $\gamma^1 \in J^1(\mathbb{R}_0, M)$.

Unless required for clarity, the superscript denoting the order of the jet and the coordinates of the base point will be suppressed. Let $(U, x)_p$ and $(\bar U, \bar x)_p$ be charts of $p \in M$. Set $X = x \circ \bar x^{-1}$ and $\bar X = \bar x \circ x^{-1}$. Then $X \circ \bar X = \mathrm{id}$ and $\bar X \circ X = \mathrm{id}$ with suitable domain restrictions. The coordinates of a 2-jet with respect to these charts, (γ_1^i, γ_2^i) and $(\bar\gamma_1^i, \bar\gamma_2^i)$, are related by

$$\begin{aligned} \bar\gamma_1^i &= \bar X^i_j \gamma_1^j , \\ \bar\gamma_2^i &= \bar X^i_j \gamma_2^j + \bar X^i_{jk} \gamma_1^j \gamma_1^k , \end{aligned} \qquad (3.8)$$

where $(\bar X^i_j, \bar X^i_{jk}) \in \mathrm{GL}^2(n)$ is the 2-jet of $\bar X$ at $x(p)$. The coordinates of a 1-jet, (γ_1^i) and $\bar\gamma_1^i$, are related by the first of equations (3.8).

Definition: A geodesic acceleration field Γ: $J^1(\mathbb{R}_0, M) \to J^2(\mathbb{R}_0, M)$ is an acceleration field for which, at each $p \in M$, there is a chart, say $(\bar U, \bar x)_p$, such that

$$\bar\Gamma(\bar\gamma_1^i) = (\bar\gamma_1^i, 0) . \qquad (3.9)$$

This definition is a modern formulation of Weyl's definition of a symmetric linear connection. (See Ref. 14b, Sec. 15, p. 114.)

Theorem 1: An acceleration field is geodesic iff relative to any given chart $(U, x)_p$

$$\Gamma^i_2(\gamma_1^j) = -\Gamma^i_{jk} \gamma_1^j \gamma_1^k , \qquad (3.10)$$

where the Γ^i_{jk} are functions only of $p \in M$.

If an acceleration field is geodesic, then relative to some chart $(\bar U, \bar x)_p$, it is given by Eq. (3.9). Then relative to $(U, x)_p$

$$\begin{aligned} \Gamma(\gamma_1^i) &= (\bar X^{-1})(\bar X)\Gamma(\bar X^{-1})(\bar X)(\gamma_1^i) \\ &= (X)\bar\Gamma(\bar\gamma_1^i) = (X)(\bar\gamma_1^i, 0) \\ &= (\gamma_1^i, X^i_{lm} \bar X^l_j \bar X^m_k \gamma_1^j \gamma_1^k) . \end{aligned} \qquad (3.11)$$

Thus (3.10) holds with $\Gamma^i_{jk} = -X^i_{lm} \bar X^l_j \bar X^m_k$ which are functions only of $p \in M$. Conversely, if an acceleration field is given by (3.10), then relative to $(\bar{U},\bar{x})_p$

$$\bar{\Gamma}^i_2(\bar{\gamma}^j_1) = (-\bar{X}^i_l\Gamma^l_{jk}X^j_pX^k_q + \bar{X}^i_{jk}X^j_pX^k_q)\bar{\gamma}^p_1\bar{\gamma}^q_1 . \quad (3.12)$$

Consequently, there exist charts in which $\bar{\Gamma}^i_2(\bar{\gamma}^j_2)$ vanishes; namely, those for which

$$\bar{X}^i_{jk} = \bar{X}^i_l\Gamma^l_{jk} . \quad (3.13)$$

Geodesic acclearation fields can all be obtained as cross sections of an AFB with PFB $\mathscr{L}^2(M)$. Consider the space of maps $\Gamma: J^1(\mathbb{R}_0,\mathbb{R}^n_0) \to J^2(\mathbb{R}_0,\mathbb{R}^n_0)$ such that $\pi^1_2\circ\Gamma = \text{id}$ and

$$\Gamma(\gamma^i_1) = (\gamma^i_1, -\Gamma^i_{jk}\gamma^j_1\gamma^k_1) , \quad (3.14)$$

where the Γ^i_{jk} are just numbers. This function space is a manifold of dimension $n^2(n+1)/2$ with the global coordinates Γ^i_{jk}. The group $GL^2(n)$ acts on this space according to

$$[(a)\Gamma(a^{-1})]^i_{jk} = (a^i_l\Gamma^l_{pq} - a^i_{pq})a_j^{-1p}a_k^{-1q} . \quad (3.15)$$

Note that this equation has nothing at all to do with the manifold M whereas (3.12) refers to a particular $p\in M$ and the Γ^i_{jk} in (3.12) are functions of p.

Denoted the space of maps defined by (3.14) by GA. Then using the $GL^2(n)$ action on GA given by (3.15) construct the AFB

$$\mathscr{G}\mathscr{A}(M) = \langle GA(M),\pi_{GA},M,GA,\mathscr{L}^2(M)\rangle . \quad (3.16)$$

Then every geodesic acceleration field on M is given by a cross section $\Gamma: M \to GA(M)$ of $GA(M)$.

4. PATH STRUCTURES

A path in M will be denoted by ξ. That a curve γ is a member of the equivalence class defining ξ will be denoted by $\gamma\in\xi$, or $\xi = [\gamma]$. The k-lift of a curve $\gamma: I \to M$ is the curve $j^k(\gamma): I \to J^k(\mathbb{R}_0, M)$ which defines a curve $j^k_{R_k}(\gamma)$: $I \to \mathbb{D}^k(M)$ by means of the right action R_k of P^k on $J^k(\mathbb{R}_0, M)$. The k-lift of the path ξ in M is the path $j^k(\xi) \equiv [j^k_{R_k}(\gamma)]$ in $\mathbb{D}^k(M)$. (Note that it is not appropriate to define $j^k(\xi)$ to be the path $[j^k(\gamma)]$ since the set of parameter transformations allowed for $[j^k(\gamma)]$ is in general the subset of those for $[\gamma]$ such that $j^k_s(\mu)$ is the identity of P^k.) A general element of $\mathbb{D}^k(M)$ will be denoted by ξ^k. General curves and paths in $\mathbb{D}^k(M)$ may be defined but will not be needed for the purposes of this paper.

Relative to a coordinate chart $(U,x)_p$ for $p\in M$, $\gamma^1 \in J^1(\mathbb{R}_0,M_p)$ and $\gamma^2 \in J^2(\mathbb{R}_0,M_p)$ are determined by the coordinates

$$\gamma^1 = (x^i(p),\gamma^{1i}_1) ,$$
$$\gamma^2 = (x^i(p),\gamma^{2i}_1,\gamma^{2i}_2) . \quad (4.1)$$

In terms of these coordinates, the right actions R_1 and R_2 defined by (2.10) are given by

$$R_1(\gamma^1,j^1_0(\mu)) = (x^i(p),(D\mu)\gamma^{1i}_1) ,$$
$$R_2(\gamma^2,j^2_0(\mu)) = (x^i(p),(D\mu)\gamma^{2i}_1,(D\mu)^2\gamma^{2i}_2 + (D^2\mu)\gamma^{2i}_1) , \quad (4.2)$$

where $D\mu$ and $D^2\mu$ are the first and second derivatives of the parameter transformation at $t = 0$. From (4.2), it is evident that the portion of $\mathbb{D}^1(M)$ over U is covered by the n coordinate charts defined by taking $D\mu = 1/\gamma^{1b}_1$ for $b = 1,...,n$. Similarly, the portion of $\mathbb{D}^2(M)$ over U is covered by the n coordinate charts defined by taking $D\mu = 1/\gamma^{2b}_1$ and

$$D^2\mu = -\gamma^{2b}_2/(\gamma^{2b}_1)^3 \quad (4.3)$$

for $b = 1,...,n$. In general, equations will only be written for the case $b = n$, a case which is particularly apt for discussing timelike paths when $n = 4$. In this case the parameter transformed coordinates are given by

$$\xi^{1i}_1 = \gamma^{1i}_1/\gamma^{1n}_1 ,$$
$$\xi^{2i}_1 = \gamma^{2i}_1/\gamma^{2n}_1 \quad \xi^{2i}_2 = (\gamma^{2n}_1\gamma^{2i}_2 - \gamma^{2i}_1\gamma^{2n}_2)/(\gamma^{2n}_1)^3 \quad (4.4)$$

and satisfy $\xi^{1n}_1 = 1$, $\xi^{2n}_1 = 1$, $\xi^{2n}_2 = 0$. In terms of local coordinates, elements $\xi^1\in\mathbb{D}^1_p(M)$ and $\xi^2\in\mathbb{D}^2_p(M)$ are given by

$$\xi^1 = (x^i(p), \xi^{1\alpha}_1) ,$$
$$\xi^2 = (x^i(p), \xi^{2\alpha}_1, \xi^{2\alpha}_2) , \quad (4.5)$$

where $\alpha = 1,...,n-1$. For convenience, the superscript denoting the order of the element and the coordinates of $p\in M$ will in general be suppressed.

Let $\gamma: I \to M$ be a curve in M. Then the lifted curves in $\mathbb{D}^1(M)$ and $\mathbb{D}^2(M)$ are given by

$$j^1_{R_1}(\gamma)(s) = (x^i\circ\gamma(s),\xi^\alpha_1(s)) ,$$
$$j^2_{R_2}(\gamma)(s) = (x^i\circ\gamma(s),\xi^\alpha_1(s),\xi^\alpha_2(s)) , \quad (4.6)$$

where

$$\xi^\alpha_1(s) = \dot{\gamma}^\alpha(s)/\dot{\gamma}^n(s) ,$$
$$\xi^\alpha_2(s) = [\dot{\gamma}^n(s)\ddot{\gamma}^\alpha(s) - \dot{\gamma}^\alpha(s)\ddot{\gamma}^n(s)]/[\dot{\gamma}^n(s)]^3 . \quad (4.7)$$

Writing $x^i(s) = x^i\circ\gamma(s) = \gamma^i(s)$, one readily obtains

$$\xi^\alpha_1(s) = \frac{dx^\alpha}{dx^n} ,$$
$$\xi^\alpha_2(s) = \frac{d^2x^\alpha}{(dx^n)^2} . \quad (4.8)$$

For the case $n = 4$, $\xi^\alpha_1(s)$ and $\xi^\alpha_2(s)$ are the 3-velocity and 3-acceleration, respectively.

Definition: A path structure (PS), $\mathscr{P}$, on M is set of paths in M such that for every element $\xi^1\in\mathbb{D}^1(M)$, there exists exactly one maximal path $\xi\in\mathscr{P}$ such that ξ^1 is on $j^1(\xi)$.

Definition: A directing field on M is a map Ξ: $\mathbb{D}^1(M) \to \mathbb{D}^2(M)$ such that $\pi^1_2\circ\Xi = \text{id}$.

In terms of local coordinates, a directing field Ξ is given by

$$\Xi(x^i(p), \xi^\alpha_1)$$
$$= (\Xi^i(x^i(p),\xi^\alpha_1),\Xi^\alpha_1(x^i(p),\xi^\alpha_1),\Xi^\alpha_2(x^i(p),\xi^\alpha_1))$$
$$= (x^i(p),\xi^\alpha_1,\Xi^\alpha_2(x^i(p),\xi^\alpha_1)) . \quad (4.9)$$

Lemma: Every PS on M defines a unique directing field on M and conversely.

Given a PS, choose $\xi^1\in\mathbb{D}^1(M)$ and let $\xi = [\gamma]$ be the unique path determined by $\xi^1 = (x^i(p),\xi^\alpha_1)$. Let $j^2_{R_2}(\gamma)(s) = (x^i\circ\gamma(s),\xi^\alpha_1(s),\xi^\alpha_2(s))$ and let ξ^α_2 be the value of $\xi^\alpha_2(s)$ at p. Then Ξ is defined at ξ^1 by

$$\Xi^\alpha_2(x^i(p),\xi^\alpha_1) = \xi^\alpha_2 . \quad (4.10)$$

Conversely, a directing field Ξ determines a PS by means of the differential equation

$$\xi^\alpha_2(s) = \Xi^\alpha_2(x^i(p),\xi^\alpha_1(s)), \quad (4.11)$$

which by means of (4.8) may be reexpressed as

$$\frac{d^2x^\alpha}{(dx^n)^2} = \Xi^\alpha_2\left(x^i,\frac{dx^\alpha}{dx^n}\right). \tag{4.12}$$

The coordinate transformation formulas (3.8) together with (4.4) yield the transformation formulas

$$\bar\xi^\alpha_1 = \frac{\bar X^\alpha_n + \bar X^\alpha_\beta \xi^\beta_1}{\bar X^n_n + \bar X^n_\gamma \xi^\gamma_1},$$

$$\bar\xi^\alpha_2 = \frac{\bar X^\alpha_\beta \xi^\beta_2 + \bar X^\alpha_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2\bar X^\alpha_{n\rho}\xi^\rho_1 + \bar X^\alpha_{nn}}{(\bar X^n_n + \bar X^n_\gamma \xi^\gamma_1)^2} - \frac{\bar X^n_\beta \xi^\beta_2 + \bar X^n_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2\bar X^n_{n\rho}\xi^\rho_1 + \bar X^n_{nn}}{(\bar X^n_n + \bar X^n_\gamma \xi^\gamma_1)^2}\bar\xi^\alpha_1. \tag{4.13}$$

Definition: A geodesic directing field $\Pi:\mathbb{D}^1(M) \to \mathbb{D}^2(M)$ is a directing field for which, at each $p\in M$, there is a chart, say $(\bar U,\bar x)_p$, such that

$$\bar\Pi(\bar\xi^\alpha_1) = (\bar\xi^\alpha_1, 0). \tag{4.14}$$

Note that every geodesic directing field corresponds to a class of symmetric linear connections which are projectively equivalent. (See Ref. 12, Sec. 22, p. 56.)

Theorem 2: A directing field is geodesic iff relative to any given chart $(U,x)_p$

$$\Pi^\alpha_2(\xi^\beta_1) = \xi^\alpha_1(\Pi^n_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2\Pi^n_{n\rho}\xi^\rho_1 + \Pi^n_{nn}) - (\Pi^\alpha_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2\Pi^\alpha_{n\rho}\xi^\rho_1 + \Pi^\alpha_{nn}), \tag{4.15}$$

where the Π^i_{jk} are functions only of $p\in M$ and $\Pi^i_{ji} = 0$ [so that $\Pi^n_{n\rho}$ and Π^n_{nn} can be eliminated from (4.15)].

Let Π be a geodesic directing field satisfying (4.14). Then relative to the chart $(U,x)_p$

$$\Pi(\xi^\alpha_1) = (X)(\bar X)\Pi(X)(\bar X)(\xi^\alpha_1) = (X)\bar\Pi(\bar\xi^\alpha_1) = (X)(\bar\xi^\alpha_1, 0). \tag{4.16}$$

Using the inverse of (4.13), one obtains

$$\Pi^\alpha_2(\xi^\alpha_1) = \frac{X^\alpha_{\rho\sigma}\bar\xi^\rho_1\bar\xi^\sigma_1 + 2X^\alpha_{n\rho}\bar\xi^\rho_1 + X^\alpha_{nn}}{(X^n_n + X^n_\gamma\bar\xi^\gamma_1)^2} - \frac{X^n_{\rho\sigma}\bar\xi^\rho_1\bar\xi^\sigma_1 + 2X^n_{n\rho}\bar\xi^\rho_1 + X^n_{nn}}{(X^n_n + X^n_\gamma\bar\xi^\gamma_1)^2}\xi^\alpha_1. \tag{4.17}$$

Substitution for $\bar\xi^\alpha_1$ in terms of ξ^α_1 gives

$$\Pi^\alpha_2(\xi^\alpha_1) = -\xi^\alpha_1(X^n_{ij}\bar X^i_\rho\bar X^j_\sigma\xi^\rho_1\xi^\sigma_1 + 2X^n_{ij}\bar X^i_n\bar X^j_\rho\xi^\rho_1 + X^n_{ij}\bar X^i_n\bar X^j_n) + (X^\alpha_{ij}\bar X^i_\rho\bar X^j_\sigma\xi^\rho_1\xi^\sigma_1 + 2X^\alpha_{ij}\bar X^i_n\bar X^j_\rho\xi^\rho_1 + X^\alpha_{ij}\bar X^i_n\bar X^j_n). \tag{4.18}$$

Since $\Gamma^i_{ji} = -X^i_{pq}\bar X^p_j\bar X^q_l$ by (3.11), (4.18) is the same as (4.15) with the Π^i_{jk} replaced by Γ^i_{jk}. However, one can define $\Gamma_i \equiv \Gamma^k_{ki}$ and

$$\Pi^i_{jk} = \Gamma^i_{jk} - [1/(n+1)](\delta^i_j\Gamma_k + \delta^i_k\Gamma_j), \tag{4.19}$$

where $\Pi^i_{ij} = 0$. The terms in (4.18) involving Γ_i cancel, giving (4.15).

Conversely, suppose a directing field is given by (4.15). Then apply (4.13) (for simplicity choose $\bar X^\alpha_\beta = \delta^\alpha_\beta$) to obtain

$$\bar\Pi^\alpha_2(\bar\xi^\beta_1) = \Pi^\alpha_2(\xi^\beta_1) + \bar X^\alpha_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2\bar X^\alpha_{n\rho}\xi^\rho_1 + \bar X^\alpha_{nn} - \xi^\alpha_1(\Pi^n_2(\xi^\beta_1) + \bar X^n_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + \bar X^n_{n\rho}\xi^\rho_1 + \bar X^n_{nn}). \tag{4.20}$$

The right-hand side vanishes for the choice $\bar X^i_{jk} = \Pi^i_{jk}$; so that, a coordinate chart exists in which (4.14) holds.

Geodesic directing fields can all be obtained as cross sections of an AFB with PFB $\mathscr{L}^2(M)$. With apologies for the multiple use of the same symbols, consider the space of maps $\Pi:\mathbb{D}^1(\mathbb{R}^n_0) \to \mathbb{D}^2(\mathbb{R}^n_0)$ such that $\pi^1_2\circ\Pi = \text{id}$ and with $\Pi^\alpha_2(\xi^\beta_1)$ given by the expression (4.15) with the understanding that ξ^α_1 denotes an element of $\mathbb{D}^1(\mathbb{R}^n_0)$ [not of $\mathbb{D}^1(M_p)$] and that the Π^i_{jk} are just numbers (not functions of $p\in M$). This function space, denoted by $G\Xi$, is a manifold of dimension $n^2(n+1)/2 - 4$ (since $\Pi^i_{ij} = 0$). Again, there are n coordinate charts. Corresponding to the chart in which $\xi^b_1 = 1$ and $\xi^b_2 = 0$, one may choose to eliminate Π^b_{bb} and $\Pi^b_{b\rho}$. An element $(a)\in GL^2(n)$ acts on $G\Xi$ according to

$$\Pi \to (a)\Pi(a)^{-1}. \tag{4.21}$$

The effect of this transformation of the Π^i_{jk} can be found by successive application of (4.13) with the $(\bar X^i_j, \bar X^i_{jk})$ replaced by (a^i_j, a^i_{jk}). Thus one can construct the AFB

$$\mathscr{G}\Xi(M) = \langle G\Xi(M), \pi_{G\Xi}, M, G\Xi, \mathscr{L}^2(M)\rangle \tag{4.22}$$

and every geodesic directing field on M is given by a cross section $\Pi:M \to G\Xi(M)$ of $G\Xi(M)$.

Finally, it is clear from Theorem 2 that if Π is a geodesic directing field, then $\Pi^\alpha_2(\xi^\beta_1)$ is a cubic polynomial in ξ^β_1 in every coordinate chart $(U,x)_p$. The converse is also true.

Theorem 3: If with respect to every coordinate chart $(U,x)_p$, the corresponding function $\Xi^\alpha_2(\xi^\beta_1)$ which determines the directing field Ξ is cubic, that is, if

$$\Xi^\alpha_2(\xi^\beta_1) = A^\alpha + B^\alpha_\rho\xi^\rho_1 + C^\alpha_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + D^\alpha_{\rho\sigma\tau}\xi^\rho_1\xi^\sigma_1\xi^\tau_1, \tag{4.23}$$

where the coefficients A, B, C, D are functions only of $p\in M$, then Ξ is geodesic.

Under a coordinate transformation, a directing field tranforms according to

$$\bar\Xi = (\bar X)\Xi(X). \tag{4.24}$$

In terms of the function $\Xi^\alpha_2(\xi^\beta_1)$, this law becomes

$$\bar\Xi^\alpha_2(\bar\xi^\beta_1) = \overline{\Xi^\alpha_2(\xi^\beta_1)} = \frac{\bar X^\alpha_\rho\Xi^\rho_2(\xi^\beta_1) + \bar X^\alpha_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2\bar X^\alpha_{n\rho}\xi^\rho_1 + \bar X^\alpha_{nn}}{(\bar X^n_n + \bar X^n_\gamma\xi^\gamma_1)^2} - \frac{\bar X^n_\rho\Xi^\rho_2(\xi^\beta_1) + \bar X^n_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2\bar X^n_{n\rho}\xi^\rho_1 + \bar X^n_{nn}}{(\bar X^n_n + \bar X^n_\gamma\xi^\gamma_1)^2}\bar\xi^\alpha_1, \tag{4.25}$$

where $\overline{\Xi^\alpha(\xi^\beta_1)}$ is given by the second equation of (4.13). The expression for $\bar\Xi^\alpha_2(\bar\xi^\beta_1)$ is obtained by substituting (4.23) into (4.25) and by expressing ξ^α_1 in terms of $\bar\xi^\alpha_1$ using the inverse of the first of Eqs. (4.13). The result is not in general a polynomial unless the coefficients $D^\alpha_{\rho\sigma\tau}$ have the form

$$D^\alpha_{\rho\sigma\tau} = \tfrac{1}{3}(\delta^\alpha_\rho D_{\sigma\tau} + \delta^\alpha_\sigma D_{\rho\tau} + \delta^\alpha_\tau D_{\rho\sigma}). \tag{4.26}$$

However, if this condition is satisfied, then (4.23) may be put into the form (4.15) by redefining the coefficients in the following way. Set

$$B^{\alpha}_{\rho} = 2\tilde{B}^{\alpha}_{\rho} + \delta^{\alpha}_{\rho} B ,$$
$$B = [1/(n+1)] B^{\alpha}_{\alpha} , \tag{4.27}$$

and

$$C^{\alpha}_{\rho\sigma} = \tilde{C}^{\alpha}_{\rho\sigma} + \delta^{\alpha}_{\rho} C_{\sigma} + \delta^{\alpha}_{\sigma} C_{\rho} ,$$
$$C_{\rho} = [1/(n+1)] C^{\alpha}_{\alpha\rho} . \tag{4.28}$$

Then it is only necessary to make the identifications

$$D_{\rho\sigma} = \Pi^{n}_{\rho\sigma}, \quad \tilde{C}^{\alpha}_{\rho\sigma} = -\Pi^{\alpha}_{\rho\sigma} ,$$
$$\tilde{B}^{\alpha}_{\rho} = -\Pi^{\alpha}_{n\rho}, \quad A^{\alpha} = -\Pi^{\alpha}_{nn} , \tag{4.29}$$

from which follow (recall $\Pi^{i}_{ij} = 0$)

$$B = \tilde{B}^{\alpha}_{\alpha} = -\Pi^{\alpha}_{\alpha n} = \Pi^{n}_{nn} ,$$
$$C_{\rho} = \tilde{C}^{\alpha}_{\alpha\rho} = -\Pi^{\alpha}_{\alpha\rho} = \Pi^{n}_{n\rho} . \tag{4.30}$$

5. SYMMETRIES OF CURVE AND PATH STRUCTURES

For the examination of differentiable manifolds and for the discussion of the symmetries of geometric objects defined on them, there are three qualitatively different scales to consider; namely, global, local, and micro. In each case, a symmetry is an invertible, active transformation of the manifold which preserves the geometric object when attention is restricted to the appropriate scale. For a given scale, the set of transformations which preserve a given geometric object form a group or pseudogroup called its global, local, or microsymmetry group, respectively. Note that the use of the term "infinitesimal symmetry group" instead of "microsymmetry group" would *incorrectly* suggest that the Lie algebra of some finite group was under consideration. The symmetry groups will be defined for the cases of curve and path structures and for the corresponding acceleration and directing fields, the geometric objects of central interest in this paper; however, similar definitions would apply to any geometric object.[10]

First, consider global symmetries of a CS $\mathscr{C}$. Let $f:M \to M$ be a diffeomorphism. Then for every $\gamma \in \mathscr{C}$, $\gamma^{f} = f \circ \gamma$ is a curve in M and $\mathscr{C}^{f} = \{\gamma^{f} | \gamma \in \mathscr{C}\}$ is a CS for M. If $\mathscr{C}^{f} = \mathscr{C}$, then f is a symmetry of $\mathscr{C}$ and the set of all diffeomorphisms $f:M \to M$ such that $\mathscr{C}^{f} = \mathscr{C}$ is the global symmetry group of $\mathscr{C}$.

Moreover, if $\mathscr{P}$ is a PS on M and $\xi = [\gamma]$ is a path, $\xi \in \mathscr{P}$, then $\xi^{f} = [f \circ \gamma]$ is a path on M and $\mathscr{P}^{f} = \{\xi^{f} | \xi \in \mathscr{P}\}$ is a PS for M. If $\mathscr{P}^{f} = \mathscr{P}$, then f is a symmetry of $\mathscr{P}$ and the set of all diffeomorphisms f such that $\mathscr{P}^{f} = \mathscr{P}$, is the global symmetry group of $\mathscr{P}$.

Because of the bijective correspondence between CS's and acceleration fields and between PS's and directing fields, the above definitions may be reformulated in terms of these fields. Let $A:J^{1}(\mathbb{R}_0, M) \to J^{2}(\mathbb{R}_0, M)$ be the acceleration field corresponding to the CS $\mathscr{C}$. Then the acceleration field A^{f} corresponding to the CS, $\mathscr{C}^{f}$ is given by

$$A^{f} = j^{2}(f) \circ A \circ j^{1}(f)^{-1} , \tag{5.1}$$

where $j^{k}(f):J^{k}(\mathbb{R}_0, M) \to J^{k}(\mathbb{R}_0, M)$ is the k-prolongation of $f:M \to M$. The condition that the CS remain invariant under f is

$$A^{f} = A . \tag{5.2}$$

If $\gamma:I \to M$ is a curve on M and μ is a parameter transformation, then since

$$j^{k}(f) \circ j^{k}_{0}(\gamma \circ \mu) = j^{k}(f \circ \gamma) \circ j^{k}_{0}(\mu) , \tag{5.3}$$

This action of the $j^{k}(f)$ can be factored by the projective transformations $j^{k}(\mu)$ to define the action on $\mathbb{D}^{k}(M)$

$$j^{k}(f) \circ j^{k}_{R_k}(\gamma) = j^{k}_{R_k}(f \circ \gamma) . \tag{5.4}$$

Consequently, if $\Xi:\mathbb{D}^{1}(M) \to \mathbb{D}^{2}(M)$ is the directing field corresponding to the PS, $\mathscr{P}$ then the directing field Ξ^{f} corresponding to the PS $\mathscr{P}^{f}$ is given by

$$\Xi^{f} = j^{2}(f) \circ \Xi \circ j^{1}(f)^{-1} \tag{5.5}$$

and the condition for invariance of the PS, $\mathscr{P}$, becomes

$$\Xi^{f} = \Xi . \tag{5.6}$$

If the global diffeomorphism is replaced by a local diffeomorphism $f:U \to V$ in the above considerations and if the invariance conditions are applied to the restrictions of curves and paths to U and V, then one refers to the local diffeomorphism f as a local symmetry and the set of such local symmetries forms a local symmetry pseudogroup. If, in addition, the local diffeomorphisms are required to leave some point $p \in M$ fixed, the terms p-local symmetry and p-local symmetry pseudogroup will be used. In this case, the invariance conditions are applied only to those curves and paths which pass through the point p.

The set $J^{k}D(M_p, M_p)$ of k-jets $j^{k}_{p}(f)$ of diffeomorphisms $f:M \to M$ which leave $p \in M$ fixed form a finite dimensional Lie group GL^{k}_{p}. The group product is k-jet composition. The group GL^{k}_{p} is isomorphic to the group $\mathrm{GL}^{k}(n)$. For $l < k$, there is a natural projection from GL^{k}_{p} to GL^{l}_{p} which maps $j^{k}_{p}(f)$ into $j^{l}_{p}(f)$. The group GL^{k}_{p} acts on $J^{k}(\mathbb{R}_0, M_p)$ according to

$$j^{k}_{p}(f) \circ j^{k}_{0}(\gamma) = j^{k}_{0}(f \circ \gamma) . \tag{5.7}$$

Again, parameter transformations commute with the action (5.7) so that the group GL^{k}_{p} also acts on $\mathbb{D}^{k}_{p}(M)$ according to

$$j^{k}_{p}(f) \circ j^{k}_{R_k}(\gamma) = j^{k}_{R_k}(f \circ \gamma) . \tag{5.8}$$

[See (4.6) and (4.7).]

As noted above, a diffemorphism f induces transformations (5.1) and (5.5) of acceleration and directing fields, respectively. If $f(p) = p$, then one may restrict these transformations to the point $p \in M$ to obtain

$$A^{f}_{p} = j^{2}_{p}(f) \circ A_{p} \circ j^{1}_{p}(f)^{-1} ,$$
$$\Xi^{f}_{p} = j^{2}_{p}(f) \circ \Xi_{p} \circ j^{1}_{p}(f)^{-1} , \tag{5.9}$$

called the microtransformations at p of the curve and path structures.

Definition: A microsymmetry of CS, $\mathscr{C}$ (or a PS, $\mathscr{P}$) at a point $p \in M$ is an element of GL^2_p which leaves the corresponding acceleration field A (or directing field Ξ) invariant at p. The set of such microsymmetries forms a group which is a Lie subgroup of GL^2_p called the microsymmetry group at p.

The invariance conditions are

$$A^f_p = A_p, \quad \Xi^f_p = \Xi_p. \tag{5.10}$$

Relative to a chart $(U,x)_p$, the microtransformation $j^2_p(f)$ is represented by

$$j^2_{x(p)}(x \circ f \circ x^{-1}) = (f^i_{\ j}, f^i_{\ jk}), \tag{5.11}$$

where $f^i_{\ j}$ is the Jacobian and $f^i_{\ jk}$ is the Hessian at $x(p)$. For $(\gamma^i_1) \in J^1(\mathbb{R}_0, M_p)$,

$$\begin{aligned} A^f(\gamma^i_1) &= (f)A(f)^{-1}(\gamma^i_1) \\ &= (f)A(f_j^{-1i}\gamma^j_1) = (f)(f_j^{-1i}\gamma^j_1, A^i_2(f_j^{-1i}\gamma^j_1)) \\ &= (\gamma^i_1, f^i_{\ j}A^j_2(f_j^{-1i}\gamma^j_1) + f^i_{\ jk} f_l^{-1j} f_m^{-1k}\gamma^l_1\gamma^m_1). \end{aligned} \tag{5.12}$$

Consequently, the transformation law is

$$A^{fi}_2(\gamma^i_1) = f^i_{\ j}A^j_2(f_j^{-1i}\gamma^j_1) + f^i_{\ jk} f_l^{-1j} f_m^{-1k}\gamma^l_1\gamma^m_1. \tag{5.13}$$

Replacement of γ^i_1 by $f^i_{\ j}\gamma^j_1$ gives for the invariance condition

$$A^i_2(f^i_{\ j}\gamma^j_1) = f^i_{\ j}A^j_2(\gamma^i_1) + f^i_{\ jk}\gamma^j_1\gamma^k_1. \tag{5.14}$$

For an infinitesimal microtransformation

$$(f) = (\delta^i_j + \epsilon F^i_{\ j}, \epsilon F^i_{\ jk}), \tag{5.15}$$

where ϵ is infinitesimal; consequently, the infinitesimal version of (5.14) is

$$F^k_{\ j}\gamma^j(\partial/\partial\gamma^k_1)A^i_2(\gamma^i_1) = F^i_{\ j}A^j_2(\gamma^i_1) + f^i_{jk}\gamma^j_1\gamma^k_1. \tag{5.16}$$

The corresponding formulas for directing fields are obtained as follows. Choose one of the n coordinate charts for $\mathbb{D}^1_p(M)$ and $\mathbb{D}^2_p(M)$ corresponding to $(U,x)_p$, say the nth. Then apply (5.9) in the form

$$\Xi^f(f) = (f)\Xi. \tag{5.17}$$

Using (4.13), one obtains for $\xi_1 \in \mathbb{D}^1_p(M)$

$$\Xi^f(f)(\xi^\alpha_1) = \Xi^f\left(\frac{f^\alpha_n + f^\alpha_\beta\xi^\beta_1}{f^n_n + f^n_\gamma\xi^\gamma_1}\right) = \left(\frac{f^\alpha_n + f^\alpha_\beta\xi^\beta_1}{f^n_n + f^n_\gamma\xi^\gamma_1}, \Xi^{f\alpha}\left(\frac{f^\alpha_n + f^\alpha_\beta\xi^\beta_1}{f^n_n + f^n_\gamma\xi^\gamma_1}\right)\right) \tag{5.18}$$

and

$$\begin{aligned} (f)\Xi(\xi^\alpha_1) &= (f)(\xi^\alpha_1, \Xi^\alpha_2(\xi^\alpha_1)) = \Bigg(\frac{f^\alpha_n + f^\alpha_\beta\xi^\beta_1}{f^n_n + f^n_\gamma\xi^\gamma_1}, \\ &\frac{f^\alpha_\beta\Xi^\beta_2(\xi^\alpha_1) + f^\alpha_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2f^\alpha_n\xi^\rho_1 + f^\alpha_{nn}}{(f^n_n + f^n_\gamma\xi^\gamma_1)^2} \\ &- \frac{f^n_\beta\Xi^\beta_2(\xi^\alpha_1) + f^n_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2f^n_{n\rho}\xi^\rho_1 + f^n_{nn}}{(f^n_n + f^n_\gamma\xi^\gamma_1)^2} \\ &\times\left(\frac{f^\alpha_n + f^\alpha_\beta\xi^\beta_1}{f^n_n + f^n_\gamma\xi^\gamma_1}\right)\Bigg). \end{aligned} \tag{5.19}$$

Using the convention $\xi^n_1 = 1$, the result may be expressed more compactly as

$$\begin{aligned} \Xi^{f\alpha}_2\left(\frac{f^\alpha_i\xi^i_1}{f^n_i\xi^i_1}\right) &= \frac{f^\alpha_\beta\Xi^\beta_2(\xi^\alpha_1)f^n_i\xi^i_1 - f^n_\beta\Xi^\beta_2(\xi^\alpha_1)f^\alpha_i\xi^i_1}{(f^n_i\xi^i_1)^3} \\ &+ \frac{f^\alpha_{jk}\xi^j_1\xi^k_1 f^n_i\xi^i_1 - f^n_{jk}\xi^j_1\xi^k_1 f^\alpha_i\xi^i_1}{(f^n_i\xi^i_1)^3}. \end{aligned} \tag{5.20}$$

The invariance condition corresponding to (5.14) is obtained by replacing Ξ^f by Ξ in (5.20). Finally, using (5.15), one obtains for the infinitesimal version of the invariance condition

$$\begin{aligned} &\frac{\partial\Xi^\alpha_2}{\partial\xi^\beta_1}(\xi^\alpha_1)[F^\beta_\gamma\xi^\gamma_1 + F^\beta_n - \xi^\beta_1(F^n_\gamma\xi^\gamma_1 + F^n_n)] \\ &\quad + 2\Xi^\alpha_2(\xi^\alpha_1)[F^n_\gamma\xi^\gamma_1 + F^n_n] \\ &\quad + \Xi^\beta_2(\xi^\alpha_1)[F^n_\beta\xi^\alpha_1 - F^\alpha_\beta] \\ &\qquad = F^\alpha_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2F^\alpha_{n\rho}\xi^\rho_1 + F^\alpha_{nn} \\ &\qquad\quad - \xi^\alpha_1[F^n_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2F^n_{n\rho}\xi^\rho_1 + F^n_{nn}]. \end{aligned} \tag{5.21}$$

6. SYMMETRIES OF GEODESIC CURVE AND PATH STRUCTURES

In this section, a number of theorems are stated and proved which serve to characterize geodesic curve and path structures geometrically in terms of their microsymmetry groups.

Theorem 4: A curve structure $\mathscr{C}$ is geodesic if and only if its microsymmetry group for every $p \in M$ is a subgroup of GL^2_p isomorphic to $GL^1(n)$.

Let A be the acceleration field corresponding to a geodesic CS, $\mathscr{C}$. Then with respect to any chart $(U,x)_p$

$$A^i_2(\gamma_1) = -\Gamma^i_{\ jk}\gamma^j_1\gamma^k_1. \tag{6.1}$$

Substitution of (6.1) into the invariance condition (5.14) gives

$$f^i_{\ jk} = f^i_l\Gamma^l_{\ jk} - \Gamma^i_{\ lm}f^l_{\ j}f^m_k. \tag{6.2}$$

Thus the microsymmetry group is the subgroup of GL^2_p of elements of the form

$$(f^i_{\ j}, f^i_l\Gamma^l_{\ jk} - \Gamma^i_{\ lm}f^l_{\ j}f^m_k). \tag{6.3}$$

It is straightforward to verify that this subgroup of GL^2_p is isomorphic to $GL^1(n)$.

Conversely, assume that the microsymmetry group is isomorphic to $GL^1(n)$. An infinitesimal element of GL^2_p has the form (5.15). For any element in the microsymmetry group, the $F^i_{\ jk}$ are determined by the $F^i_{\ j}$. The product of two such elements is

$$(\delta^i_j + \epsilon(F^i_{\ j} + G^i_{\ j}), \epsilon(\alpha^i_{\ jk}(F^r_s) + \alpha^i_{\ jk}(G^r_s)). \tag{6.4}$$

Closure requires linearity

$$\alpha^i_{\ jk}(F^r_s + G^r_s) = \alpha^i_{\ jk}(F^r_s) + \alpha^i_{\ jk}(G^r_s) \tag{6.5}$$

and $\alpha^i_{jk}(F^r_s)$ vanish when F^r_s vanish since the identity element is $(\delta^i_j, 0)$. Thus

$$\alpha^i_{jk}(F^r_s) = \alpha^i_{jkr}{}^s F^r_s , \quad (6.6)$$

where the $\alpha^i_{jkr}{}^s$ depend only on the point $p \in M$.

Now assume that $A^i_2(\gamma^i_1)$ is at least C^4 and set

$$A^i_2(\gamma^i_1) = A^i + B^i_j \gamma^j_1 + C^i_{jk}\gamma^j_1\gamma^k_1 + w^i(\gamma^i_1) , \quad (6.7)$$

where A^i, B^i_j, and C^i_{jk} depend only on $p \in M$ and $w^i(\gamma^i_1)$ is of order $(\gamma^i_1)^3$. Substitute (6.7) and (6.6) into (5.1) and note that the F^i_j are arbitrary. Set the coefficients of F^i_j equal to zero to obtain

$$\gamma^s[B^i_r + 2C^i_{jr}\gamma^j_1 + w^i_{,r}(\gamma^i_1)] = \delta^i_r[A^s + B^s_j\gamma^j_1 + C^s_{jk}\gamma^j_1\gamma^k_1 + w^s(\gamma^i_1)] + \alpha^i_{jkr}{}^s\gamma^j_1\gamma^k_1 . \quad (6.8)$$

Equating the coefficients of terms of corresponding order, one obtains

$$A^s = 0 ,$$
$$\delta^s_j B^i_r = -\delta^i_r B^s_j ,$$
$$\alpha^i_{jkr}{}^s = -\delta^i_r C^s_{jk} + \delta^s_j C^i_{rk} + \delta^s_k C^i_{jr} ,$$
$$\gamma^s_1 w^i_{,r}(\gamma^i_1) = \delta^i_r w^s(\gamma^i_1) . \quad (6.9)$$

From the last equation of (6.9), $w^i_{,r} = 0$ for $i \neq r$; so that, $\forall i$ the ith component of w depends only on the ith component of γ_1. But then choosing $r = i$ and $s \neq i$, w^s would have to be of first order in γ^s_1 contrary to assumption; consequently, $w^i(\gamma_1) = 0$. From the second equation of (6.9) by contracting on s and j

$$B^i_r = \delta^i_r(1/n)B^s_s = \delta^i_r B . \quad (6.10)$$

The third equation of (6.9) shows that the $\alpha^i_{jkr}{}^s$ have the form required in order that $(\delta^i_j + \epsilon F^i_j, \epsilon\alpha^i_{jkr}{}^s F^r_s)$ is a microsymmetry group element which is the infinitesimal version of (6.3) where

$$\alpha^i_{jk} = C^i_{jk} = -\Gamma^i_{jk} . \quad (6.11)$$

Using these results in (6.9), one obtains

$$A^i_2(\gamma^i_1) = B\gamma^i_1 - \Gamma^i_{jk}\gamma^j_1\gamma^k_1 . \quad (6.12)$$

The CS defined by (6.12) is geodesic since the term containing B can be eliminated by a suitable choice of parameter.

The fact that the microsymmetry group of a geodesic CS is isomorphic to $GL^1(n)$ is closely related to the existence of affine normal coordinates[11] and the fact that such coordinates are unique up to a $GL^1(n)$ transformation.

The next theorem characterizes the maximal microsymmetry group of a geodesic path structure $\mathscr{P}$ with corresponding directing field Ξ.

Theorem 5: If a path structure $\mathscr{P}$ is geodesic then its microsymmetry group for every $p \in M$ is a subgroup of GL^2_p isomorphic to the subgroup of $GL^2(n)$ with elements of the form

$$(a^i_j, a^i_j a_k + a^i_k a_j) . \quad (6.13)$$

The proof of this theorem is tedious but straightforward. Consider an arbitrary infinitesimal element (5.15) of GL^2_p. To be an element of the microsymmetry group $\mathscr{P}$, the parameters F^i_j and F^i_{jk} must satisfy the invariance condition (5.21) for arbitrary ξ^α_1 where the $\Xi^\alpha_2(\xi^\alpha_1)$ are given by the expression (4.15) for $\Pi^\alpha_2(\xi^\alpha_1)$ in terms of the Π^i_{jk} which are known and depend only on $p \in M$. Carrying out the substitutions, one obtains a polynomial in ξ^α_1 which must vanish for arbitrary ξ^α_1. Equating the coefficients of this polynomial to zero yields the various components of the relation

$$F^i_{jk} = F^i_l\Pi^l_{jk} - F^l_j\Pi^i_{lk} - F^l_k\Pi^i_{jl} + \delta^i_j F_k + \delta^i_k F_j . \quad (6.14)$$

Consequently, the F^i_{jk} are determined in terms of the parameters F^i_j and F_i which may be chosen arbitrarily. To organize the computation for (6.14), it is useful to define

$$F_i = [1/(n+1)]F^l_{li} ,$$
$$\hat{F}^i_{jk} = F^i_{jk} - (\delta^i_j F_k + \delta^i_k F_j), \quad (6.15)$$

so that $\hat{F}^l_{li} = 0$. After substituting for F^i_{jk} in terms of $\hat{F}^i_{jk}$ and F_i, the terms containing F_i drop out and the $\hat{F}^i_{jk}$ are determined by the first part of (6.14) involving the Π^i_{jk}. It is also useful to recall that $\Pi^l_{li} = 0$.

The finite form of the microtransformation (5.15) with F^i_{jk} given by (6.14) is

$$(f^i_j, f^i_l\Pi^l_{jk} - \Pi^i_{lm}f^l_j f^m_k + f^i_j f_k + f^i_k f_j) . \quad (6.16)$$

It is a straightforward matter to verify that the subgroup of GL^2_p of such elements is isomorphic to the projective subgroup of $GL^2(n)$ with elements given by (6.13).

Corresponding to the normal coordinates of a space with a geodesic curve structure, there are for a space with a geodesic path structure special projective normal coordinates[12] determined up to a projective transformation.

Consider the action of $GL^1(n)$ on a flat n-dimensional affine space. Straight lines through the origin are mapped into straight lines through the origin. Moreover, the dilatation subgroup of $GL^1(n)$ of elements $(e^s\delta^i_j)$ for $s \in \mathbb{R}$ (one might also include reflections) maps each straight line through the origin into itself. The following theorem states that if the paths of a path structure are straight to second order at every point $p \in M$, then the path structure is geodesic.

Theorem 6: If a PS, $\mathscr{P}$, admits at every $p \in M$ a microsymmetry $j^2_p(f) \in GL^2_p$ with $j^1_p(f) = (\lambda\delta^i_j)$ and $\lambda \neq 1$, then $\mathscr{P}$ is geodesic and conversely.

The converse follows from Theorem 5. Let $j^2_p(f)$ be a microsymmetry of $\mathscr{P}$ and let the corresponding directing field be Ξ. The invariance condition is given by (5.20) with $\Xi^{f\alpha}_2 = \Xi^\alpha_2$. Since $f^i_j = \lambda\delta^i_j$ with $\lambda \neq 0$ and $\lambda \neq 1$,

$$\Xi^\alpha_2(\xi^\alpha_1) = [1/(\lambda^2 - \lambda)][f^\alpha_{jk}\xi^j_1\xi^k_1 - \xi^\alpha_1 f^n_{jk}\xi^j_1\xi^k_1], \quad (6.17)$$

which is of the required form (4.15).

The following theorem states that if a path structure $\mathscr{P}$ is microisotropic to first order, then it is geodesic.

Theorem 7: If a PS, $\mathscr{P}$, admits at every $p \in M$ a microsymmetry group $G_p(\mathscr{P})$, a subgroup of GL^2_p, which induces a transitive action on $\mathbb{D}^1_p(M)$, then $\mathscr{P}$ is geodesic and conversely.

Again, the converse follows easily from Theorem 5. Suppose, then, that $G_p(\mathscr{P})$ induces a transitive action on

$\mathbb{D}_p^1(M)$. An arbitrary infinitesimal element of GL_p^2 is given by (5.15). For every such element that is an element of the microsymmetry group $G_p(\mathscr{P})$, the directing field Ξ of $\mathscr{P}$ satisfies the constraints (5.21). Since, $G_p(\mathscr{P})$ acts transitively on $\mathbb{D}_p^1(M)$, an n-dimensional projective space, and since dilatations do not affect points of $\mathbb{D}_p^1(M)$, the $F^i{}_j$ in (5.21) may be chosen arbitrarily up to a dilatation. In particular, if

$$F^i{}_j = \chi^i{}_j + (1/n)\delta^i_j\,\chi\,, \qquad (6.18)$$
$$\chi^i_i = 0, \quad \chi = F^i{}_i.$$

Then the χ^n_α,χ^α_n and χ^α_β may be chosen arbitrarily. Now, assume that $\Xi^\alpha_2(\xi_1)$ is at least C^6 and expand in a Taylor series about $\xi^\alpha_1 = 0$,

$$\Xi^\alpha_2(\xi_1) = A^\alpha + B^\alpha_\rho\xi^\rho_1 + C^\alpha_{\rho\pi}\xi^\rho_1\xi^\pi_1 + D^\alpha_{\rho\pi\sigma}\xi^\rho_1\xi^\pi_1\xi^\sigma_1 + E^\alpha_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1 + w^\alpha(\xi_1)\,, \qquad (6.19)$$

where it is assumed that $w^\alpha(\xi_1)$ is of order five in the variables $\xi^\alpha_1(\alpha = 1,...,n-1)$. Substitute (6.19) into (5.21) and pick out the terms of order at least four in ξ^α_1. Expressed in terms of the χ^n_α,$\chi^\alpha_n\chi^\alpha_\beta$, and χ, the result, which does not depend on the $F^i{}_{jk}$, is

$$\begin{aligned}
&\chi^\beta_n[w^\alpha_{,\beta}(\xi_1)] + \chi^n_\beta[-\xi^\beta_1\xi^\gamma_1 w^\alpha_{,\gamma}(\xi_1) + 2\xi^\beta_1 w^\alpha(\xi_1)\\
&+\xi^\alpha_1 w^\beta(\xi_1) - 2\xi^\beta_1 E^\alpha_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1 + \xi^\alpha_1 E^\beta_{\rho\pi\sigma\tau}\\
&\times\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1 + \xi^\alpha_1 D^\beta_{\rho\pi\sigma}\xi^\rho_1\xi^\pi_1\xi^\sigma_1 - \xi^\beta_1 D^\alpha_{\rho\pi\sigma}\xi^\rho_1\xi^\pi_1\xi^\sigma_1]\\
&+\chi^\beta_\gamma[-\xi^\gamma_1 w^\alpha_{,\beta}(\xi_1) + \delta^\gamma_\beta\xi^\eta_1 w^\alpha_{,\eta}(\xi_1)\\
&-\delta^\alpha_\beta w^\gamma(\xi_1) - \delta^\gamma_\beta w^\alpha(\xi_1) + 4\xi^\gamma_1 E^\alpha_{\rho\pi\sigma\beta}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\\
&-\delta^\alpha_\beta E^\gamma_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1 + 2\delta^\gamma_\beta E^\alpha_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1]\\
&+(1/n)\chi[-2\xi^\gamma_1 w^\alpha_{,\gamma}(\xi_1) + E^\alpha_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1] = 0\,.
\end{aligned} \qquad (6.20)$$

Since χ^β_n,χ^n_β, and χ^β_γ may be chosen arbitrarily, for arbitrary parameters λ_β,μ^β, and ν^γ_β which depend only on $p\in M$, one obtains the relations

$$w^\alpha_{,\beta}(\xi_1) = \lambda_\beta[-2\xi^\gamma_1 w^\alpha_{,\gamma}(\xi_1) + E^\alpha_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1]\,, \qquad (6.21)$$

$$\begin{aligned}
&-\xi^\beta_1\xi^\gamma_1 w^\alpha_{,\gamma}(\xi_1) + 2\xi^\beta_1 w^\alpha(\xi_1) + \xi^\alpha_1 w^\beta(\xi_1)\\
&-2\xi^\beta_1 E^\alpha_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1 + \xi^\alpha_1 E^\beta_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1\\
&+\xi^\alpha_1 D^\beta_{\rho\pi\sigma}\xi^\rho_1\xi^\pi_1\xi^\sigma_1 - \xi^\beta_1 D^\alpha_{\rho\pi\sigma}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\\
&\quad = \mu^\beta[-2\xi^\gamma_1 w^\alpha_{,\gamma}(\xi_1) + E^\alpha_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1]\,,
\end{aligned} \qquad (6.22)$$

and

$$\begin{aligned}
&-\xi^\gamma_1 w^\alpha_{,\beta}(\xi_1) + \delta^\gamma_\beta\xi^\eta_1 w^\alpha_{,\eta}(\xi_1) - \delta^\alpha_\beta w^\gamma(\xi_1) - \delta^\gamma_\beta w^\alpha(\xi_1)\\
&\quad +4\xi^\gamma_1 E^\alpha_{\rho\pi\sigma\beta}\xi^\rho_1\xi^\pi_1\xi^\sigma_1 - \delta^\alpha_\beta E^\gamma_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1\\
&\quad +2\delta^\gamma_\beta E^\alpha_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1\\
&\quad = \nu^\gamma_\beta[-2\xi^\eta_1 w^\alpha_{,\eta}(\xi_1) + E^\alpha_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1]\,.
\end{aligned} \qquad (6.23)$$

The terms of order four of (6.22) give

$$\xi^\alpha_1 D^\beta_{\rho\pi\sigma}\xi^\rho_1\xi^\pi_1\xi^\sigma_1 - \xi^\beta_1 D^\alpha_{\rho\pi\sigma}\xi^\rho_1\xi^\pi_1\xi^\sigma_1 = \mu^\beta E^\alpha_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1\,. \qquad (6.24)$$

Thus

$$\mu^\alpha E^\beta_{\rho\pi\sigma\tau} + \mu^\beta E^\alpha_{\rho\pi\sigma\tau} = 0\,. \qquad (6.25)$$

Suppose $\exists\alpha\; E^\alpha_{\rho\pi\sigma\tau}\neq 0$. Then (6.25) for $\beta=\alpha$ gives $\mu^\alpha = 0$ and then (6.25) for $\beta\neq\alpha$ gives $\mu^\beta = 0$. On the other hand, suppose $\exists\alpha\;\mu^\alpha\neq 0$. Then (6.25) for $\beta=\alpha$ gives $E^\alpha_{\rho\pi\sigma\tau} = 0$ and (6.25) for $\beta\neq\alpha$ gives $E^\beta_{\rho\pi\sigma\tau} = 0$.

Consequently, the right side of (6.24) must vanish, and it follows that

$$D^\alpha_{\rho\pi\sigma} = \tfrac{1}{3}\,[\delta^\alpha_\rho D_{\pi\sigma} + \delta^\alpha_\pi D_{\rho\sigma} + \delta^\alpha_\sigma D_{\rho\pi}]\,. \qquad (6.26)$$

Moreover, one must have $\forall_\alpha E^\alpha_{\rho\pi\sigma\tau} = 0$, for if $\exists\delta$ $E^\delta_{\rho\pi\sigma\tau}\neq 0$ then $\forall_\alpha\;\mu^\alpha = 0$ and the right side of (6.22) vanishes. The terms of order five of (6.22) then give for any α, β

$$-2\xi^\beta_1 E^\alpha_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1 + \xi^\alpha_1 E^\beta_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1 = 0 \qquad (6.27)$$

and for $\alpha = \beta = \delta$ this gives

$$\xi^\delta_1 E^\delta_{\rho\pi\sigma\tau}\xi^\rho_1\xi^\pi_1\xi^\sigma_1\xi^\tau_1 = 0 \qquad (6.28)$$

whence $E^\delta_{\rho\pi\sigma\tau} = 0$ which contradicts the assumption. Thus $\forall_\alpha E^\alpha_{\rho\pi\sigma\tau} = 0$.

Next (6.21) gives

$$w^\alpha_{,\beta}(\xi_1) = -2\lambda_\beta\xi^\gamma_1 w^\alpha_{,\gamma}(\xi_1)\,. \qquad (6.29)$$

Contraction with ξ^β_1 gives

$$(1 + 2\xi^\beta_1\lambda_\beta)\xi^\gamma_1 w^\alpha_{,\gamma}(\xi_1) = 0\,. \qquad (6.30)$$

Consequently, since $w^\alpha_{,\beta}(\xi_1)$ is C^1,

$$\xi^\gamma_1 w^\alpha_{,\gamma}(\xi_1) = 0\,, \quad w^\alpha_{,\beta}(\xi_1) = 0\,. \qquad (6.31)$$

Finally, (6.23) gives

$$\delta^\alpha_\beta w^\gamma(\xi_1) + \delta^\gamma_\beta w^\alpha(\xi_1) = 0\,, \qquad (6.32)$$

and by contraction of γ and β

$$w^\alpha(\xi_1) = 0\,. \qquad (6.33)$$

Since it has been shown that $\Xi^\alpha_2(\xi_1)$ is a polynomial of degree at most three, the proof may be completed by appealing to Theorem 3 above in Sec. 4. Alternatively, since (6.26) is just (4.26), the redefinition argument following (4.26) may be applied directly.

7. DECIDABILITY OF THE CONSTRUCTIVE AXIOMS OF GRT

Recent criticisms of the geodesic method of EPS were outlined in the Introduction. Before proving their invalidity, we shall briefly analyze their philosophical basis and contrast the latter with the conceptual motivation, significance, and aim of the constructive axiomatics of EPS. This will clarify to what extent the work of EPS constitutes a solution to the controversy between realism and geometric conventionalism in favor of realism.

Einstein suggested the distinction between *principle* theories and *constructive* theories.[13] The aim of a constructive theory is to reduce a wide class of diverse complex physical processes to simpler ones. Our understanding of the former is constructed out of hypotheses concerning the latter; for example, the kinetic theory of gases constructs mechanical, thermal, and diffusional processes from the hypothesis of molecular motion. On the other hand, a principal theory postulates abstract structural constraints which events are held to satisfy. Einstein's example is the classical theory of thermodynamics.

The special and general theories of relativity are principle theories of spacetime structure. The four dimensional pseudo-Riemannian manifold is the mathematical model of the physical spacetime of the theory of general relativity. It was Weyl who first distinguished between two more primitive structures of the model: the *conformal* structure, and the *projective* structure of paths defined by the set of all unparametrized geodesics.[14]

Weyl suggested that the conformal structure represents the *causal* structure and may be identified with the propagation of light, and that the projective structure represents the *inertial* structure of spacetime that is revealed by the *path* structure of free fall motions of suitable test particles.

Using these structures and their compatibility relation, Ehlers, Pirani, and Schild[1] have derived a unique Riemannian spacetime metric solely as a consequence of a set of "geometry free" axioms concerning the incidence and differential-topological properties of light propagation and free fall.

The "geometry free" axioms are propositions about a few general qualitative assumptions concerning free fall motion and light propagation that can be verified directly through experience in a way that does not presuppose the full blown edifice of the theory of general relativity. From these axioms, the theoretical basis of the theory is reconstructed step by step. Following Reichenbach,[15] EPS call their approach *constructive axiomatics*.

The aim of a constructive axiomatic approach to a principle theory of space–time is to exhibit the physical basis for the particular structural constraints which the principle theory postulates certain events must satisfy. The structures contained in the mathematical model of a principle theory should all have in principle a link to physical experience. Spacetime models with inherent structures that do not relate to experience (e.g., absolute time) are defective for that reason.[16] Hence, it must be theoretically possible, that is, possible in principle, to relate the various structures to experience in a way that is consistent with the theory.

Hence a constructive axiomatic approach should satisfy the basic requirement of any *proper* and *complete* theory. *Completeness* requires that the reconstruction of the various structures inherent in the mathematical model of a principle theory of spacetime be realizable by means of relatively simple physical systems that are themselves well defined within the specific theory being considered, that is, that can be considered as an interpretation of the inherent structures of the spacetime model and are consistent with the theoretical consequences of the theory which presupposes that model. Einstein was well aware of this problem and considered the use of clocks and rigid rods an undesirable makeshift.[17] Unlike light propagation and freely falling particles, rigid rods and ideal clocks are relativistically ill defined and are thus unsuitable for the determination of the inherent structures of the spacetime of general relativity. The concepts of a theory, its formulation and measuring devices should all lead to a unified, self-sufficient and conceptually coherent world picture. There are essentially two types of conventionalist viewpoints. The less radical type may be called *epistemological conventionalism*. On this view, observationally indistinguishable theories may utilize alternative geometries, but it is in principle not possible to single out that theory whose underlying geometry is the true geometry of the world. Any such decision, whether or not it is guided by criteria of simplicity, is essentially epistemically conventional. Epistemological conventionality permits the existence of a true geometry, but access to it is not possible in a nonconventional manner.

Ontological conventionalism asserts that the continuous spacetime manifold is metrically amorphous. All nontopological structures are *extrinsic* to spacetime and are stipulated by means of the behavior of material entities such as clocks, light rays, and geodesic particles; that is, the metric structure of spacetime is always relative to which class of material entities is chosen as the standard of measurement (which choice is arbitrary). According to this view; metrical relations within spacetime reduce to the relations of the chosen material standards of measurement; that is, the latter are ontologically *constitutive* of the former.

We are now able to see what the criticisms leveled against EPS really amount to. The charge of epistemic circularity is directed against the geodesic method because the latter employs the concept of free fall as a standard of inertial motion. The criticism is thus essentially about the status of the infinitesimal law of inertia. Since, as the argument goes, the inertial law does not by itself furnish *independent* criteria by which one can decide when a test particle is free, it is considered to be conventional in character. But this reasoning rests on a serious misunderstanding of both the law of inertia and the geodesic method which employs it.

First, the essential idea of the geodesic method is to *discover* through the behavior of physical systems various intrinsic, primitive geometrical spacetime structures. It is in spirit analogous to Helmholtz's procedure of deducing the existence and form of the metric of physical space.[18] Helmholtz asked "what must the geometric structure of space be in order that a mechanics of rigid bodies is realizable in that space?" Thus Helmholtz is essentially asking what abstract structural constrainst must a principle theory of mechanics postulate that certain events must satisfy. According to Helmholtz, the structure of space follows from the possibility of congruent transport of rigid bodies; that is, the structure of space constitutes a necessary condition for the possibility of the realizability of certain physical processes and operations within that space; in particular, whether or not space possesses a constant curvature, or whether space is a general Riemannian space depends on whether or not physics allows the introduction of ideal rigid bodies.

The structure of space is, according to Helmholtz, the framework for possible physical laws. Certain types of laws presuppose certain types of spaces. Hence, on this view, the law of inertia presupposes an affine structure and may thus be regarded as a geometrical statement.

The conventionalist view that considers the behavior of material entities as being ontologically constitutive of the metrical structure of spacetime is clearly at variance with the notion of a principle theory. It is clear that the views of Weyl and Helmholtz are directly opposed to those of ontological conventionalism. According to Weyl, "... the behavior of rigid bodies and clocks is almost exclusively determined through the metric structure, as is the pattern of the motion of a force free mass point and the propagation of a light source. And only through these effects on the concrete natural processes can we recognize this structure."[19] Thus according to Weyl we *discover* through the behavior of physical phenomena an *already determined* metrical structure of spacetime; that is, the metrical relations of physical objects are determined by the second rank *physical* metric tensor field which is only revealed by, not defined by, those relations. Although distinct from physical objects in space–time, the metric tensor *explains* the geometric relations between them.

Secondly, Newton's first law and the corresponding infinitesimal version thereof, is physically realized by a suitable class of objects in free motion. These laws are geometrical statements concerning the underlying spacetime structure. The inertial laws serve to define an affine structure on the spacetime manifold. It is the affine structure that plays the essential role in the formulation of all physical laws that are expressed in terms of differential equations. In both Newtonian physics and general relativity, all dynamical laws presuppose that structure. Now, inability to identify or single out a class of suitable test objects in an epistemologically noncircular way whose free motion exhibit the projective structure of spacetime means only that the truth of the axioms concerning free fall is epistemically undecidable. But any argument from the epistemic inaccessibility of free test particles—even if this inaccessibility has a sound logical and physical basis—does not establish that the structures derived from the axioms are ontologically conventional. The most that is entailed is epistemological conventionality.

However, epistemological conventionality permits the assertion of the truth of the axioms and hence the inference from them to a unique metric structure at least in this conditional sense:

> If the geometry-free axioms are true of the world and are hence satisfied by an actual or possible nonempty class of suitable test objects (light rays and symmetric, nonrotating, neutral, freely falling particles), then there exists a unique and intrinsic spacetime metric.

The truth of this conditional claim is *incompatible* with the truth of ontological conventionalism, for if the latter were true, then there could be *no factual* reasons, known or unknown, for preferring one metric over another. But EPS have at least shown that certain facts, *if known*, would single out a unique intrinsic metric. That we may not perhaps avail ourselves of these facts in an epistemically noncircular way supports only epistemological conventionalism.

We shall now show that one does have epistemic access to freely falling particles in a way that does not beset the geodesic method of EPS with either logical or epistemological circularity. First, note that freely falling particles are *not* required to construct the radar coordinate systems. For this purpose, any massive particles may be employed. Then, relative to such a coordinate system the trajectory of any other particle may be determined.

If the motion of a particle is governed by a directing field Ξ, then, by definition, such a particle's spacetime trajectory is determined uniquely by an event on the trajectory and its direction at that event. Assume that there are many particles governed by a given directing field Ξ if there are any at all. Then collections of particles corresponding to various directing fields can be built up by means of the following comparison procedure. Two particles belong to the same directing field class if and only if whenever they are launched from infinitesimally neighboring spacetime events with directions which differ only infinitesimally, their subsequent spacetime trajectories remain infinitesimally near. Here, the notion of near does not require a metric. Only an appeal to the differentiable structure of the manifold is required. The fact that in practice such a differentiable topological concept of nearness would require limiting sequences of experiments[20] would only complicate the matching procedure. Note that requiring the directions to differ only infinitesimally does not presuppose a connection since the infinitesimal transformation has been left arbitrary. This matching procedure permits the separation of particles into classes, each class associated with a distinct directing field. The EPS axiom regarding the existence of freely falling particles asserts the existence of at least one such class.

Particles with higher order gravitational multipole moments can almost be eliminated from consideration at this point. One would expect that their spacetime trajectories would not be uniquely determined solely by an event on the trajectory and the direction at the event but would also depend on the orientation of the multipole moment as is the case for particles with higher electromagnetic multipole moments. The motion of such particles would not be governed by a directing field and the above matching procedure would fail. The analyses of the motion of particles with gravitational multipole moments,[21] both relativistic and nonrelativistic, indicate that the motion of such particles is indeed not governed by directing fields; however, it is not possible to rely on such analyses here because they presuppose a metric. Consequently, the conceivable degenerate case in which only the scalar magnitude of such a particle's multipole moment interacts with the gravitational field must be considered.

For each class of particles, the corresponding directing field Ξ could be measured at any given spacetime event as follows. Take a large number of the particles and launch them from many different directions in such a way that they all pass through an infinitesimal neighborhood of the given spacetime event. Track each of the particles in some radar coordinate system. Then by curve fitting and differentiation (4.8), the one and two directions (ξ_1, ξ_2) for each of the particles may be determined at the given event. These pairs in turn determine the directing field

$$\xi_2^\alpha = \Xi_2^\alpha(\xi_1^\beta) \tag{7.1}$$

at the event in the given coordinate system. By repeating the procedure for many spacetime events the directing *fields* for the given class of particles may be measured.

Having measured the directing field with sufficient accuracy at a large number of spactime points, the analytic criterion (4.15) of Theorem 2 may be used to determine whether or not it is geodesic. Assume a polynomial form for the functions Ξ_2^α in (7.1) of degree greater that three, say five or six. Then use the measured data pairs (ξ_1, ξ_2) at the given spacetime event to determine the coefficients by, for example, the method of least squares. Then if the coefficients of the terms of degree greater than three are essentially zero and if the third degree terms other than $\xi_1^\alpha D_{\rho\sigma} \xi_1^\rho \xi_1^\sigma$ are also essentially zero, and if this turned out to be the case for every spacetime event considered, then one would conclude that the directing field was geodesic. If it turned out that Ξ_2^α were not cubic polynomials of the desired form even at a single spacetime event, then one would conclude that the directing field was not geodesic. This curve fitting technique also serves to determine the projective coefficients Π^i_{jk} ($\Pi^i_{ik} = 0$) as functions of the spacetime event. In turn, these coefficients uniquely determine a geodesic path structure.

The determination and measurement of the conformal tensor density $\mathscr{G}_{ab}$ and the conformal connection coefficients

$$K^i_{jk} = \tfrac{1}{2}\mathscr{G}^{il}(\mathscr{G}_{lj,k} + \mathscr{G}_{lk,j} - \mathscr{G}_{jk,l}) \qquad (7.2)$$

is adequately discussed elsewhere in the literature. Ehlers, Pirani, and Schild have shown that the necessary and sufficient condition that a geodesic path structure determined by Π^i_{jk} is compatible with the conformal structure determined by $\mathscr{G}_{ab}$ is that[1]

$$\Delta^i_{jk} \equiv \Pi^i_{jk} - K^i_{jk} = 5\mathscr{G}_{jk}\mathscr{G}^{il}q_l - \delta^i_j q_k - \delta^i_k q_j\,, \qquad (7.3)$$

where the coefficients q_i depend only on the spacetime event. The Eqs. (7.3) form a system of $n^2(n+1)/2$ linear equations in the n unknowns q_i. The structures are compatible if and only if a solution exists for every spacetime event. If (7.3) holds then the q_i are given by

$$q_i = \tfrac{1}{18}\mathscr{G}_{il}\mathscr{G}^{pq}\Delta^l_{pq} \qquad (7.4)$$

(for four dimensional spacetime); so that, the compatibility conditions that must be satisfied by the Δ^i_{jk} may be obtained by substituting (7.4) into the right-hand side of (7.3). If the structures are compatible, the unique symmetric linear connection which preserves nullity of vectors is given by

$$\Gamma^i_{jk} = K^i_{jk} + 5\mathscr{G}^{il}(\mathscr{G}_{jk}q_l - \mathscr{G}_{lj}q_k - \mathscr{G}_{lk}q_j)$$
$$= \Pi^i_{jk} - 4(\delta^i_j q_k + \delta^i_k q_k)\,. \qquad (7.5)$$

It is clear from this relation that it is possible to have any number of distinct projective structures all compatible with the same conformal structure.

If extensive investigation failed to reveal even a single class of particles governed by a geodesic directing field, then the EPS construction would fail to demonstrate the existence of a unique Riemannian metric. Such a structure might still exist, but other means would have to be sought to establish evidence for its existence.

If one or more classes of particles governed by geodesic directing fields were found and if none of the projective structures were compatible with the conformal structure, the construction would fail as before. If two or more projective structures were found which were compatible with the conformal structure, then not even a unique Weyl structure would exist let alone a unique Riemannian structure. There remains the case in which exactly one class of particles governed by a geodesic directing field compatible with the conformal structure is found. Then the projective path structure revealed by these particles and the conformal structure revealed by light propagation together determine a unique Weyl structure. As discussed by EPS, parallel transport along non-null curves is then well defined. Finally, the absence of the second clock effect is then the necessary and sufficient condition for the existence of a unique Riemannian metric.

In conclusion, the truth of the constructive axioms of EPS is epistemically decidable in a noncircular manner, and the metric structure derived from the conformal and projective structures and their compatibility relation is therefore not even epistemologically conventional but constitutes an intrinsic feature of the spacetime manifold that is revealed through light propagation and free fall.

[1] J. Ehlers, F.A.E. Pirani, and A. Schild, "The Geometry of Free Fall and Light Propagation," in *General Relativity, Papers in Honour of J.L. Synge*, edited by L. O'Raifeartaigh (Oxford U.P., New York, 1972), pp. 63–84.

[2] J. Ehlers and E. Köhler, "Path Structures on Manifolds," J. Math. Phys. **18**, 2014 (1977).

[3] C. Ehresmann, "Les Prolongements d'une Variété Différentiable I-V," C. R. Acad. Sci. **233**, 598, 777, 1081 (1951); **234**, 1028, 1424 (1952).

[4] F.A.E. Pirani, Symposia Mathematica, Rome, pp. 67–83 (see p. 74) (1973); J. Ehlers, "Survey of General Relativity Theory," in *Relativity, Astrophysics and Cosmology*, edited by W. Israel (Reidel, Dordrecht, (1973), pp. 1–125, see p. 30.

[5] A. Grübaum, *Philosophical Problems of Space and Time*, (Reidel, Dordrecht, 1973), Chap. 22.

[6] J. Stachel, "Space-Time Problems" (A review of the Synge Festschrift cf. 1) in Science **180**, 292 (1973).

[7] L. Sklar, "Facts, Conventions and Assumptions," in *Foundations of Space-Time Theories*, Minnesota Studies in the Philosophy of Science, VIII.

[8] J.A. Winnie, "The Causal Theory of Space-Time" in *Foundations of Space–Time Theories* (cf., Ref. 7).

[9] S. Kobayashi, and K. Nomizu, *Foundations of Differential Geometry*, (Interscience, New York, 1963), Vol. I.

[10] J.A. Schouten, *Ricci-Calculus* (Springer, Berlin, 1954), p. 67.

[11] Ref. 10, p. 155.

[12] L.P. Eisenhart, "Non-Riemannian Geometry," Am. Math. Soc. Colloq. Publ. **VIII**, 111 (1964).

[13] A. Einstein, "What is the Theory of Relativity?" in *Essays in Science* (Philosophical Library, New York, 1934).

[14] (a) H. Weyl, "Zur Infinitessimal Geometry: Einordnug der Projectiven und Konformen Auffassung," in *Gessammelte Abhandlungen*, edited by K. Chandrasekharan (Springer, New York, 1968), Vol. II, p. 195. (b) H. Weyl, *Raum, Zeit, Materie* (Springer, Berlin, 1970).

[15] H. Reichenbach, *Axiomatization of the Theory of Relativity* (Univeristy of California, Berkely, 1969).

[16] A. Trautman, "Invariance of Lagrangian Systems" in *General Relativity*, edited by L. O'Raifeartaigh, (Oxford U.P., New York, 1972).

[17] A. Einstein, "Reply to Criticisms," in *Albert Einstein Philosopher–Scientist*, edited by P.A. Schilpp (Open Court, LaSalle, Illinois, 1970), Vol. II.

[18] H. von Helmholtz *Über die Tatsachen, Welche der Geometrie Zugrunde Liegen*, Nachr. Ges. Wiss. Göttingen (1868).

[19] H. Weyl, "Geometrie und Physik," in *Gesammelte Abhandlungen*, Vol. III, p. 336 (cf., Ref. 14).

[20] J. Ehlers, and A. Schild, "Geometry in a Manifold with Projective Structure," Comm. Math. Phys. **32**, 119 (1973), see p. 146.

[21] L. Infeld, and A. Schild, Rev. Mod. Phys. **21**, 408 (1949); A. Papapetrou, Proc. R. Soc. London A **209**, 248 (1951); J.A. Wheeler, Rev. Mod. Phys. **33**, 63 (1961).

A global theory of supermanifolds

Alice Rogers

Theoretical Physics Group, Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ

(Received 5 April 1979; accepted for publication 27 July 1979)

A mathematically rigorous definition of a global supermanifold is given. This forms an appropriate model for a global version of superspace, and a class of functions is defined which corresponds to superfields. This new construction is compared with several pre-existing definitions of supermanifold and graded manifold; it is shown to include all these definitions and to go beyond them, particularly in admitting the possibility of nontrivial topology in the anticommuting sector. Local differential geometry and potential applications to supergravity are considered.

1. INTRODUCTION

In the context of supersymmetry Salam and Strathdee[1] have introduced the concept of "superspace," a space parametrized by eight coordinates $(x_1,x_2,x_3,x_4,\theta_1,\theta_2,\theta_3,\theta_4)$ with the first four coordinates taking their values in the even part of a Grassman algebra and the second four in the odd part. Until now superspace has been used as a mathematical device for generating local structure. However, the question arises, do we actually "live" in superspace? More precisely, is superspace a more appropriate model for the universe than the conventional four-dimensional space time? Such a question cannot immediately be answered, but, bearing in mind the importance of global topology in quantum field theory to monopoles, instantons, anomalies, etc. (in both a Yang–Mills and a gravitational context), it seems natural to attempt to construct a global "supermanifold," with local coordinates of the kind described above (so that it is locally equivalent to the many local definitions of superspace), modeled on the conventional definition of a differentiable manifold with real local coordinates.

Such a construction, referred to as a G^∞ supermanifold, is described in this paper. The definition is a mathematically rigorous one, and embraces the definitions of supermanifold and "graded manifold" given by several other authors in a manner made explicit in Secs. 3 and 4 of this paper. It is simpler than these definitions, but also much broader in scope, admitting a wider class of topologies, including a topologically interesting contribution from the "odd" part of the supermanifold, whereas all the other definitions are essentially trivial in the θ sector. Also, it allows for the case where the Grassman algebra is infinite dimensional. In applications to quantized superfields it is essential to use an infinite-dimensional algebra if one is to avoid placing undesirable restrictions on Green's functions.

Having constructed a G^∞ supermanifold, it is possible to make local constructions such as tangent vectors very much as on an ordinary C^∞ manifold. The G^∞ supermanifold formalism seems to be an appropriate global formalism for the various local formulations of differential geometry on superspace given by Arnowitt and Nath, Bedding, Downes-Martin and Taylor, Brink *et al.*, Ogievetsky and Sokatchev, Siegel and Gates, Wess and Zumino,[2] and other authors.

As well as superspace, there are other objects one considers which have even and odd local coordinates, notably the local "super Lie group" obtained by exponentiating the graded algebra of supersymmetry (cf. Salam and Strathdee[1]). A G^∞ supermanifold seems to be the appropriate global framework for this situation, with the vector field structure allowing one to identify the infinitesimal algebra with the set of left invariant vector fields, and the exponential map becoming an integral curve exactly as in the classical case.

The plan of the paper is to present in Sec. 2 a definition of differentiability of functions defined on a Grassman algebra, and various results of an analytical nature; Sec. 3 contains the definition of a G^∞ supermanifold and a comparison with the supermanifolds of DeWitt[3] and Batchelor[4]; in Sec. 4 (with details in the appendix) the connection between G^∞ supermanifolds and the graded manifolds of Kostant[5] (or the related supermanifold of Berezin and Leĭtes[6]) is analyzed; Sec. 5 contains a discussion of vector fields on supermanifolds and in Sec. 6 potential applications to supergravity are considered, and a comparison made with the work of Dell and Smolin.[7,8] Section 7 summarizes the paper and includes a table comparing the various definitions of supermanifolds discussed here.

Throughout this paper all vector spaces, algebras, etc., are over the real field; extension to the complex field is relatively straightforward.

2. REAL ANALYSIS EXTENDED TO GRASSMAN ALGEBRAS

In this section classes of differentiable functions on Grassman algebras are defined; they are a key ingredient in the definition of a G^∞ supermanifold. The definition is modeled as closely as possible on the usual definition of a C^r function, but using the multiplicative structure of the Grassman algebra rather than that of the real numbers; it agrees with the usual heuristic concept of superfield, and makes this concept more rigorous.

First some definitions and results of an algebraic nature are required. Let L be a finite positive integer and B_L denote the real Grassman algebra over $\mathbb{R}^L$. Employing Kostant's notation,[5] let M_L denote the set of sequences

$$\{\mu|\mu = (\mu_1,\mu_2,...,\mu_k), 1\leqslant\mu_1 < \mu_2 .. \mu_k \leqslant L,\ \mu_i \text{ an integer for } 1\leqslant i\leqslant k\},$$

Let Ω represent the empty sequence in M_L, and (j) denote the sequence with just one element j; then a basis of B_L exists of the form $\{\beta_\mu|\mu\in M_L\}$ with

$$\beta_\Omega = 1 \text{ (the unit of } B_L)$$

$$\beta_\mu = \beta_{(\mu_1)}\beta_{(\mu_2)}...\beta_{(\mu_k)}, \quad \text{for all } \mu\in M_L, \tag{2.1}$$

and

$$\beta_{(i)}\beta_{(j)} = -\beta_{(j)}\beta_{(i)} \text{ for } 1\leqslant i,j\leqslant L.$$

A complete norm $||\ ||$ is now defined on B_L as follows: given $x\in B_L$, with $x = \Sigma_{\mu\in M_L} x_\mu\beta_\mu$ (where the x_μ belong to $\mathbb{R}$ for all μ in M_L),

$$\text{let } ||x|| := \sum_{\mu\in M_L} |x_\mu|. \tag{2.2}$$

With this norm B_L becomes a Banach space; it is easily shown that B_L is a Banach algebra, i.e., $||1|| = 1$ and $||ab||\leqslant||a||\,||b||$ for all a,b in B_L. This algebraic structure of B_L is crucial in much of the analysis which follows, and it is for this reason that the norm defined above is used rather than the usual Euclidean norm, which is an equivalent norm but does not make B_L into a Banach algebra. B_L is an example of a Z_2-graded commutative algebra which, again following Kostant,[5] is defined as follows:

Definition 2.1: (a) A Z_2-graded vector space is a vector space B which is the direct sum of two subspaces B_0 and B_1. (b) A Z_2-graded algebra is a Z_2-graded vector space B which is an algebra such that $1\in B_0, B_0B_0\subset B_0, B_0B_1\subset B_1, B_1B_0 \subset B_1$, and $B_1B_1\subset B_0$. (c) A Z_2-graded algebra is graded commutative if $ab = (-1)^{pq}ba$ whenever a belongs to B_p and b belongs to B_q.

B_0 is called the even part of B, and B_1 the odd part; an element b of B is said to be homogeneous if it belongs to B_0 or to B_1. If b is homogeneous and nonzero, the degree of b is defined by $|b| := 0$ if b belongs to B_0 and $|b| := 1$ if b belongs to B_1.

All the gradings which occur in this paper are Z_2 gradings, and the word "graded" may be taken to mean Z_2 graded throughout. Statements which are made about homogeneous elements can be extended to general elements by linearity.

Definition 2.2: A graded left B module is a graded vector space W which is also a left B module in the normal sense with $B_0W_0\subset W_0, B_1W_0\subset W_1, B_0W_1\subset W_1$, and $B_1W_1\subset W_0$.

When one considers Green's functions of quantized superfields (see Sec. 6) it is desirable to consider an infinite-dimensional analog of B_L. Let B_∞ be the vector space l_1 of infinite sequences of real numbers $(x_1,x_2,...)$ such that $\Sigma_{i=1}^\infty |x_i| < \infty$. With the usual l_1 norm, i.e., $||(x_1,x_2,...)|| := \Sigma_{i=1}^\infty |x_i|$, B_∞ becomes a Banach space.

A multiplication is now defined on B_∞, under which B_∞ becomes a Banach algebra. To do this it is convenient to index the elements of each sequence with elements of $M_\infty := \cup_{N=1}^\infty M_N$. This involves setting up the following one-to-one correspondence between the positive integers and M_∞: $1\leftrightarrow(1), 2\leftrightarrow(2), 3\leftrightarrow(1,2)$, etc., and in general $r\leftrightarrow\mu$ [where r is a positive integer and $\mu = (\mu_1,...,\mu_k)$ is an element of M_∞] if $r = \frac{1}{2}(2^{\mu_1} + 2^{\mu_2} + \cdots + 2^{\mu_k})$. The sequence $(x_1,x_2,x_3,...)$ is then written $(x_{(1)},x_{(2)},x_{(1,2)},...)$. Given μ in M_∞, define β_μ to be the sequence in B_∞ with $x_\mu = 1$ and all other elements zero. Then an arbitrary element of B_∞ may be expressed as a norm convergent sum in the following manner:

$$(x_{(1)},x_{(2)},x_{(1,2)},...) = \sum_{\mu\in M_\infty} x_\mu\beta_\mu. \tag{2.3}$$

Multiplication on $\{\beta_\mu|\mu\in M_\infty\}$ is defined by

$$\beta_\Omega\beta_\mu := \beta_\mu\beta_\Omega := \beta_\mu, \quad \text{for all } \mu \text{ in } M_\infty,$$

$$\beta_{(i)}\beta_{(j)} := -\beta_{(j)}\beta_{(i)}, \text{ for all positive integers } i,j, \tag{2.4}$$

and

$$\beta_{(\mu_1)}...\beta_{(\mu_k)} := \beta_\mu, \text{ for all } \mu \text{ in } M_\infty.$$

The following proposition establishes that this multiplication can be extended by linearity and continuity to the whole of B_∞, and that B_∞ is a Banach algebra under this multiplication.

Proposition 2.3: Suppose a,b belong to B_∞ with $a = \Sigma_{\mu\in M_\infty} a_\mu\beta_\mu, b = \Sigma_{\mu\in M_\infty} b_\mu\beta_\mu$ (where the a_μ and b_μ are real numbers). Then, if $ab := \Sigma_{\mu\in M_\infty}\Sigma_{\nu\in M_\infty} a_\mu b_\nu\beta_\mu\beta_\nu$, $ab\in B_\infty$ and $||ab||\leqslant||a||\,||b||$. Thus, B_∞ is a Banach algebra.

Proof: For each positive integer s let $s(a,b) := ||\Sigma_{\mu\in M_s}\Sigma_{\nu\in M_s} a_\mu b_\nu\beta_\mu\beta_\nu||$. Then $s(a,b)\leqslant||a||\,||b||$ and thus $[s(a,b)]$ is a monotonically increasing sequence which is bounded above, and so tends to a limit as s tends to infinity. Thus $ab\in B_\infty$ and $||ab|| = \lim_{s\to\infty} s(a,b)\leqslant||a||\,||b||$. □

In the rest of the paper, except where indicated otherwise, L may be a finite positive integer or ∞; m,n will always denote finite positive integers.

Let $B_L^{m,n}$ denote the set $B_{L,0}^m\times B_{L,1}^n$, i.e., the Cartesian product of m copies of the even part of B_L and n copies of the odd part. A typical element of this set will be written $(a_1,...,a_m,b_1,...,b_n) = (\mathbf{a},\mathbf{b})$, or $(c_1,...,c_{m+n}) = (\mathbf{c})$. The norm on $B_L^{m,n}$ is defined by $||(\mathbf{c})|| := ||c_1|| + \cdots + ||c_{m+n}||$, and the topology on $B_L^{m,n}$ is the topology induced by this norm (which is also the product topology). It is of course a Hausdorff topology and, in the the case where L is finite, is the usual topology on $B_L^{m,n}$ regarded as a $2^{L-1}(m+n)$-dimensional vector space over the real numbers. A crucial difference between the supermanifolds defined in this paper and those of Batchelor[4] and DeWitt[3] is that Batchelor and DeWitt both use a coarser, non-Hausdorff topology on $B_L^{m,n}$.

Proposition 2.4: Suppose $\alpha_r\in B_L (r = 1,...,m+n)$ satisfy $\Sigma_{i=1}^m h_i\alpha_i + \Sigma_{j=1}^n k_j\alpha_{j+m} = 0$ for all $(\mathbf{h},\mathbf{k})$ in $B_L^{m,n}$. Then, (a) if $1\leqslant i\leqslant m$, $\alpha_i = 0$; (b) if $1\leqslant j\leqslant n$ and L is finite, $\alpha_{j+m} = \lambda\beta_{(1,2,...,L)}$, where λ is a real number; (c) if $1\leqslant j\leqslant n$ and $L = \infty, \alpha_{j+m} = 0$.

Proof: (a) The result follows immediately on letting $\mathbf{k} = \mathbf{0}$ and $\mathbf{h} = (1,0,...,0),(0,1,0,...,0),...,(0,...,0,1)$ in turn. (b) and (c) suppose $\alpha_{j+m} = \Sigma_{\mu\in M_L}\alpha_{j+m}^\mu\beta_\mu$ (where the α_{j+m}^μ are real numbers).

Putting $\mathbf{h} = \mathbf{0}, k_j = \beta_{(s)} (1\leqslant s\leqslant L)$ and $k_{j'} = 0 (j'\neq j)$

shows that $\Sigma_{\mu\in M_L}\alpha^{\mu}_{j+m}\beta_{(s)}\beta_{\mu}=0$ for all $s, 1\leqslant s\leqslant L$, and thus that $\alpha^{\mu}_{j+m}=0$ if $L=\infty$, or if L is finite and $\mu\neq(1,2,...,L)$. Hence, $\alpha_{j+m}=\lambda\beta_{(1,2,...,L)}$ if L is finite, and $\alpha_{j+m}=0$ if $L=\infty$. □

The definition of a differentiable function is now given; this is of vital importance in the definition of a G^{∞} supermanifold. It is a simpler definition than Batchelor's,[4] and more mathematically rigorous than DeWitt's.[3]

Definition 2.5: Let U be an open set in $B_L^{m,n}$ and $f:U\to B_L$. Then, (a) f is said to be G^0 on U iff f is continuous on U. (b) f is said to be G^1 on U if there exist $m+n$ functions $G_k f:U\to B_L, k=1,...,m+n$ and a function $\eta:B_L^{m,n}\to B_L$ such that, if $(\mathbf{a},\mathbf{b})$ and $(\mathbf{a}+\mathbf{h},\mathbf{b}+\mathbf{k})\in U$,

$$f(\mathbf{a}+\mathbf{h},\mathbf{b}+\mathbf{k})=f(\mathbf{a},\mathbf{b})+\sum_{i=1}^{m}h_i(G_i f)(\mathbf{a},\mathbf{b})+\sum_{j=1}^{n}k_j(G_{j+m}f)(\mathbf{a},\mathbf{b})+||(\mathbf{h},\mathbf{k})||\eta(\mathbf{h},\mathbf{k}) \tag{2.5}$$

and

$$||\eta(\mathbf{h},\mathbf{k})||\to 0 \text{ as } ||(\mathbf{h},\mathbf{k})||\to 0.$$

(It follows from proposition 2.4 that the "partial derivatives" $G_k f$ may not be unique if L is finite and $m+1\leqslant k\leqslant n$; it is shown later that there is a natural choice of $G_k f$ in such cases.) (c) The definition of G^p, where p is a finite positive integer, is made inductively. f is said to be G^p on U iff f is G^1 on U and it is possible to choose $G_k f (1\leqslant k\leqslant m+n)$, which are G^{p-1} on U. (d) f is said to be G^{∞} on U iff f is G^p on U for any positive integer p. (e) f is said to be G^{ω} on U if, given any $(\mathbf{p})$ in U, there exists a neighborhood N_p of $(\mathbf{p})$ such that, for all $(\mathbf{q})\in N_p$, $f(\mathbf{q})$ is equal to the sum of an absolutely convergent power series in $(\mathbf{p}-\mathbf{q})$ of this form:

$$f(\mathbf{q})=\sum_{k_1=0...k_{m+n}=0}^{\infty}a_{k_1...k_{m+n}}(q_1-p_1)^{k_1}\cdots(q_{m+n}-p_{m+n})^{k_{m+n}} \quad (\text{with the } a_{k_1...k_{m+n}} \text{ in } B_L). \tag{2.6}$$

(f) Let s be a finite positive integer and $g:U\to B_L^s$ (i.e., the Cartesian product of s copies of B_L); also let p_k denote the kth projection function [i.e., $p_k(c_1,...,c_s)=c_k$]. Then, g is said to be G^r on U if $p_k\circ g$ is G^r on U for $k=1,...,s$. (r may be a positive integer or ∞ or ω.)

The definition of a G^r function bears an obvious resemblance to the usual definition of a C^r function (although the replacement of multiplication in the real numbers by multiplication in the Grassman algebra means that it is in fact a much more restrictive definition, as is made clear below). The form of "differentiation" introduced is inspired by Salam and Strathdee[9] and seems to be effectively the same as that used by DeWitt[3] in the case where L is finite; but it is difficult to compare the definitions directly because DeWitt does not use a norm on the Grassman algebra, and the infinite-dimensional algebra which he uses, being an algebra of formal power series, does not admit a Banach algebra structure. After proposition 2.11, a detailed comparison of the classes of G^{∞} functions and differentiable functions (in the DeWitt sense) is made.

In the rest of this paper attention is focused on G^{∞} functions. Many of the results can be generalized to G^r functions, where r is a finite positive integer or ω, but G^{∞} functions are physically the most useful class, corresponding as they do to a natural choice of superfield.

A simple example of a G^{∞} function is

$$f:B_L^{2,2}\to B_L \quad (L>1), \quad (fa_1,a_2,b_1,b_2):=ca_1a_2^2b_1b_2 \tag{2.7}$$

(with c some fixed element of $B_{L,0}$); by differentiation "from first principles" one calculates that

$$G_1 f(a_1,a_2,b_1,b_2)=ca_2^2b_1b_2,$$
$$G_2 f(a_1,a_2,b_1,b_2)=2ca_1a_2b_1b_2,$$
$$G_3 f(a_1,a_2,b_1,b_2)=ca_1a_2^2b_2, \text{ and}$$
$$G_4 f(a_1,a_2,b_1,b_2)=-ca_1a_2^2b_1.$$

In fact, any function which is a finite or absolutely convergent power series of the form $f:U\to B_L$ (where U is open in $B_L^{m,n}$),

$$f(\mathbf{c})=\sum_{k_1...k_{m+n}=0}^{\infty}a_{k_1...k_{m+n}}c_1^{k_1}\cdots c_{m+n}^{k_{m+n}}, \quad a_{k_1...k_{m+n}}\in B_L, \tag{2.8}$$

is G^{∞} on U (and thus all G^{ω} functions are G^{∞}), but functions exist which are G^{∞} but not G^{ω} (as is evident from proposition 2.11).

The two algebraic lemmas which now follow are used to prove the crucial proposition 2.8, which establishes that a G^{∞} function is also a C^{∞} function.

Lemma 2.6: Let p be a finite positive integer and let $f:(B_L^{m,n})^p\to B_L$ be of the form

$$f[(\mathbf{a}^{(1)}),(\mathbf{a}^{(2)}),...,(\mathbf{a}^{(p)})]=\sum_{k_1=1...k_p=1}^{m+n}a_{k_1}^{(1)}\cdots a_{k_p}^{(p)}\alpha_{k_1...k_p}, \tag{2.9}$$

where the $\alpha_{k_1...k_p}$ belongs to B_L. Then f belongs to the space $\mathscr{L}[(B_L^{m,n})^p,B_L]$ of continuous p-linear maps of $(B_L^{m,n})^p$ into B_L.

Proof: f is easily seen to be p linear and it follows from the Banach algebra property of B_L that

$$||f[(\mathbf{a}^{(1)}),...,(\mathbf{a}^{(p)})]||\leqslant||(\mathbf{a}^{(1)})||\cdots||(\mathbf{a}^{(p)})||\sum_{k_1=1...k_p=1}^{m+n}||\alpha_{k_1...k_p}||.$$

(Note that this result is true whether L is finite or infinite.) □

The following projection maps (referred to by DeWitt as "body" and "soul", the former being more conventionally known as the augmentation map) are used in the next Lemma and frequently thereafter:

$$\epsilon:B_L\to\mathbb{R}, \epsilon\Big(\sum_{\mu\in M_L}a_\mu\beta_\mu\Big):=a_\Omega \quad (\text{where the } a_\mu\in\mathbb{R}),$$
$$\epsilon:B_L^{m,n}\to\mathbb{R}^m, \epsilon(\mathbf{a},\mathbf{b}):=[\epsilon(a_1),...,\epsilon(a_m)], \tag{2.10}$$
$$s:B_L\to B_L, s(a):=a-\epsilon(a)1,$$
$$s:B_L^{m,n}\to B_L^{m,n},$$
$$s(\mathbf{a},\mathbf{b}):=[s(a_1),...,s(a_m),s(b_1),...,s(b_n)].$$

[Note that if b is an odd element of B_L, $\epsilon(b)=0$ and $s(b)=b$.]

In the case where L is finite, $B_L\simeq\mathbb{R}\oplus N$, where N is the subspace of B_L consisting of nilpotent elements; ϵ is the projection onto $\mathbb{R}$, and s onto N. In the case where $L=\infty$, the elements of $s(B_L)$ are not actually nilpotent, but very nearly so as the next Lemma shows.

Lemma 2.7: Let $d\in B_L$, and $c = s(d)$. Then, (a) if L is finite, c is nilpotent (i.e, $c^q = 0$ for some finite positive integer q); (b) if $L = \infty$, there exist $\lambda, \theta \in \mathbb{R}$ with $\theta < 1$ such that $\|c^t\| \leqslant \lambda\theta^t$, $t = 1,2,\ldots$.

Proof: (a) If L is finite, $c^{L+1} = 0$. (b) Suppose $L = \infty$, and that $c = \Sigma_{\mu\in M_\infty} c_\mu \beta_\mu$ (where the $c_\mu \in \mathbb{R}$). Then, because $\Sigma_{\mu\in M_\infty} |c_\mu| < \infty$, it is possible to choose $a,b\in s(B_L)$ such that $c = a + b$, $a = \Sigma_{\mu\in M_\infty} a_\mu\beta_\mu$ (where N is some finite integer, and $a_\Omega = 0$) and $b = \Sigma_{\mu\in M_\infty} b_\mu \beta_\mu$ with $\Sigma_{\mu\in M_\infty} |b_\mu| = \delta < 1$; i.e., $c = a + b$ with $a^{N+1} = 0$ and $\|b\| = \delta < 1$. If $b = 0$, the result follows immediately. Suppose $b \neq 0$ (and thus $N > 1$). Then

$$\begin{aligned}\|c^{h+N}\| &= \|(a+b)^{h+N}\| \\ &\leqslant \sum_{r=0}^{N} \binom{h+N}{r} \|a^r\| \delta^{h+N-r} \\ &\leqslant \delta^h \sum_{r=0}^{N} \binom{N}{r} \|a\|^r \delta^{N-r} \\ &\quad \times \frac{(N+1)\ldots(N+h)}{(N-r+1)\ldots(N-r+h)} \\ &\leqslant \delta^h \left(\prod_{k=1}^{h} \frac{N+k}{k} \right) \sum_{r=0}^{N} \binom{N}{r} \|a\|^r \delta^{N-r} \\ &= \delta^h \left(\prod_{k=1}^{h} \frac{N+k}{k} \right) (\|a\| + \delta)^N.\end{aligned}$$

Now let m be the integer such that $m/(N+m) \leqslant \delta < [(m+1)/(N+m+1)]$, and let $\theta = \delta(N+m+1)/(m+1)$; then, $0 < \theta < 1$ and

$$\|c^{h+N}\| \leqslant \delta^m \left[\prod_{k=1}^{m} (N+k)/k \right] \theta^{h-m} (\|a\| + \delta)^N.$$

Hence, if

$$\lambda = \max\left[\frac{\|c\|}{\theta}, \frac{\|c^2\|}{\theta^2}, \ldots, \frac{\|c^N\|}{\theta^N}, \frac{\delta^m \left(\prod_{k=1}^{m} \frac{N+k}{k} \right)(\|a\| + \delta)^N}{\theta^{N+m}} \right],$$

then $\|c^t\| \leqslant \lambda\theta^t$. □

The next proposition establishes that a G^∞ function is also a C^∞ function (regarding $B_L^{m,n}$ and B_L as Banach spaces) and gives the expression for the total derivative of a G^∞ function in terms of the partial derivatives G_k ($k = 1, \ldots, m+n$). The converse is certainly not true: it is easy to construct C^∞ functions which are not G^∞ functions, such as

$$f: B_L^{1,0} \rightarrow B_L, f(a) = \epsilon(a)1\,. \tag{2.11}$$

Propostion 2.8: If U is open in $B_L^{m,n}$ and $f\in G^\infty(U)$, then $f\in C^\infty(U, B_L)$, the space of C^∞ maps of U into B_L. Also, regarding the pth total derivative of f as an element of $\mathscr{L}[(B_L^{m,n})^p, B_L]$,

$$\begin{aligned}&[D^p f(\mathrm{c})]\,[(\mathrm{l}^{(1)}), (\mathrm{l}^{(2)}), \ldots, (\mathrm{l}^{(p)})] \\ &\quad = \sum_{k_1=1\ldots k_p=1}^{m+n} l_{k_1}^{(1)} \ldots l_{k_p}^{(p)} (G_{k_p} G_{k_{p-1}} \ldots G_{k_1} f)(\mathrm{c}),\end{aligned} \tag{2.12}$$

for all (c) in U, $[(l^{(1)}), \ldots, (l^{(p)})]$ in $(B_L^{m,n})^p$.

Proof: Let (c), (c + h)$\in U$. Then, since $f\in G^\infty(U)$,

$$f(\mathrm{c}+\mathrm{h}) = f(\mathrm{c}) + \sum_{k=1}^{m+n} h_k G_k f(\mathrm{c}) + \|(\mathrm{h})\| \eta(\mathrm{h})$$

[where $\|\eta(\mathrm{h})\| \rightarrow 0$ as $\|(\mathrm{h})\| \rightarrow 0$]. Thus, f is C^1 and

$$D^1 f(\mathrm{c})(\mathrm{l}^{(1)}) = \sum_{k=1}^{m+n} l_k^{(1)} (G_k f)(\mathrm{c})\,.$$

Proceeding by induction, suppose that f is C^p and

$$D^p f(\mathrm{c}')[(\mathrm{l}^{(1)}), \ldots, (\mathrm{l}^{(p)})] = \sum_{k_1=1\ldots k_p=1}^{m+n} l_{k_1}^{(1)} \ldots l_{k_p}^{(p)} (G_{k_p} \ldots G_{k_1} f)(\mathrm{c}')$$

for all (c′) in U; then,

$$\begin{aligned}D^p f(\mathrm{c}+\mathrm{h})[(\mathrm{l}^{(1)}), \ldots, (\mathrm{l}^{(p)})] &= \sum_{k_1=1\ldots k_p=1}^{m+n} l_{k_1}^{(1)} \ldots l_{k_p}^{(p)} (G_{k_p} G_{k_{p-1}} \ldots G_{k_1} f)(\mathrm{c}+\mathrm{h}) \\ &= D^p f(\mathrm{c})[(\mathrm{l}^{(1)}), \ldots, (\mathrm{l}^{(p)})] + \sum_{k_1=1\ldots k_{p+1}=1}^{m+n} l_{k_1}^{(1)} \ldots l_{k_p}^{(p)} h_{k_{p+1}} (G_{k_{p+1}} G_{k_p} \ldots G_{k_1} f)(\mathrm{c}) \\ &\quad + \|(\mathrm{h})\| \sum_{k_1=1\ldots k_p=1}^{m+n} l_{k_1}^{(1)} \ldots l_{k_p}^{(p)} \eta_{k_1\ldots k_p}(\mathrm{h}),\end{aligned}$$

where $\|\eta_{k_1\ldots k_p}(\mathrm{h})\| \rightarrow 0$ as $\|(\mathrm{h})\| \rightarrow 0$. Hence, $D^p f(\mathrm{c}+\mathrm{h}) = D_p f(\mathrm{c}) + F(\mathrm{c})(\mathrm{h}) + \|\mathrm{h}\| \eta(\mathrm{h})$, where $F(\mathrm{c}) \in \mathscr{L}[(B_L^{m,n})^{p+1}, B_L]$ with

$$F(\mathrm{c})[(\mathrm{h}), (\mathrm{l}^{(1)}), \ldots, (\mathrm{l}^{(p)})] = \sum_{k_1=1\ldots k_{p+1}=1}^{m+n} l_{k_1}^{(1)} \ldots l_{k_p}^{(p)} h_{k_{p+1}} (G_{k_{p+1}} G_{k_p} \ldots G_{k_1} f)(\mathrm{c}),$$

and $\eta(\mathrm{h}) \in \mathscr{L}[(B_L^{m,n})^p, B_L]$ with

$$\eta(\mathrm{h})[(\mathrm{l}^{(1)}), \ldots, (\mathrm{l}^{(p)})] = \sum_{k_1=1\ldots k_p=1}^{m+n} l_{k_1}^{(1)} \ldots l_{k_p}^{(p)} \eta_{k_1\ldots k_p}(\mathrm{h})\,.$$

Now it follows from the Banach algebra property of B_L that

$$\|[\eta(\mathrm{h})]\,[(\mathrm{l}^{(1)}), \ldots, (\mathrm{l}^{(p)})]\| \leqslant \|(\mathrm{l}^{(1)})\| \ldots \|(\mathrm{l}^{(p)})\| \sum_{k_1=1\ldots k_p=1}^{m+n} \|\eta_{k_1\ldots k_p}(\mathrm{h})\|\,.$$

Hence, $\|\eta(\mathrm{h})\| \rightarrow 0$ as $\|(\mathrm{h})\| \rightarrow 0$. Thus, f is C^{p+1}, $F(\mathrm{c}) = D^{p+1} f(\mathrm{c})$, and the theorem is proved by induction. □

Corollary 2.9:

Given $(\mathbf{c}) \in B_L^{m,n}$, and q a finite positive integer, let $(\mathbf{c})^q := [(\mathbf{c}), (\mathbf{c}), \ldots, (\mathbf{c})]$ (q terms). Then, (a) if $(\mathbf{a},\mathbf{b})$, $(\mathbf{a}+\mathbf{h}, \mathbf{b}+\mathbf{k}) \in U$, $\epsilon(h) = 0$, and L is finite,

$$f(\mathbf{a}+\mathbf{h}, \mathbf{b}+\mathbf{k}) = f(\mathbf{a},\mathbf{b}) + \sum_{q=1}^{L} \frac{1}{q!} D^q f(\mathbf{a},\mathbf{b})[(\mathbf{h},\mathbf{k})^q] ; \tag{2.13}$$

(b) if $(\mathbf{a},\mathbf{b})$ and $(\mathbf{a},\mathbf{b}+\mathbf{k}) \in U$,

$$f(\mathbf{a},\mathbf{b}+\mathbf{k}) = f(\mathbf{a},\mathbf{b}) + \sum_{q=1}^{n} \frac{1}{q!} D^q f(\mathbf{a},\mathbf{b})[(\mathbf{0},\mathbf{k})]^q . \tag{2.14}$$

Proof: An immediate consequence of Taylor's theorem in this form:

$$f(\mathbf{a}+\mathbf{h}, \mathbf{b}+\mathbf{k}) = f(\mathbf{a},\mathbf{b}) + \sum_{q=1}^{P} \frac{1}{q!} D^q f(\mathbf{a},\mathbf{b})[(\mathbf{h},\mathbf{k})^q] + \frac{1}{(p+1)!} D^{p+1} f(\mathbf{a}+\theta\mathbf{h}, \mathbf{b}+\theta\mathbf{k})[(\mathbf{h},\mathbf{k})^{p+1}] \tag{2.15}$$

(where $0 \leqslant \theta \leqslant 1$), together with the fact that (a) if L is finite, the product of $L+1$ (or more) nilpotent elements of B_L is always zero and (b) for all values of L, the square of an odd element of B_L is always zero. □

The fact that Taylor series in nilpotent elements always terminate has many uses; as far as nilpotent elements are concerned, the distinction between G^ω and G^∞ disappears, and makes possible two kinds of "analytic continuation" of differentiable functions when L is finite, described in definition 2.10 and proposition 2.11(c). It also allows one to specify a natural choice of kth partial derivative when ambiguities arise. The first kind of continuation is now defined.

Definition 2.10:

Let L be finite, $L \geqslant n$, let U be open $B_L^{m,n}$, and V be open in $\mathbb{R}^m$ with $V = \epsilon(U)$. Define $z: C^\infty(V, B_L) \to \{$functions of U into $B_L\}$ by

$$z(f)(a_1, \ldots, a_m, b_1, \ldots, b_n) := \sum_{i_1 = 0 \ldots i_m = 0}^{L} \frac{1}{i_1! \ldots i_m!} \times (\partial_1^{i_1} \ldots \partial_m^{i_m} f)[\epsilon(\mathbf{a})] s(a_1)^{i_1} \ldots s(a_m)^{i_m} , \tag{2.16}$$

where ∂_j denotes the unusual jth partial derivative.

The definition allows one to extend any function in $C^\infty(V, B_L)$ to an element of $G^\infty(U)$, as is proved in the next proposition. The definition is motivated by the Taylor series expansion of f, which, in terms of partial derivatives, is

$$f(\mathbf{x}+\mathbf{y}) = \sum_{\substack{i_1 = 0 \ldots i_m = 0 \\ i_1 + \cdots + i_m \leqslant p}} \frac{1}{i_1! \ldots i_m!} (\partial_1^{i_1} \ldots \partial_m^{i_m} f)(\mathbf{x}) y_1^{i_1} \ldots y_m^{i_m} + O(\|\mathbf{y}\|^{p+1}) ,$$

where $\mathbf{x}, \mathbf{x}+\mathbf{y} \in V \subset \mathbb{R}^m$. $\mathbf{x}$ and $\mathbf{y}$ are replaced by $\epsilon(\mathbf{a})$ and $s(\mathbf{a})$, respectively, and the nilpotence of $s(\mathbf{a})$ is used. As well as establishing that $z(f)$ is a G^∞ continuation of the C^∞ function f, proposition 2.11 shows how a G^∞ function may be uniquely continued.

Proposition 2.11: Suppose L is finite. (a) If U is open in $B_L^{m,n}$ and $f \in C^\infty[\epsilon(U)]$, then $z(f) \in G^\infty(U)$, $G_i z(f) = z(\partial_i f)$ ($i = 1, \ldots, m$) and a possible choice of $G_{j+m} z(f)$ is $G_{j+m} z(f) = 0$ ($j = 1, \ldots, n$). (b) z is an isomorphism of $C^\infty[\epsilon(U), B_L]$ onto its image. (c) Let W also be open in $B_L^{m,n}$ with $U \subset W$ and $\epsilon(U) = \epsilon(W)$, and let $f \in G^\infty(U)$. Then there exists a unique $f' \in G^\infty(W)$ such that $f'|_U = f$, i.e., f' is the unique continuation of f to W. (d) Let $v_j: U \to B_L$, $v_j(\mathbf{a},\mathbf{b}) := b_j$ ($j = 1, \ldots, n$); also let $v_\mu := v_{\mu_1} \ldots v_{\mu_k}$, where $\mu = (\mu_1, \ldots, \mu_k) \in M_n$, $\mu \neq \Omega$, and let $v_\Omega := 1$. Then, given $f \in G^\infty(U)$, there exist uniquely determined $f_\mu \in C^\infty[\epsilon(U), B_L]$ such that $f = \Sigma_{\mu \in M_n} v_\mu z(f_\mu)$. (This will be referred to as the z expansion of f.) Conversely, if $g: U \to B_L$ and $g = \Sigma_{\mu \in M_n} v_\mu z(g_\mu)$, where the g_μ belong to $C^\infty[\epsilon(U), B_L]$, then g belongs to $G^\infty(U)$. (The z expansion is useful in defining a natural choice of partial derivative G_{j+m}, $j = 1, \ldots, n$).

Outline of proof: (a) This is proved essentially by "differentiation from first principles". (b) This follows directly from the properties of Taylor series. (c) Let $(\mathbf{a},\mathbf{b}) \in W$; choose $(\mathbf{c},\mathbf{d}) \in U$ such that $\epsilon(\mathbf{a},\mathbf{b}) = \epsilon(\mathbf{c},\mathbf{d})$.

Define $f': W \to B_L$ by

$$f'(\mathbf{a},\mathbf{b}) := \sum_{p=0}^{L} \frac{1}{p!} [D^p f(\mathbf{c},\mathbf{d})] [(\mathbf{a}-\mathbf{c}, \mathbf{b}-\mathbf{d})^p] . \tag{2.17}$$

This map is well defined because if $(\mathbf{e}, \mathbf{f})$ is another point in U such that $\epsilon(\mathbf{a},\mathbf{b}) = \epsilon(\mathbf{e}, \mathbf{f})$, then

$$\begin{aligned}
&\sum_{p=0}^{L} \frac{1}{p!} (D^p f)(\mathbf{e}, \mathbf{f})[(\mathbf{a}-\mathbf{e}, \mathbf{b}-\mathbf{f})]^p \\
&= \sum_{p=0}^{L} \sum_{p=0}^{L} \frac{1}{p!} \frac{1}{q!} [D^{p+q} f(\mathbf{c},\mathbf{d})][(\mathbf{e}-\mathbf{c}, \mathbf{f}-\mathbf{d})^q (\mathbf{a}-\mathbf{e}, \mathbf{b}-\mathbf{f})^p] \\
&= \sum_{p=0}^{L} \sum_{p=0}^{L} \left[\binom{p+q}{q} / (p+q)!\right] [D^{p+q} f(\mathbf{c},\mathbf{d})] \, [(\mathbf{e}-\mathbf{c}, \mathbf{f}-\mathbf{d})^q (\mathbf{a}-\mathbf{e}, \mathbf{b}-\mathbf{f})^p] \\
&= \sum_{N=0}^{L} \sum_{q=0}^{N} \left[\binom{N}{q} / N!\right] [D^N f(\mathbf{c},\mathbf{d})] \, [(\mathbf{e}-\mathbf{c}, \mathbf{f}-\mathbf{d})^q (\mathbf{a}-\mathbf{e}, \mathbf{b}-\mathbf{f})^{N-q}] \\
&= \sum_{N=0}^{L} \frac{1}{N!} [D^N f(\mathbf{c},\mathbf{d})] \, [(\mathbf{a}-\mathbf{c}, \mathbf{b}-\mathbf{d})^N] \quad [\text{using symmetry properties of } D^N f(\mathbf{c},\mathbf{d})] \\
&= f'(\mathbf{a},\mathbf{b}).
\end{aligned}$$

and thus $f'|_U = f$.

If f'' is another element of $G^\infty(W)$ such that $f''|_U = f$, then (by Corollary 2.9)

$$f''(\mathbf{a},\mathbf{b}) = \sum_{q=0}^{L} \frac{1}{q!} D^q f'(\mathbf{c},\mathbf{d}) \, [(\mathbf{a}-\mathbf{c}, \mathbf{b}-\mathbf{d})^q] = f'(\mathbf{a},\mathbf{b})$$

Direct calculation shows that $f' \in G^\infty(W)$. If

$$(\mathbf{a},\mathbf{b}) \in U, f'(\mathbf{a},\mathbf{b}) = \sum_{p=0}^{L} \frac{1}{p!} [D^p f(\mathbf{a},\mathbf{b})] \, [(\mathbf{0},\mathbf{0})^p] f(a,b) ,$$

and thus $f' = f''$, so that f' is unique.

(d) Choose $W = \epsilon^{-1}[\epsilon(U)] \cap B_L^{m,n}$ and, given $f \in G^\infty(U)$, let f' be the unique element of $G^\infty(W)$ such that $f'|_U = f$. [Such an f' exists by part (c) above.] Then, given

$$(\mathbf{a},\mathbf{b}) \in U, f(\mathbf{a},\mathbf{b}) = \sum_{p=1}^{n} \frac{1}{p!} D^p f'(\mathbf{a},\mathbf{0})[(\mathbf{0},\mathbf{b})^p]$$
$$= \sum_{\mu \in M_n} b_\mu \left[G_{\mu_k+m} \ldots G_{\mu_1+m} f'(\mathbf{a},\mathbf{0})\right].$$

Define ρ: $G^\infty(U) \rightarrow C^\infty[\epsilon(U), B_L]$ by [given $g \in G^\infty(U)$] $\rho(g)(x_1,\ldots,x_m) := g'(x_1 1,\ldots,x_m 1)$ for all $(x_1,\ldots,x_m)$ in $\epsilon(U)$, where g' is the continuation of g to W. Then $z[\rho(g)](\mathbf{a},\mathbf{b}) = g'(\mathbf{a},\mathbf{0})$ and thus

$$f(\mathbf{a},\mathbf{b}) = \sum_{\mu \in M_n} b_\mu z\left[\rho(G_{\mu_1+m} \ldots G_{\mu_k+m} f')\right](\mathbf{a},\mathbf{b}).$$

Now let

$$f_\Omega = \rho(f') \text{ and } f_\mu = \pi_L \circ \rho(G_{\mu_1+M} \ldots G_{\mu_k+M} f') \tag{2.18}$$

for all μ in M_n, where π_L is the projection of B_L onto the subspace spanned by $\{\beta_\mu | \mu \in M_L, \mu \neq (1,2,3,\ldots,L)\}$. Then $f(\mathbf{a},\mathbf{b}) = \Sigma_{\mu \in M_n} b_\mu z(f_\mu)(\mathbf{a},\mathbf{b})$, and finally $f = \Sigma_{\mu \in M_n} v_\mu z(f_\mu)$, where the f_μ are uniquely determined by Eq. (2.18). The converse is easily proved. □

Although the preceding proposition is not valid when $L = \infty$, indeed the z map is not defined in this case, a similar proposition, with entire functions instead of G^∞ and C^∞, holds for all values of L; this may be proved using Lemma 2.7.

The result of proposition 2.11(d) can be used to specify a natural choice of G_k in the cases where ambiguities arise, i.e., when L is finite and $m+1 \leqslant k \leqslant m+n$.

Let $f \in G^\infty(U)$ and $f = \Sigma_{\mu \in M_n} v_\mu z(f_\mu)$ be the z expansion of f. Given an integer q with $m+1 \leqslant q \leqslant m+n$, define

$$[q](v_\mu) := v_{\mu_2} \ldots v_{\mu_k}, \quad \text{if } q = \mu_1 + m,$$
$$[q](v_\mu) := (-1)^{s+1} v_{\mu_1} \ldots v_{\mu_{s-1}} v_{\mu_{s+1}} \ldots v_{\mu_k}, \quad \text{if } q = \mu_s + m, \tag{2.19}$$
$$[q](v_\mu) := 0, \quad \text{otherwise.}$$

Then a natural choice of $G_k f$, where $m+1 \leqslant k \leqslant n+m$, is

$$G_k f := \sum_{\mu \in M_n} [k](v_\mu) z(f_\mu). \tag{2.20}$$

All future references to G_k will be to this particular G_k; it is evident that $G_k f \in G^\infty(U)$ if $f \in G^\infty(U)$. Although it is satisfactory to have eliminated the ambiguity in the G_k, it should be emphasized that the ambiguity was unimportant because any two possible choices of G_k would give the same result when substituted into an expression of the form (2.5).

At this stage the definition of a G^∞ function can be compared with DeWitt's definition of a differentiable function[3]. DeWitt works with the real part of a complex Grassman algebra, rather than a real Grassman algebra, but this is not a significant difference. The topology DeWitt uses on $B_L^{m,n}$ (L finite) is a coarser topology than the usual topology, and is even non-Hausdorff. A subset U of $B_L^{m,n}$ is open in the DeWitt topology if and only if $U = \epsilon^{-1}(V)$ for some V open in $\mathbb{R}^m$. If U is open in $B_L^{m,n}$ with the DeWitt topology (and thus also open with the usual topology) and L is finite, the class of functions of U into B_L defined by DeWitt as "differentiable" does in fact coincide with $G^\infty(U)$ (so that it is the different topology on $B_L^{m,n}$, rather than the use of a different class of functions, which distinguishes a G^∞ supermanifold from a DeWitt supermanifold).

The infinite-dimensional algebra used by DeWitt is not B_∞, but rather the algebra generated by a countably infinite number of odd generators, which will be denoted W_∞. DeWitt effectively defines a differentiable function mapping an open set in $W_\infty^{m,n}$ (with the DeWitt topology) into W_∞ to be a function which can be expressed in terms of a z expansion (which now becomes a formal power series); this approach avoids some of the difficulties that might be raised by the absence of a norm on W_∞.

Many of the usual properties of C^∞ functions have direct analogies in the G^∞ framework; but the G^∞ structure is richer in that it inherits a grading from the grading on B_L. Proposition 2.12 contains the graded form of many of the classical properties of C^∞ functions.

Proposition 2.12: Let U be open in $B_L^{m,n}$, $f, g \in G^\infty(U)$, $a \in B_L$, and $\lambda \in \mathbb{R}$. Then, (a) $f + g \in G^\infty(U)$ and $G_k(f+g) = G_k f + G_k g$ $(k = 1,\ldots,m+n)$. (b) $\lambda f \in G^\infty(U)$ and $G_k(\lambda f) = \lambda G_k f$ $(k = 1,\ldots,m+n)$. (c) If E, Q represent projection maps of B_L onto the even and odd parts of B_L, repectively, then $E \circ f$ and $Q \circ f$ belong to $G^\infty(U)$. (d) $G^\infty(U)$ is a graded vector space with grading defined by $G^\infty(U)_0 := \{f | f \in G^\infty(U), f(U) \subset B_{L,0}\}$ and $G^\infty(U)_1 := \{f | f \in G^\infty(U), f(U) \subset B_{L,1}\}$.
(e) $af \in G^\infty(U)$ with $G_i(af) = aG_i f$ $(i = 1,\ldots,m)$, and $G_{j+m}(af) = (-1)^{|a|} a G_{j+m} f$ $(j = 1,\ldots,n)$. (f) $fg \in G^\infty(U)$ with $G_i(fg) = (G_i f)g + f G_i g$ $(i = 1,\ldots,m)$, and $G_{j+m}(fg) = (G_{j+m} f)g + (-1)^{|f|} f G_{j+m} g$ $(j = 1,\ldots,n)$ (the graded Leibnitz property). Suppose also that V is open in $B_L^{m',n'}$, and h is a G^∞ function of V into $B_L^{m,n}$. Let $H_l := p_l \circ h$. Then (g) (chain rule) $f \circ h \in G^\infty[h^{-1}(U) \cap V]$, and

$$G_k(f \circ h)(\mathrm{a}) = \sum_{l=1}^{m+n} (G_k h_l)(\mathrm{a})(G_l f)[h(\mathrm{a})]$$

for all $\mathrm{a} \in h^{-1}(U) \cap V$, $k = 1,\ldots,m'+n'$. Finally suppose that I is open in $\mathbb{R}$ and $h' \in C^\infty(I, B_L^{m,n})$. Then, (h) (another chain rule) $f \circ h' \in C^\infty[h'^{-1}(U) \cap I, B_L^{m,n}]$ and

$$\frac{\partial}{\partial t}(f \circ h') = \sum_{k=1}^{m+n} \frac{\partial g_k(t)}{\partial t} (G_k f)[g(t)], \quad \text{for } t \text{ in } I.$$

Proof: (a), (b), (e), (f), (g), and (h) may be proved very much as in the classical case; it is here that the Banach algebra property of B_L is essential. (c) It is easily shown that $E \circ f$ and $Q \circ f$ are G^1 with $G_i(E \circ f) = E \circ (G_i f)$, $G_i(Q \circ f) = Q \circ (G_i f)$ $(i = 1,\ldots,m)$ and $G_{j+m}(E \circ f) = Q \circ (G_{j+m} f)$, $G_{j+m}(Q \circ f) = E \circ (G_{j+m} f)$ $(j = 1,\ldots,n)$, and thus by induction $E \circ f$ and $Q \circ f$ are G^∞.
(d) It follows from (a) and (b) that $G^\infty(U)$ is a vector space. Also $f = E \circ f + Q \circ f$, and $E \circ f \in G^\infty(U)_0$, $Q \circ f \in G^\infty(U)_1$. Thus, $G^\infty(U) = G^\infty(U)_0 \oplus G^\infty(U)_1$. □

This very long section concludes with the definition of a useful subset of $G^\infty(U)$. The definition appears a rather arbitrary one at this stage, but is useful in comparing G^∞ supermanifolds with the supermanifolds and graded manifolds of other authors, particularly in Sec. 4.

Definition 2.13: Suppose U is open in $B_L^{m,n}$, and L is finite. The space $C^\infty[\epsilon(U)]$ of C^∞ maps of $\epsilon(U)$ into $\mathbb{R}$ may be identified with a subset of $C^\infty[\epsilon(U),B_L]$ by the identification map

$$i: C^\infty[\epsilon(U)] \to C^\infty[\epsilon(U),B_L], i(f)(p) := f(p)1, \quad \text{for all } p \text{ in } U.$$

Define $H^\infty(U): \{f \mid f: U \to B_L$, there exists $f_\mu \in C^\infty[\epsilon(U)]$ such that $f = \Sigma_{\mu \in M_L} z \circ i(f_\mu) v_\mu\}$.

It follows from proposition 2.11(d) that $H^\infty(U)$ is a subset of $G^\infty(U)$; clearly, there will always be many functions in $G^\infty(U)$ which are not in $H^\infty(U)$, and this is one of the reasons why a G^∞ supermanifold is a broader concept than a Kostant graded manifold.

3. DEFINITION OF A G^∞ SUPERMANIFOLD

Definition 3.1, which defines a G^∞ supermanifold, is modeled on a standard definition of a C^∞ manifold.[10] Just as an m-dimensional C^∞ manifold looks like $\mathbb{R}^m$ locally and has local coordinates $[x_1(p),...,x_m(p)]$ in $\mathbb{R}^m$, an (m,n)-dimensional G^∞ supermanifold over B_L looks like $B_L^{m,n}$ locally and has local coordinates $[u_1(p),...,u_m(p), v_1(p),...,v_n(p)]$ in $B_L^{m,n}$. As the definition uses the concept of the G^∞ function, which is the natural mathematical form for a superfield, a G^∞ supermanifold is an ideal vehicle for superfields. Both the structure of a supermanifold, and the classes of functions which can be defined on one, depend crucially on the nature of the transition functions. It is here that the analysis developed in the preceding section is used.

Definition 3.1: Let Y be a Hausdorff topological space. (a) An (m,n) open chart on Y over B_L is a pair (U,ψ) with U an open subset of Y and ψ a homeomorphism of U onto an open subset of $B_L^{m,n}$. (b) An (m,n) G^r structure on Y over B_L is a collection $\{(U_\alpha,\psi_\alpha) \mid \alpha \in \Lambda\}$ of open charts on Y such that (i) $Y = \cup_{\alpha\in\Lambda} U_\alpha$, (ii) for each pair α, β in Λ, the mapping $\psi_\beta \circ \psi_\alpha^{-1}$ is a G^r mapping of $\psi_\alpha(U_\alpha \cap U_\beta)$ onto $\psi_\beta(U_\alpha \cap U_\beta)$, (iii) the collection $\{(U_\alpha,\psi_\alpha) \mid \alpha\in\Lambda\}$ is a maximal collection of open charts for which (i) and (ii) hold. [A collection of open charts satisfying (i) and (ii) is called an (m,n) G^r subatlas on Y over B_L.] (c) An (m,n)-dimensional G^r supermanifold over $B_L^{m,n}$, is a Hausdorff topological space Y with an $(m,n)G^r$ structure on Y over B_L. (d) Each U_α is called a coordinate neighborhood, and each ψ_α is a coordinate map. For each $\alpha\in\Lambda$, $m + n$ local coordinate functions are defined by

$$u_i := p_i \circ \psi_\alpha, \quad i = 1,...,m,$$
$$v_j := p_{j+m} \circ \psi_\alpha, \quad j = 1,...,n. \tag{3.1}$$

In the rest of this paper attention is confined to the case where $r = \infty$. It will be useful in Sec. 4 to define an H^∞ supermanifold over B_L (where L is finite) by repeating definition 3.1 with H^∞ everywhere replacing G^r. An H^∞ supermanifold can always be given a G^∞ structure and thus made into a G^∞ supermanifold, but the converese is not true.

Some examples of G^∞ supermanifolds are now given: (a) $B_L^{m,n}$ is itself an (m,n)-dimensional G^∞ supermanifold over B_L. (b) Any open subset of $B_L^{m,n}$ is an (m,n)-dimensional G^∞ supermanifold. (c) The two-dimensional torus T^2 can be given the structure of a (1,1)-dimensional G^∞ supermanifold over B_1 as follows:

Let $\{1,\beta\}$ be a basis of B_1 with $\beta^2 = 0$. T^2 is defined to be I_2/R, where I_2 is the unit square $\{(x,y) \mid (x,y)\in\mathbb{R}^2, 0\leqslant x\leqslant 1, 0\leqslant y \leqslant 1\}$ with the usual topology, and R is the usual equivalence relation, defined by $(x,y) \sim (x',y')$ if and only if one of the following is true:

$$\begin{aligned}
&\text{(i) } (x,y) = (x',y'),\\
&\text{(ii) } x = x' \text{ and } |y - y'| = 1,\\
&\text{(iii) } y = y' \text{ and } |x - x'| = 1,\\
&\text{(iv) } |x - x'| = |y - y'| = 1.
\end{aligned} \tag{3.2}$$

Let $\overline{(x,y)}$ denote the equivalence class of (x,y).

A collection of (1,1) open charts $\{(U_\alpha,\psi_\alpha) \mid \alpha\in\{1,2,3,4\}\}$ on T^2 over B_1 is defined by

$$U_1 = \{\overline{(x,y)} \mid \tfrac{1}{5} < x < \tfrac{4}{5}, \tfrac{1}{5} < y < \tfrac{4}{5}\},$$
$$\psi_1[\overline{(x,y)}] = (x1, y\beta),$$

$$U_2 = \{\overline{(x,y)} \mid \tfrac{1}{5} < x < \tfrac{4}{5}, y < \tfrac{2}{5}\} \cup \{\overline{(x,y)} \mid \tfrac{1}{5} < x < \tfrac{4}{5}, y > \tfrac{3}{5}\},$$
$$\psi_2[\overline{(x,y)}] = (x1, y\beta), \quad \text{when } y < \tfrac{2}{5},$$
$$\psi_2[\overline{(x,y)}] = [x1, (y-1)\beta], \quad \text{when } y > \tfrac{3}{5},$$

$$U_3 = \{\overline{(x,y)} \mid x < \tfrac{2}{5}, \tfrac{1}{5} < y < \tfrac{4}{5}\} \cup \{\overline{(x,y)} \mid x > \tfrac{3}{5}, \tfrac{1}{5} < y < \tfrac{4}{5}\},$$
$$\psi_3[\overline{(x,y)}] = (x1, y\beta), \quad \text{when } x < \tfrac{2}{5},$$
$$\psi_3[\overline{(x,y)}] = [(x-1)1, y\beta], \quad \text{when } x > \tfrac{3}{5},$$

$$U_4 = \{\overline{(x,y)} \mid x < \tfrac{2}{5}, y < \tfrac{2}{5}\} \cup \{\overline{(x,y)} \mid x < \tfrac{2}{5}, y > \tfrac{3}{5}\} \cup \{\overline{(x,y)} \mid x > \tfrac{3}{5}, y < \tfrac{2}{5}\} \cup \{\overline{(x,y)} \mid x > \tfrac{3}{5}, y > \tfrac{3}{5}\},$$
$$\psi_4[\overline{(x,y)}] = (x1, y\beta), \quad \text{when } x < \tfrac{2}{5} \text{ and } y < \tfrac{2}{5},$$
$$\psi_4[\overline{(x,y)}] = [x1, (y-1)\beta], \quad \text{when } x < \tfrac{2}{5} \text{ and } y > \tfrac{3}{5},$$
$$\psi_4[\overline{(x,y)}] = [(x-1)1, y\beta], \quad \text{when } x > \tfrac{3}{5} \text{ and } y < \tfrac{2}{5},$$
$$\psi_4[\overline{(x,y)}] = [(x-1)1, (y-1)\beta], \quad \text{when } x > \tfrac{3}{5} \text{ and } y > \tfrac{3}{5}.$$

This collection of open charts is a $(1,1)G^\infty$ subatlas on T^2 over B_1. A typical transition function is

$$\psi_2 \circ \psi_1^{-1}: \{(\lambda 1, \mu\beta) \mid \tfrac{1}{5} < \lambda < \tfrac{4}{5}, \tfrac{1}{5} < \mu < \tfrac{2}{5}\} \cup \{(\lambda 1, \mu\beta) \mid \tfrac{1}{5} < \lambda < \tfrac{4}{5}, \tfrac{3}{5} < \mu < \tfrac{4}{5}\}$$
$$\to \{(\lambda 1, \mu\beta) \mid \tfrac{1}{5} < \lambda < \tfrac{4}{5}, \tfrac{1}{5} < \mu < \tfrac{2}{5}\} \cup \{(\lambda 1, \mu\beta) \mid \tfrac{1}{5} < \lambda < \tfrac{4}{5}, -\tfrac{2}{5} < \mu < -\tfrac{1}{5}\},$$
$$\psi_2 \circ \psi_1^{-1}(a,b) = (a,b), \quad \text{when } (a,b) \in \{(\lambda 1, \mu\beta) \mid \tfrac{1}{5} < \lambda < \tfrac{4}{5}, \tfrac{1}{5} < \mu < \tfrac{2}{5}\},$$
$$\psi_2 \circ \psi_1^{-1}(a,b) = (a, b - \beta), \quad \text{when } (a,b) \in \{(\lambda 1, \mu\beta) \mid \tfrac{1}{5} < \lambda < \tfrac{4}{5}, \tfrac{3}{5} < \mu < \tfrac{4}{5}\}.$$

This function is evidently G^∞. The subatlas can be extended to form a G^∞ structure on T^2, which then becomes a (1,1)-dimensional G^∞ supermanifold over B_1.

This last example shows that a G^∞ supermanifold can be contructed which both is compact and involves patching in the anticommuting sector. In contrast, neither of these is possible for a "differentiable supermanifold" as defined by DeWitt.[3] His definition may be summarized as follows (in the notation of this paper): Given a set M, (a) an (m,n) chart on M over B_L (L finite) or W_∞ is a pair (U,ψ) with U a subset of M and ψ an injective mapping of U into an open set (in the DeWitt topology) in $B_L^{m,n}$ (or $W_\infty^{m,n}$); (b) an (m,n) DeWitt differentiable structure on M over B_L (or W_∞) is a collection $\{(U_\alpha,\psi_\alpha) \mid \alpha \in \Lambda\}$ of (m,n) charts on M over B_L (or W_∞) such that

(i) $\bigcup_{\alpha\in\Lambda} U_\alpha = M$,
(ii) $\forall \alpha,\beta \in \Lambda$, $\psi_\alpha \circ \psi_\beta^{-1}$ is differentiable (in the sense defined by DeWitt),
(iii) the collection of charts is a maximal collection satisfying (i) and (ii).
(c) an (m,n)-dimensional DeWitt supermanifold M over B_L (or W_∞) is a set M together with an (m,n) DeWitt differentiable structure on M over B_L (or W_∞).

A DeWitt supermanifold M over B_L, where L is finite, is also a C^∞ manifold [of dimension $2^L(m+n)$] and the topology of M is defined by DeWitt to be the topology of M qua C^∞ manifold. The crucial difference between the definitions of a DeWitt supermanifold and a G^∞ supermanifold is the different topology used on $B_L^{m,n}$. Because DeWitt's topology is coarser, any (m,n)-dimensional DeWitt supermanifold can always be given an $(m,n)G^\infty$ structure, but the converse is not true—there are G^∞ supermanifolds (such as the torus example above) which cannot be given a DeWitt differentiable structure. It is a consequence of proposition 3.4 that any (m,n)-dimensional DeWitt supermanifold is noncompact and contractible to an m-dimensional C^∞ manifold, but this is certainly not the case with the torus example.

Because W_∞ is not a Banach space, a DeWitt supermanifold over W_∞ is not a C^∞ manifold; the topology of such a supermanifold is not defined by DeWitt. It is possible to define various topologies on W_∞, such as the non-Hausdorff topology of DeWitt, or the standard topologies on the ring $C[[X]]$ of formal power series,[11] and then define the topology of the supermanifold to be the topology it inherits from W_∞. The disadvantage of this approach, as opposed to that taken in the present paper using B_∞, is that W_∞ is not a Banach algebra, and thus any notion of differentiation on $W_\infty^{m,n}$ must be much more involved than that of definition 2.5.

DeWitt makes the interesting observation that with every (m,n)-dimensional DeWitt supermanifold M one can associate an m-dimensional C^∞ manifold which he calls "the body of M". His construction (which resembles Batchelor's construction of the underlying manifold of a Batchelor supermanifold[4]) is essentially to define on M the equivalence relation S as follows: Suppose $\{(U_\alpha,\psi_\alpha) \mid \alpha \in \Lambda\}$ is the DeWitt differentiable structure on M. Then, given x,y in M, $x \sim y$ if any only if there exist α,β in Λ such that x belongs to U_α, y belongs to U_β, and

$$\psi_1^{-1}\{\epsilon^{-1}[\epsilon \circ \psi_\alpha(x)] \cap B_L^{m,n}\} = \psi_\beta^{-1}\{\epsilon^{-1}[\epsilon \circ \psi_\beta(x)] \cap B_L^{m,n}\}. \tag{3.3}$$

M_B, the body of M, is then defined to be M/S. Given x in M, let $\bar{x}$ denote the equivalence class of x. The C^∞ structure on M_B is $\{(\bar{U}_\alpha,\bar{\psi}_\alpha) \mid \alpha \in \Lambda'\}$, where

$$\bar{U}_\alpha := \{\bar{x} \mid x \in U_\alpha\} \text{ and } \bar{\psi}_\alpha(\bar{x}) := \epsilon \circ \psi_\alpha(x). \tag{3.4}$$

DeWitt proves that, if $x \in U_\alpha \cap U_\beta$,

$$\psi_\alpha^{-1}\{\epsilon^{-1}[\epsilon \circ \psi_\alpha(x)] \cap B_L^{m,n}\} = \psi_\beta^{-1}\{\epsilon^{-1}[\epsilon \circ \psi_\beta(x)] \cap B_L^{m,n}\}.$$

Thus, $\bar{\psi}_\alpha$ is well defined because if $x \sim y$ and x belongs to U_α, then y belongs to $\psi_\alpha^{-1}\{\epsilon^{-1}[\epsilon \circ \psi_\alpha(x)] \cap B_L^{m,n}\}$. Thus, y belongs to U_α and $\epsilon \circ \psi_\alpha(y) = \epsilon \circ \psi_\alpha(x)$. [This argument also shows that S can be more simply expressed by $x \sim y$ if and only if there exists α in Λ such that x and y belong to U_α and $\epsilon \circ \psi_\alpha(x) = \epsilon \circ \psi_\alpha(y)$.]

A similar process applied to an (m,n)-dimensional G^∞ supermanifold also yields an m-dimensional C^∞ manifold; but an important difference between a G^∞ supermanifold and a DeWitt supermanifold is that a DeWitt supermanifold always has the structure of a vector bundle over its body as is proved in proposition 3.4 while there exist G^∞ supermanifolds (such as the torus example above) where this is not the case. As a result, a DeWitt supermanifold, despite being a $2^{L-1}(m+n)$-dimensional topological manifold, is topologically little more interesting than its body which is an m-dimensional manifold; a much less restricted class of topologies is possible for G^∞ supermanifolds.

Proposition 3.2: If M is an (m,n)-dimensional DeWitt supermanifold over B_L, the triple $(M,\pi,M/S)$ (where S is the equivalence relation defined above and π is the natural projection map from M onto M/S) can be given the structure of a $[2^{L-1}(m+n) - m]$-dimensional vector bundle.

Outline of proof: It can be shown that $\pi^{-1}(\bar{x}) = \psi_\alpha^{-1}\{\epsilon^{-1}[\epsilon \circ \psi_\alpha(x)] \cap B_L^{m,n}\}$. Let $(1,\gamma_1,\ldots,\gamma_{2^{L-1}-1})$ be a basis of $B_{L,0}$ and $(\delta_1,\delta_2,\ldots,\delta_{2^{L-1}})$ be a basis of $B_{L,1}$. Defind $h_x : \pi^{-1}(\bar{x}) \to R^{2^{L-1}(m+n)-m}$ by

$$h_x\Big(\psi_\alpha^{-1}\Big\{\epsilon[p_1 \circ \psi_\alpha(x)]1 + \sum_{k=1}^{2^{L-1}} b_{1,k}^1 \gamma_k, \ldots, + \epsilon[p_{m+n} \circ \psi_\alpha(x)]1 + \sum_{k=1}^{2^{L-1}} b_{m+n,k}\delta_k\Big\}\Big) = (b_{1,1,\ldots,} b_{m+n,2^{L-1}}).$$

h_x is evidently a homeomorphism; also, it is easily proved that $\pi^{-1}(\bar{U}_\alpha)$ is homeomorphic to $\bar{U}_\alpha \times R^{2^{L-1}(m+n)-m}$ for all $\alpha \in \Lambda$; and that the necessary continuity conditions are satisfied. □

Another definition of a supermanifold is given by Batchelor.[4] Her definition again superficially resembles that of a G^∞ supermanifold, and of a DeWitt supermanifold, over B_L where L is finite, but differs crucially (a) in that the ψ_α are required to be homeomorphisms of open subsets U_α of the supermanifold onto $\psi_\alpha(U_\alpha)$, where $\psi_\alpha(U_\alpha)$ is open in $B_L^{m,n}$ with the non-Hausdorff DeWitt topology (as in DeWitt's definition), and the topology on $\psi_\alpha(U_\alpha)$ is the restriction to $\psi_\alpha(U_\alpha)$ of the DeWitt topology on $B_L^{m,n}$ (whereas DeWitt uses the usual topology); also, (b) in that the transition functions $\psi_\alpha \circ \psi_\beta^{-1}$ are required to be "smooth" in a sense which is defined by a rather involved process that appears to be equivalent to requiring them to be H^∞. It is possible to define a finer topology on a Batchelor supermanifold S by specifying a subset U of S to be open if and only if $\psi_\alpha(U \cap U_\alpha)$ is open in $B_L^{m,n}$ (with the usual topology) for all α in Λ. A Batchelor supermanifold with this topology is evidently both a DeWitt differentiable supermanifold and an H^∞ supermanifold; the rather restricted definition of smooth used by Batchelor is motivated by her comparison with the graded manifolds of Kostant,[5] which is discussed in Sec. 4 of the present paper. (Batchelor observes that a less restricted definition is possible, which appears to be equivalent to requiring the function to be G^∞, so that her alternative definition is equivalent to DeWitt's. Future references to Batchelor supermanifolds will be to the more restricted definition.)

Because of the coarser topology used on $B_L^{m,n}$ by DeWitt and Batchelor, and the more restricted definition of smooth used by Batchelor, the class of G^∞ supermanifolds is a wider class than either of these other two classes, and embraces both of them. In the next section it is shown that the graded manifold formalism of Kostant[5] may also be subsumed withing the G^∞ supermanifold formalism.

4. GRADED MANIFOLDS AND G^∞ SUPERMANIFOLDS

The definition of a graded manifold given by Kostant[5], and the related definition of supermanifold given independently by Berezin and Leĭtes,[6] at first sight bears little resemblance to the definition of a G^∞ supermanifold. However, it will be shown that there is a connection.

An ordinary C^∞ manifold X has defined on it the sheaf of commutative algebras C^∞ [i.e., with each U open in X, one associates the commutative algebra $C^\infty(U)$]; a graded manifold of dimension (m,n) is defined by Kostant[5] to be a pair (X,A) where X an m-dimensional C^∞ manifold and A a sheaf of graded commutative algebras over X with specified properties—briefly, X has a covering of open sets $X = \cup_{\alpha \in \Lambda} U_\alpha$ such that $A(U_\alpha) \simeq C^\infty(U_\alpha) \otimes B_n$ for all α in Λ. Thus, this definition extends the definition of an ordinary manifold by extending the properties of the algebra $C^\infty(X)$ rather than the properties of the point set X itself.

The set of graded manifolds of dimension (m,n) can be identified with a subset of the set of (m,n)-dimensional G^∞ supermanifolds over B_L (where L is a fixed finite integer not less than n) in the following sense: Given an (m,n)-dimensional Kostant graded manifold (X,A), it is possible to construct an (m,n)-dimensional G^∞ supermanifold Y over B_L (which is in fact an H^∞ supermanifold) such that (a) $Y_B = X$, and (b) the sheaf A is isomorphic to the sheaf $H^\infty \circ \pi^{-1}$ (where π is the projection map from Y onto Y_B, and, given V open in Y,

$$H^\infty(V): \{f \mid f: V \to B_L, f \circ \psi_\alpha^{-1} \in H^\infty[\psi_\alpha(V \cap U_\alpha)] \forall \alpha \in \Lambda\}). \tag{4.1}$$

The construction is described in detail in the Appendix. It should be compared to the interesting work of Batchelor,[4] who proves that the category of Batchelor supermanifold (briefly described in Sec. 3 of the present paper) is equivalent to the category of Kostant graded manifold. Batchelor's work seems thus to implicity contain the construction given in the Appendix, but the approach taken here is somewhat simpler and more direct than Batchelor's and may readily be used when assessing the suitability of the graded manifold formalism for applications to supersymmetry (see Sec. 6).

Suppose $Y(X,A)$ is the G^∞ supermanifold constructed from a Kostant graded manifold (X,A). Then, $Y(X,A)$ is a restricted type of G^∞ supermanifold in two senses: It has a subatlas $\{(U_\alpha, \psi_\alpha) \mid \alpha \in \chi\}$ (a) with all transition functions H^∞, not merely G^∞, and (b) with all the sets $\psi_\alpha(U_\alpha)$, α in χ, open subsets of $B_L^{m,n}$ in the DeWitt topology, not merely in the usual topology. Thus, $Y(X,A)$ has more limited topological possibilities than an arbitrary G^∞ supermanifold. This is consistent with Kostant's proof that the de Rham cohomology of (X,A) is isomorphic to the Cech cohomology

5. VECTOR FIELDS ON A G^∞ SUPERMANIFOLD

Having constructed a G^∞ supermanifold, one can build on it local structure very much as on a C^∞ manifold. As an example, vector fields are considered in this section. Just as a vector field on an m-dimensional C^∞ manifold can be expressed locally as $\Sigma_{i=1}^m f_i(\mathbf{x})(\partial/\partial x_i)$, where the f_i are C^∞ functions, a vector field on an (m,n)-dimensional G^∞ supermanifold can be expressed locally as $\Sigma_{i=1}^m f_i(\mathbf{u},\mathbf{v})(\partial/\partial u_i) + \Sigma_{j=1}^n f_{m+j}(\mathbf{u},\mathbf{v})(\partial/\partial v_j)$, where the f_i and f_{m+j} are G^∞ functions. It is here that the formalism ties in with the local formulation of differential geometry on superspace.[2]

Suppose that Y is an (m,n)-dimensional G^∞ supermanifold with atlas $\{(U_\alpha, \psi_\alpha) \mid \alpha \in \Lambda\}$ and subatlas $\{(U_\alpha, \psi_\alpha) \mid \alpha \in \chi\}$ (with of course $\chi \subset \Lambda$). If U is an open set in Y, then a set of functions of U into B_L denoted $G^\infty(U)$ may be defined [corresponding to the definition of $C^\infty(V)$ when V is an open subset of a C^∞ manifold].

Definition 5.1: (a) Given U an open subset of Y,

$$G^\infty(u) := \{f \mid f: U \to B_L,$$

with

$$f \circ \psi_\alpha^{-1} \in G^\infty[\psi_\alpha(U \cap U_\alpha)] \forall \alpha \in \Lambda\}. \tag{5.1}$$

(b) Given p belonging to Y, $G^\infty(p) := \{f \mid$ there exists an open neighborhood N of p such that $f \in G^\infty(N)\}$. (5.2)

Exactly as in the classical case, a sufficient condition for f to belong to $G^\infty(U)$ is that $f \circ \psi_\alpha^{-1}$ belong to $G^\infty[\psi_\alpha(U \cap U_\alpha)]$ for all α in the subatlas index set χ. If V is an open subset of a C^∞ manifold, $C^\infty(V)$ is a commutative algebra; the analogous result for a G^∞ supermanifold is as follows.

Proposition 5.2: Given U open in Y, (a) $G^\infty(U)$ is a graded commutative algebra over $\mathbb{R}$, with

$$G^\infty(U)_0 := \{ f \mid f \in G^\infty(U), f(U) \subset B_{L,0} \},$$
$$G^\infty(U)_1 := \{ f \mid f \in G^\infty(U), f(U) \subset B_{L,1} \}. \quad (5.3)$$

(b) $G^\infty(U)$ is a graded left B_L module.

Outline of proof: The results follow directly from proposition 2.12. □

The algebras $G^\infty(U)$ are the equivalent in the G^∞ supermanifold formalism of the algebras $A(U)$ in the Kostant graded manifold formalism,[5] of the algebras $M^\infty(U, B_L)$ of smooth functions in the Batchelor supermanifold formalism, and of the algebras $F(U)$ of differentiable functions in the DeWitt supermanifold formalism. Only the algebras $G^\infty(U)$ and $F(U)$ have a graded left B_L-modules structure as well as a graded algebra structure.

A G^∞ function is the natural form for a superfield, as is explained in the next section; Taylor's theorem (corollary 2.9) leads easily to the conventional superfield expansion in powers of θ.

Let End $[G^\infty(U)]$ denote the set of vector space endomorphisms of $G^\infty(U)$. Then, End $[G^\infty(U)]$ is a graded algebra (over $\mathbb{R}$), with an element α of End $G^\infty(U)$ belonging to End $[G^\infty(U)]_0$ if $|\alpha(f)| = |f|$ for all homogeneous f in $G^\infty(U)$, and belonging to End $[G^\infty(U)]_1$ if $|\alpha(f)| = |f| + 1 \pmod 2$.

A vector field on an open subset V of a C^∞ manifold may be defined as a derivation of $C^\infty(V)$. Motivated by this, and Kostant's use of derivations of the graded algebras $A(U)$,[5] the natural definition of a vector field on an open set U in Y is a derivation of $G^\infty(U)$, with an extra condition relating to the left B_L-module structure of $G^\infty(U)$ (for which there is no analog in the C^∞ or graded manifold structures).

Definition 5.3: Let U be open in Y. A *vector field* on U is an element X of End $[G^\infty(U)]$ such that

$$\text{(a)} \quad X(fg) = (Xf)g + (-1)^{|f||X|} fXg, \quad \text{for all } f, g \text{ in } G^\infty(U),$$
$$\text{(b)} \quad X(af) = (-1)^{|X||a|} aXf, \quad \text{for all } f \text{ in } G^\infty(U),\ a \text{ in } B_L. \quad (5.4)$$

The set of vector fields on U is denoted $D^1(U)$.

As is to be expected, the vector fields defined here give zero when they act on a constant map. $D^1(U)$ is both a graded left B_L module and a graded Lie algebra (over the real numbers). These two structures interplay in a manner which gives $D^1(U)$ the structure of a "graded Lie left B_L module", defined as follows:

Definition 5.4: A *graded Lie left* B_L *module* is a graded Lie algebra W (over the real numbers) which is also a graded left B_L module such that

$$[aX_1, X_2] = a[X_1, X_2], \quad \text{for all } a \text{ in } B_L,\ X_1, X_2 \text{ in } W. \quad (5.5)$$

Proposition 5.5: $D^1(U)$ is a graded Lie left B_L module, with bracket operation defined by

$$[X_1, X_2] := X_1X_2 - (-1)^{|X_1||X_2|} X_2X_1. \quad (5.6)$$

Proof: Let the grading on $D^1(U)$ be defined by

$$D^1(U)_p := D^1(U) \cap \text{End}[G^\infty(U)]_p, \quad \text{for } p = 0,1.$$

It is a standard result that, with the above bracket operation and grading, $D^1(U)$ forms a graded Lie algebra (see Kostant[5]). Let X belong to $D^1(U)$ and a belong to B_L. Then,

$$\begin{aligned} aX(fg) &= a(Xf)g + (-1)^{|X||f|} a fXg \\ &= (aXf)g + (-1)^{(|X| + |a|)|f|} faXg, \quad \text{for all } f, g \text{ in } G^\infty(U). \end{aligned}$$

Also, $aX(bg) = (-1)^{(|X| + |a|)} baXg$, for all g in $G^\infty(U)$, b in B_L. Thus, aX belongs to $D^1(U)$ and $|aX| = |a| + |X|$; hence, $D^1(U)$ is a graded left B_L module. Finally,

$$\begin{aligned} [aX_1, X_2] &= aX_1X_2 f - (-1)^{|aX_1||X_2|} X_2 aX_1 f \\ &= aX^1X_2 f - (-1)^{|aX_1||X_2| + |a||X_2|} aX_2X_1 f \\ &= a[X_1, X_2] f, \quad \text{for all } f \text{ in } G^\infty(U). \end{aligned}$$

Thus, $D^1(U)$ is a graded Lie left B_L module. □

Coordinate derivatives can be defined on a G^∞ supermanifold in the same way as on a C^∞ manifold.

Definition 5.6: Let (U, ψ) be a chart on Y. Define, for $i = 1, ..., m$,

$$\frac{\partial}{\partial u_i} : G^\infty(U) \to \{\text{functions of } U \text{ into } B_L\},$$
$$\frac{\partial f}{\partial u_i} := [G_i(f \circ \psi^{-1})] \circ \psi \text{ for all } f \text{ in } G^\infty(U); \quad (5.7)$$

Also, for $j = 1, ..., n$,

$$\frac{\partial}{\partial v_j} : G^\infty(U) \to \{\text{functions of } U \text{ into } B_L\},$$
$$\frac{\partial f}{\partial v_j} := [G_{j+m}(f \circ \psi^{-1})] \circ \psi, \quad \text{for all } f \text{ in } G^\infty(U). \quad (5.8)$$

The next two propositions contain useful and important properties of the $\partial / \partial u_i$ and $\partial / \partial v_j$. These correspond closely to the properties of coordinate derivatives on C^∞ manifolds, and to theorem 2.8 of Kostant (Ref. 5, p. 197) which establishes that, given a graded manifold (X, A) of dimension (m,n), the set of derivations of each algebra $A(U)$ is a free $A(U)$ module with a basis consisting of m even and n odd elements.

Proposition 5.7: $\partial/\partial u_i$ belongs to $D^1(U)_0$ for $i = 1, ..., m$, and $\partial/\partial v_j$ belongs to $D^1(U)_1$ for $j = 1, ..., n$.

Proof: Let f, g belong to $G^\infty(U)$. Then, $\partial f/\partial u_i \circ \psi^{-1} = G_i(f \circ \psi^{-1})$, which belongs to $G^\infty[\psi(U)]$. Thus $\partial f/\partial u_i \in G^\infty(U)$. Similarly, $\partial f/\partial v_j \in G^\infty(U)$. It follows from proposition 2.12 that $\partial/\partial u_i$ and $\partial/\partial v_j$ belong to End $[G^\infty(U)]$. Suppose f belongs to $G^\infty(U)_p$, where $p = 0$ or 1. Then, $f \circ \psi^{-1}$ belongs to $G^\infty[\psi(U)]_p$ and $G_i(f \circ \psi^{-1})$ belongs to $G^\infty[\psi(U)]_p$ for $i = 1, ..., m$, $G_{j+m}(f \circ \psi^{-1})$ belongs to $G^\infty[\psi(U)]_{p+1 \pmod 2}$ for $j = 1, ..., n$. Hence, $|\partial f/\partial u_i| = |f|$, $i = 1, ..., m$ and $|\partial f/\partial v_j| = |f| + 1 \pmod 2$, $j = 1, ..., n$. Thus, $\partial/\partial u_i$ belongs to End $[G^\infty(U)]_0$ and $\partial/\partial v_j$ belongs to End $[G^\infty(U)]_1$. Suppose a belongs to B_L. Then, by Proposition 2.12(e),

$$\frac{\partial a f}{\partial u_i} = G_i(a f \circ \psi^{-1}) \circ \psi = a(G_i f \circ \psi^{-1}) \circ \psi$$
$$= (-1)^{|a||\partial/\partial u_i|} \frac{\partial f}{\partial u_i} \tag{5.9}$$

and similarly

$$\frac{\partial a f}{\partial v_j} = (-1)^{|a||\partial/\partial v_j|} \frac{\partial f}{\partial v_j}. \tag{5.10}$$

Finally, by proposition 2.12 (f),

$$\frac{\partial (fg)}{\partial u_i} = G_i[(f \circ \psi^{-1})(g \circ \psi^{-1})] \circ \psi$$
$$= \{[G_i(f \circ \psi^{-1})](g \circ \psi^{-1}) + (f \circ \psi^{-1})[G_i(g \circ \psi^{-1})]\} \circ \psi$$
$$= \frac{\partial f}{\partial u_i} g + (-1)^{|\partial/\partial u_i||f|} f \frac{\partial g}{\partial u_i}. \tag{5.11}$$

Thus, $\partial/\partial u_i$ belongs to $D^1(U)$. It can be shown in a similar manner that $\partial/\partial v_j$ belongs to $D^1(U)$. □

Having established that coordinate derivatives are vector fields, it can be shown that they form a basis for $D^1(U)$ [regarded as a free left $G^\infty(U)$ module], in complete analogy with the classical case.

Proposition 5.8: (a) $D^1(U)$ is a graded left $G^\infty(U)$ module. (b) If (U,ψ) is a coordiante chart on Y, $D^1(U)$ is a free left $G^\infty(U)$ module with basis $\{(\partial/\partial u_i)|i=1,...,m\} \cup \{(\partial/\partial v_j)|j=1,...,n\}$.

Outline of proof: (a) Given f in $G^\infty(U)$, X in $D^1(U)$, define

$$fX : G^\infty(U) \to \{\text{functions of } U \text{ into } B_L\}$$

by

$$fX(g) := fXg \text{ for all } g \text{ in } G^\infty(U). \tag{5.12}$$

It can be verified by straightforward calculation that fX belongs to $D^1(U)$ and $|fX| = |f| + |X|$. Thus, $D^1(U)$ is a graded left $G^\infty(U)$ module. (b) A straightforward extension of the proof of the classical case[10] establishes the result. □

Tangent vectors can be defined on a G^∞ supermanifold exactly as on a C^∞ manifold.

Definition 5.9: Let p belong to Y and X belong to $D^1(Y)$. Define

$$X_p : G^\infty(U) \to B_L, X_p f := Xf(p), \text{ for all } f \text{ in } G^\infty(p). \tag{5.13}$$

The set $\{X_p | X \in D^1(Y)\}$ is called the tangent module at p and denoted Y_p.

It is an easy consequence of proposition 5.8 that Y_p is a free graded left B_L module with basis

$$\left\{\left(\frac{\partial}{\partial u_i}\right)_p \Big| i=1,...,m\right\} \cup \left\{\left(\frac{\partial}{\partial v_j}\right)_p \Big| j=1,...,n\right\},$$

where the $\partial/\partial u_i$ and $\partial/\partial v_j$ are defined with respect to any coordinate chart containing p.

Suppose that Z is an H^∞ supermanifold and that U is an open subset of Z. It is easily seen that the set of derivations of $H^\infty(U)$ is a graded left $H^\infty(U)$ module [free, with basis $\{(\partial/\partial u_i)|i=1,...,m\} \cup \{(\partial/\partial v_j)|j=1,...,n\}$ in the case where (U,ψ) is an H^∞ open chart], and also a graded Lie algebra (but not a graded left B_L module). This should be compared with Kostant's result[5] that [given (X,A) an (m,n)-dimensional graded manifold, U open in X, and $\mathrm{Der}(U) := \{\text{derivations of } A(U)\}$] $\mathrm{Der}(U)$ is a graded left $A(U)$ module (free in the case where U is an odd and even coordinate neighborhood) and a graded Lie algebra. Such a similarity is to be expected in the light of the results of Sec. 4.

In view of the structure of vector fields on G^∞ supermanifolds, it seems clear that if one defines a super Lie group to be a group G which is also a G^∞ supermanifold with group operations which are G^∞, then the set of left invariant vector fields will be a free graded Lie left B_L module of dimension (m,n); and that conversely given any free graded Lie left B_L module g, of dimension (m,n), there will exist a nonempty set of (locally isomorphic) super Lie groups of dimension (m,n) each of whose set of left invariant vector fields is isomorphic to the given module g. An example of this is the group obtained from the (4,4)-dimensional graded supersymmetry algebra with generators $\{P_\mu, S_\alpha\}$ by exponentiation using the Hausdorff formula. There may be other, locally isomorphic but globally nonisomorphic, super Lie groups each of whose set of left invariant vector fields is isomorphic to this algebra.

In this section attention has been concentrated on vector fields. It is clear that many other constructions which are conventionally made on C^∞ manifolds, such as forms, can also be made on G^∞ supermanifolds.

6. POTENTIAL APPLICATIONS TO SUPERGRAVITY

The definition of a G^∞ supermanifold was motivated by the definition of superspace given by Salam and Strathdee[1] and subsequently taken up by many authors. An obvious global definition of superspace, locally equivalent to the local definition of Salam and Strathdee, is to define superspace to be a (4,4)-dimensional G^∞ supermanifold. Local coordinates will then be of the form $(x_1, x_2, x_3, x_4, \theta_1, \theta_2, \theta_3, \theta_4,)$. In curved superspace, as opposed to the homogeneous superspace of rigid supersymmetry, the θ's are anticommuting coordinates but not Lorentz spinors. The action of the Lorentz group is defined on the tangent module at each point, and it is the odd part of each tangent vector which is a spinor.

In the heuristic local formalism, a superfield is defined as a function Φ mapping superspace into the even part of the Grassman algebra (or sometimes simply into the Grassman algebra) with

$$\Phi(x_1,x_2,x_3,x_4,\theta_1,\theta_2,\theta_3,\theta_4) = \sum_{\mu \in M_4} \phi_4(x_1,x_2,x_3,x_4)\theta_\mu, \tag{6.1}$$

where implicitly the $\phi_\mu(\mathbf{x})$ are differentiable functions in some unspecified sense, and are required to reduce to ordinary fields when restricted to the "body" of superspace.

Corresponding to the global definition of superspace as a (4,4)-dimensional G^∞ supermanifold Y, it is natural to define a superfield to be an element of $G^\infty(Y)_0$ [or simply of $G^\infty(Y)$]. Let Φ be a superfield on Y and (U,ψ) be an open chart on Y. Then, $\Phi \circ \psi^{-1}$ is G^∞ on $\psi(U)$ and the z expansion of $\Phi \circ \psi^{-1}$ (cf. Proposition 2.11),

TABLE I. Global definitions of supermanifold.

		Type of definition	Topology on $B_L m,n$	Transition functions	Nature of "superfield"	Possible values of L	Topology	Equivalent to	Contained in
(A)	Graded manifold (Kostant, Berezin and Leĭtes)	Manifold + sheaf (or bundle)	...	(H^∞)	H^∞	...	...	(B)	(C),(D),(E)
(B)	Supermanifold (Batchelor)	Set with atlas	Coarse	H^∞	H^∞	Finite	Non-Hausdorff	(A)	(C),(D),(E)
(C)	Super bundle (Smolin)	Manifold + vector bundle	...	(H^∞)	G^∞	...	...	...	(E)
(D)	Supermanifold (DeWitt)	Set with atlas	Coarse	G^∞	G^∞	Finite or infinite	Noncompact vector bundle over body	...	(E)
(E)	G^∞ Supermanifold	Set with atlas	Fine	G^∞	G^∞	Finite or infinite	Hausdorff; may be compact or noncompact	...	...

$$\Phi\circ\psi^{-1} = \sum_{\mu\in M_4} v_\mu z(f_\mu) \tag{6.2}$$

(where the f_μ belong to $C^\infty[\epsilon\circ\psi(U), B_L]$), is the equivalent of the usual superfield expansion (6.1). There is an apparent inconsistency in the expansion (6.1): The superfield is required to be infinitely differentiable in the even part but analytic in the odd part, and it is not evident that this property will be preserved under supersymmetry trnasformations which mix up the odd and even parts. However, as this "mixing up" only involves nilpotent elements, and Taylor's theorem (corrollary 2.9) shows that the distinction between infinitely differentiable and analytic disappears for nilpotent elements, there is in fact no inconsistency.

In most work on superspace the dimension of the Grassman algebra used is unspecified; clearly, L must be at least as great as the odd dimension n if the superfield expansion is not to have trivial terms. If L is finite, then no product

$$\prod_{i=1}^{N} \Phi_i(\mathbf{x}^{(i)}, \theta^{(i)}),$$

where the Φ_i are superfields with

$$\Phi_i(\mathbf{x}^{(i)}, \theta^{(i)}) = \sum_{\mu_i\in M_4} \phi_{i,\mu_i}(\mathbf{x}^{(i)})\theta^{(i)}_{\mu_i},$$

can contain terms with more than L factors $\phi_{i,\mu_i}(\mathbf{x}^{(i)})$ with none of the μ_i equal to Ω. This would seem to place an undesirable restriction on N-point Green's functions, which can only be lifted by using an infinite-dimensional "Grassman" algebra such as B_∞.

It is evident that the graded manifold formalism (or the equivalent supermanifold formalisms such as that developed by Batchelor[4] or by the author in the Appendix to the present paper) is inappropriate for applications to supersymmetry because the coefficients $\phi_{i,\mu}$ in the superfield expansion (6.1) would all be commuting, which excludes the possibility of fermions. An alternative approach to applying Kostant's graded manifold formalism is given in a recent paper by Dell and Smolin.[7] They make use of Batchelor's theorem[12] that, given a Kostant graded manifold (X,A), it is possible to find a vector bundle E over X such that $[X,\Gamma(\Lambda E)]$ is isomorphic to (X,A) (where $\Gamma(\Lambda E)$ represents the sheaf of cross sections of the exterior bundle ΛE associated with E), and consider the exterior bundle of the spin bundle over a four-dimensional space–time. The approach neatly incorporates the spinorial character of the θ-s but still has the drawback that anticommuting classical fields are excluded.

In a second paper,[8] again motivated by Batchelor's theorem, Smolin constructs a further vector bundle X from E (denoted sup E) which has cross sections which may be expanded locally in the form (6.1) (with ϕ_μ even when the sequence μ has an even number of elements and odd otherwise). This appears to be equivalent to regarding superspace as a G^∞ supermanifold of the type which can be constructed from a Kostant graded manifold by the construction given in the Appendix, and a superfield as an even G^∞ function.

However, there is no obvious physical reason why superspace should be restricted to being a supermanifold of this type, or to being a DeWitt supermanifold. On the contrary, it seems very desirable to consider the full class of G^∞ supermanifolds, admitting as it does the possibility of nontrivial topology in the fermion sector.

7. SUMMARY

A mathematically rigorous definition of supermanifold has been developed in this paper. Local constructions on G^∞ supermanifolds agree with the local differential geometry on superspace of other authors.[2] The definition of a G^∞ supermanifold includes other "global" definitions of supermanifold in a way which is described in detail in Sec. 4 and 5, and is summarized in Table I. It also includes possibilities not allowed for in any other global formalism, particularly in making possible patching and nontrivial topology in the anticommuting sector.

The G^∞ supermanifold formalism can be applied naturally to superspace and supersymmetry. It shows that the heuristic formulation of superspace and superfields can be made rigorous, and how superspace can be enabled to have a variety of global topologies.

ACKNOWLEDGMENT

I am extremely grateful to Chris Isham for many helpful discussions and suggestions, and for critically reading the manuscript.

APPENDIX

Given a Kostant graded manifold[5] (X,A), a G^∞ supermanifold Y is constructed such that (a) $Y_B = X$; (b) the sheaf A is isomorphic to the sheaf $H^\infty \circ \pi^{-1}$ (c.f. Sec. 4, and Batchelor's proof[6] that the category of Kostant graded manifold is equivalent to the category of Batchelor supermanifold).

Using the notation and terminology of Kostant, let (X,A) be an (m,n)-dimensional graded manifold, and L be an integer with $n \leqslant L < \infty$. Let $\cup_{\alpha\in A} X_\alpha = X$ be an open covering of X by odd and even coordinate neighborhoods. For each $\alpha\in A$, let $r_i^\alpha (i=1,\ldots,m)$ $s_j^\alpha (j=1,\ldots,n)$ be odd and even coordinates on X_α, and $C(X_\alpha)$ be the function factor corresponding to the coordinates $\{r_i^\alpha\}$. Also, given

$$\alpha, \beta\in A, \text{ let } r_i^{\alpha\beta} := \rho_{X_\alpha, X_\alpha\cap X_\beta} r_i^\alpha, \quad r_i^{\beta\alpha} := \rho_{X_\beta, X_\alpha\cap X_\beta} r_i^\beta, \tag{A1}$$

with $s_j^{\alpha\beta}$ and $s_j^{\beta\alpha}$ defined similarly. Also, given $U \subset X_\alpha$, let $C_\alpha(U)$ be the function factor on U corresponding to the even coordinates $\{\rho_{X_\alpha,U} r_i^\alpha | i=1,\ldots,m\}$.

Now there exist unique

$$P_{i\mu}^{\alpha\beta}\in C_\alpha(X_\alpha\cap X_\beta), \quad i=1,\ldots,m, \mu\in M_n,$$

and (A2)

$$Q_{j\mu}^{\alpha\beta}\in C_\alpha(X_\alpha\cap X_\beta), \quad j=1,\ldots,n, \mu\in M_n,$$

such that
$$r_i^{\beta\alpha} = \sum_{\mu\in M_n} P_{i\mu}^{\alpha\beta} s_\mu^{\alpha\beta}, \quad i=1,\ldots,m$$

(A3)

$$\text{and } s_j^{\beta\alpha} = \sum_{\mu\in M_n} Q_{j\mu}^{\alpha\beta} s_\mu^{\alpha\beta}, \quad j=1,\ldots,n.$$

Given $\alpha\in A$, let ϕ_α be the coordinate map on X corresponding to $\{r_i^\alpha\}$, i.e., $\phi_\alpha \in C^\infty(X_\alpha,\mathbb{R}^m)$ and, for $i=1,\ldots,m$, $p_i\circ\phi_\alpha = \tilde{r}_i\alpha$. Given $\alpha,\beta\in A$,
let $S_{\alpha\beta} : \epsilon^{-1}[\phi_\alpha(X_\alpha\cap X_\beta)]\cap B_L^{m,n}$.

A set of functions will now be defined which are of great importance in the construction of Y, and will eventually be shown to be transition functions on Y.

Definition A.1: Given $\alpha,\beta\in A$, let $\tau_{\beta\alpha} : S_{\alpha\beta} \to S_{\beta\alpha}$ be defined by

$$p_i\circ\tau_{\beta\alpha}(\mathbf{a},\mathbf{b}) := \sum_{\mu\in M_n} z(\tilde{P}_{i\mu}^{\alpha\beta}\circ\phi_\alpha^{-1})(\mathbf{a})p_\mu(\mathbf{b}), \quad i=1,\ldots,m,$$
(A4)
$$p_{j+m}\circ\tau_{\beta\alpha}(\mathbf{a},\mathbf{b}) := \sum_{\mu\in M_n} z(\tilde{Q}_{j\mu}^{\alpha\beta}\circ\phi_\alpha^{-1})(\mathbf{a})p_\mu(\mathbf{b}), \quad j=1,\ldots,n.$$

Note that $\tau_{\beta\alpha}$ is defined on $S_{\alpha\beta}$ because ϕ_α^{-1} is defined on $\phi_\alpha(X_\alpha\cap X_\beta)$ and $\tilde{P}_{i\mu}^{\alpha\beta}$ and $\tilde{Q}_{j\mu}^{\alpha\beta}$ are defined on $X_\alpha\cap X_\beta$. Also, $\epsilon\circ\tau_{\beta\alpha} = \phi_\beta\circ\phi_\alpha^{-1}\circ\epsilon$, and thus $\tau_{\beta\alpha}(S_{\alpha\beta})\subset S_{\beta\alpha}$. It will be shown below that $\tau_{\beta\alpha}$ is in fact a homeomorphism of $S_{\alpha\beta}$ onto $S_{\beta\alpha}$.

Having defined the $\tau_{\beta\alpha}$, the next step is to show (in proposition A.3) that these functions have the necessary properties of transition functions; the proof of this proposition requires the following lemma, which is the cornerstone of the entire construction:

Lemma A.2: Suppose U is open in X, $f\in C_\alpha(U\cap X_\beta\cap X_\alpha)$, and $(\mathbf{a},\mathbf{b})\in\epsilon^{-1}[\phi^\alpha(U\cap X_\alpha\cap X_\beta)]\cap B_L^{m,h}$ (where h is an integer, $0\leqslant h\leqslant n$). Then

$$z(\tilde{f}\circ\phi_\beta^{-1})\Big[\sum_{\nu\in M_h} z(\tilde{P}_{1\nu}^{\alpha\beta}\circ\phi_\alpha^{-1})(a_1,\ldots,a_m)p_\nu(b_1,\ldots,b_h),\ldots,$$
$$\sum_{\nu\in M_h} z(\tilde{P}_{m\nu}^{\alpha\beta}\circ\phi_\alpha^{-1})(a_1,\ldots,a_m)p_\nu(b_1,\ldots,b_h)\Big]$$
$$= \sum_{\nu\in M_h} z(\tilde{\partial}_\nu f\circ\phi_\alpha^{-1})(a_1,\ldots,a_m)p_\nu(b_1,\ldots,b_h), \tag{A5}$$

where

$$\partial_\nu := \frac{\partial}{\partial s_{\nu_k}^{\alpha'}}\cdots\frac{\partial}{\partial s_{\nu_1}^{\alpha'}} \text{ and } s_j^{\alpha'} := \rho_{X_\alpha, U\cap X_\alpha\cap X_\beta} s_j^\alpha.$$

This lemma may be proved by induction on h. Note that in the case where $h=n$ the lemma states that

$$z(\tilde{f}\circ\phi_\beta^{-1})\circ\tau_{\beta\alpha}(\mathbf{a},\mathbf{b}) = \sum_{\nu\in M_n} z(\tilde{\partial}_\nu f\circ\phi_\alpha^{-1})(\mathbf{a})p_\nu(\mathbf{b}). \tag{A6}$$

Proposition A.3: (a) Suppose $\alpha,\beta,\gamma,\in A$. Then $\tau_{\gamma\beta}\circ\tau_{\beta\alpha} = \tau_{\gamma\alpha}$, where both functions are defined, i.e., on $\tau_{\beta\alpha}^{-1}(S_{\beta\gamma})\cap S_{\alpha\gamma}$. (b) $\forall\,\alpha\in A$, $\tau_{\alpha\alpha}$ is the identity map on $S_{\alpha\alpha}$. (In future, $S_{\alpha\alpha} := \epsilon^{-1}[\phi_\alpha(X_\alpha)]\cap B_L^{m,n}$ will simply be denoted S_α.) (c) $\forall\,\alpha,\beta\in A$, $\tau_{\alpha\beta}\circ\tau_{\beta\alpha}$ is the identity function on $S_{\alpha\beta}$. (d) $\tau_{\alpha\beta}$ is a homeomorphism of $S_{\alpha\beta}$ and $S_{\beta\alpha}$.

Indication of proof: The proof of (a) is too long to be included here. The method involves showing that the interrelation between the $\tilde{P}_{i\mu}^{\alpha\beta}$, $\tilde{Q}_{j\mu}^{\alpha\beta}$, $\tilde{P}_{i\mu}^{\beta\gamma}$, $\tilde{Q}_{j\mu}^{\beta\gamma}$, $\tilde{P}_{i\mu}^{\alpha\gamma}$, $\tilde{Q}_{j\mu}^{\alpha\gamma}$, may be expressed in a form which after application of Lemma A.2 gives exactly the required relationship between $p_k\circ\tau_{\gamma\beta}\circ\tau_{\beta\alpha}$ and $p_k\circ\tau_{\gamma\alpha}$. (b) is proved quite simply after noting that

$$P_{i\Omega}^{\alpha\alpha} = r_i^\alpha, \quad P_{i\mu}^{\alpha\alpha} = 0 \quad (\mu\neq\Omega),$$
$$Q_{j(j)}^{\alpha\alpha} = 1_{A(X_\alpha)}, \quad Q_{j\mu}^{\alpha\alpha} = 0 \quad [\mu\neq(j)].$$

(c) is a straightfoward consequence of (a) and (b). (d) follows from (c) (which shows that $\tau_{\alpha\beta}$ and $\tau_{\beta\alpha}$ are inverses) and from the fact that $\tau_{\alpha\beta}$ and $\tau_{\beta\alpha}$ and G^∞ functions and thus a fortiori continuous. □

Having established the necessary properties of the $\tau_{\alpha\beta}$, the G^∞ supermanifold can now be constructed by a standard patching technique.[13]

For each α in A, let $Z_\alpha := \{\alpha\}\times S_\alpha$, and define $\Omega_\alpha : Z_\alpha \to S_\alpha$

$$\text{by } \Omega_\alpha[(\alpha,p)] := p, \quad \text{for all } p \text{ in } S_\alpha . \tag{A7}$$

Let $Z := \cup_{\alpha\in\Lambda} Z_\alpha$ and define the following relation (R) on Z: given z_1, z_2, in Z, let $z_1 R\, z_2$ if and only if $\tau_{\beta\alpha}\circ\Omega_\alpha(z_1) = \Omega_\beta(z_2)$ (where α is the unique element of Λ such that z_1 is in Z_α and β is the unique element of Λ such that z_2 is in Z_β). It can be proved (using proposition A.3) that R is an quivalence relation. The set Y is then defined to be Z/R.

Let $Y_\alpha := \{[z] | z\in Z_\alpha\}$ (where $[z]$ is the equivalence class of z), and define

$$\psi_\alpha : Y_\alpha \to S_\alpha \text{ by } \psi_\alpha([z]) := \Omega_\alpha(z)$$

where z is the unique element in $[z]$ which is an element of Z_α). Then, ψ_α is a bijective mapping and $\psi_\alpha(Y_\alpha) = S_\alpha$.

Also, if $[z]\in Y_\alpha \cap Y_\beta$, $\psi_\alpha([z]) = \Omega_\alpha(z)$ and $\psi_\beta([z]) = \Omega_\beta(z')$, where $z\in Z_\alpha$, $z'\in Z_\beta$ and zRz'. Thus, $\psi_\beta([z]) = \tau_{\beta\alpha}\circ\Omega_\alpha(z) = \tau_{\beta\alpha}\circ\psi_\alpha([z])$. Hence, $\psi_\beta\circ\psi_\alpha^{-1} = \tau_{\beta\alpha}$ and thus $\psi_\beta\circ\psi_\alpha^{-1}\in H^\infty[\psi_\alpha(Y_\alpha\cap Y_\beta)] \subset G^\infty[\psi_\alpha(Y_\alpha\cap Y_\beta)]$. Evidently, $Y = \cup_{\alpha\in\Lambda} Y_\alpha$ and thus $\{(Y_\alpha,\psi_\alpha)|\alpha\in\Lambda\}$ gives Y the structure of a topological manifold and of a G^∞ supermanifold.

Proposition A.4: Y_B is diffeomorphic to X.

Outline of proof: Recall that $Y_B = Y/S$ (c.f. Sec. 3). It can be shown that, for all $\alpha\in\Lambda$, the mapping $h_\alpha : [Y_\alpha] \to X_\alpha$, $h_\alpha([y]) := \phi_\alpha^{-1}\circ\epsilon\circ\psi_\alpha(y)$ defines a homeomorphism of $[Y_\alpha]$ and X_α. Also, $\bar\psi_\alpha([y]) = \phi_\alpha[h_\alpha([y])]$. □

Before defining and proving the isomorphism of the sheaves A and $H^\infty\circ\pi^{-1}$, it is necessary to introduce further notation. If U is an open subset of X_α, define $z_{\alpha(U)} : C^\infty(U) \to \{$functions on $\pi^{-1}(U)\}$ (where π is the projection map of Y onto Y/S, and Y/S and X are now identified by

$$z_{\alpha(U)}(f) := z(f\circ\phi_\alpha^{-1})\circ\psi_\alpha, \quad \text{for all } f \text{ in } C^\infty(U). \tag{A8}$$

Let $\mathbf{C}_\alpha^\infty[\pi^{-1}(U)] := z_\alpha[C^\infty(U)]$. The, $z_{\alpha(U)}$ defines an isomorphism of $C^\infty(U)$ and $\mathbf{C}_\alpha^\infty[\pi^{-1}(U)]$. (The $\mathbf{C}_\alpha^\infty[\pi^{-1}(U)]$ correspond to Kostant's various function factors.) Let

$$\phi_\alpha(U) := \rho_{X_\alpha,X_\alpha\cap U}(\phi_\alpha), \qquad \psi_{\alpha(U)} := \rho_{Y_\alpha,\pi^{-1}(X_\alpha\cap U)}(\psi_\alpha); \tag{A9}$$

$$u_i^{\alpha(U)} := p_i\circ\psi_{\alpha(U)}, \quad i = 1,\ldots,m, \qquad v_j^{\alpha(U)} := p_{j+m}\circ\psi_{\alpha(U)}, \quad j = 1,\ldots,n. \tag{A10}$$

Then, $H^\infty[\pi^{-1}(U)]$ is equal to the set of functions $f:\pi^{-1}(U)\to B_L$ such that, for all $\alpha\in\Lambda$, $\mu\in M_n$, there exist $f_\mu^{\alpha(U)}\in\mathbf{C}_\alpha^\infty[\pi^{-1}(U\cap X_\alpha)]$ with $\rho_{U,U\cap X_\alpha} f = \Sigma_{\mu\in M_n} f_\mu^{\alpha(U)} v_\mu^{\alpha(U)}$.

Now, $H^\infty\circ\pi^{-1}$ has the sheaf properties specified by Kostant (Ref. 8, p. 187), i.e., (a) suppose U, V are open in X and $U\subset V$. Define

$$\rho'_{U,V} : H^\infty\circ\pi^{-1}(U) \to H^\infty\circ\pi^{-1}(V)$$

$$\text{by} \quad \rho'_{U,V}(f) := f|_{\pi^{-1}(V)} . \tag{A11}$$

(b) If W is a further open set in X and $W\subset V$, $\rho'_{V,W}\circ\rho'_{V,W} = \rho'_{U,W}$. (c) If $U = \cup_{i\in\Gamma} U_i$ is an open covering of U, and f, $g\in H^\infty\circ\pi^{-1}(U)$, then $\rho'_{U,U_i} f = \rho'_{U,U_i} g$ for all $i\in\Gamma$ implies $f = g$. (d) If $h_i\in H^\infty\circ\pi^{-1}(U)$ is given for all i in Γ such that $\rho'_{U_i,U_i\cap U_j} h_i = \rho'_{U_j,U_i\cap U_j} h_j$ for all i,j in Γ, then there exists h in $H^\infty\circ\pi^{-1}(U)$ such that $h_i = \rho'_{U,U_i} h$ for all i in Γ. (These sheaf properties of $H^\infty\circ\pi$ are a direct consequence of the properties of the sheaf C^∞ over X.)

The isomorphism of the sheaves $H^\infty\circ\pi^{-1}$ and A will now be established: (a) Let U be open in X with $U\subset X_\alpha$. Define

$$K_U^\alpha : A(v) H^\infty\circ\pi^{-1}(U)$$

by

$$K_U^\alpha\Big(\sum_{\mu\in M_n} f_\mu^{\alpha(U)} s_\mu^{\alpha(U)}\Big) := \sum_{\mu\in M_n} z_\alpha(\tilde f_\mu^{\alpha(U)}) v_\mu^{\alpha(U)}\Big), \tag{A12}$$

where each $f_\mu^{\alpha(U)}\in C_\alpha(U)$ and $s_\mu^{\alpha(U)} := \rho_{X_\alpha,U} s_\mu^\alpha$. Then K_U^α is an isomorphism of $A(U)$ and $H^\infty\circ\pi^{-1}(U)$, since $L\geqslant n$. (b) Suppose $U\subset X_\alpha\cap X_\beta$. Then the isomorphisms K_U^α and K_U^β can be shown to be equal (using Lemma A.2) and both isomorphisms referred to unambiguously as K_U. (c) Suppose $V\subset U\subset X_\alpha$, with U and V open. Then it can easily be shown that $K_V\circ\rho_{U,V} = \rho'_{U,V}\circ K_U$. (d) Suppose U is an arbitrary open set in X. Define

$$K_U : A(U) \to H^\infty\circ\pi^{-1}(U)$$

by

$$\rho'_{U,U\cap X_\alpha} K_U f := K_{U\cap X_\alpha}\,\rho_{U,U\cap X_\alpha} f, \quad \text{for all } \alpha \text{ in } \Lambda. \tag{A13}$$

Using the sheaf properties of A and $H_\infty\circ\pi^{-1}$, it can be proved that K_U is well defined, and is an isomorphism of $A(U)$ and $H^\infty\circ\pi^{-1}(U)$; and also that if V is open in X and $V\subset U$, then $\rho'_{U,V}\circ K_U = K_V\circ\rho_{U,V}$.

Thus, the set $K := \{K_U | U \text{ open in } X\}$ defines an isomorphism of the sheaves A and $H^\infty\circ\pi^{-1}$.

[1]Abdus Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974).

[2]R. Arnowitt and Pran Nath, Gen. Relativity Gravitation **7**, 89 (1976); S. Bedding, S. Downes-Martin, and J.G. Taylor, "A Superfield analysis of supergravity," preprint (King's College, London, July 1978); Lars Brink, Murray Gell-Mann, Pierre Ramond, and John H. Schwarz, Phys. Lett. B **74**, 336 (1978); V.I. Ogievetsky and E. Sokatchev, "Supergravity as a theory of an axial superfield," talk presented at the seminar on quantum gravity, Moscow, December 1978; Warren Siegel and S. James Gates Jr., Nucl. Phys. B **147**, 77 (1979); J. Wess and B. Zumino, Phys. Lett. B **66**, 361 (1977).

[3]B. DeWitt, "Supermanifolds" (in preparation).

[4]Marjorie Batchelor, "Two approaches to supermanifolds," Trans. Am. Math. Soc. (to appear).

[5]B. Kostant, "Graded Manifolds, Graded Lie Theory, and Prequantization," in *Differential Geometric Methods in Mathematical Physics*, proceedings of the Symposium held at Bonn July 1975, Lecture Notes in Mathematics 570 (Springer, Berlin, 1977).

[6]F.A. Berezin and D.A. Leĭtes, Sov. Math. Dokl. **16**, 1218 (1975).

[7]John Dell and Lee Smolin, Commun. Math. Phys. **66**, 197 (1979).

[8]Lee Smolin, "Superspace unmasked: the physical interpretation of supergravity theory", preprint HUTP-78/A043 (Harvard, 1978).

[9]Abdus Salam and J. Strathdee, "Supersymmetry and superfields," preprint IC/76/12 (Trieste, 1976).

[10]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York and London, 1962).

[11]F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York and London, 1967).

[12]Marjorie Batchelor, "The Structure of Supermanifolds," Trans. Am. Math. Soc. (to appear).

[13]J. Dieudonné, *A Treatise on Analysis* (Academic, New York and London, 1970), Vol. II, Chap. XII.

Extended Numerov method for the numerical solution of the Hartree–Fock equations

Edward A. Burke
Department of Physics, Adelphi University, Garden City, New York 11530

(Received 7 August 1979; accepted for publication 19 October 1979)

To match the intrinsic accuracy of an expanded basis set solution of atomic orbitals by numerical orbitals requires integration formulas of comparable intrinsic accuracy. Integration of the coupled, integro-differential, Hartree–Fock equations is most easily done using Numerov's formula. To increase the accuracy of Numerov's formula without increasing the number of mesh points, one must extend the basic formula to include more points, Herein is presented the derivation of the general $(2N+1)$-point Numerov formula, including the error term, as well as suggestions for numerical determination with specific formulas of 3-, 5-, 7-, and 9-point fits.

I. INTRODUCTION

The Hartree–Fock method[1,2] for numerical atomic structure calculations requires the solution of coupled integrodifferential equations. The general form of the equations is

$$P''(r) + f(r)P(r) = g(r). \tag{1}$$

The most widely-used method[3] for the solution of Eq. (1) is the Numerov formula[4]:

$$\begin{aligned} &12P_{i-1} - 24P_i + 12P_{i+1} - h^2(P''_{i-1} + 10P''_i + P''_{i+1}) \\ &\quad = -\tfrac{1}{20}h^6 P_i^{VI} + \theta(h^8). \end{aligned} \tag{2}$$

In Eq. (2), $P_i = P(r)$ at $r = r_i$, $h = r_{i+1} - r_i$ and is constant for all i, and P_i^{VI} is the sixth derivative of $P(r)$ evaluated at $r = r_i$. As indicated in Eq. (2) the neglected terms are of the order of h^8. Thus, the accuracy of Numerov's formula is dependent on the magnitude of h.

Another popular method for atomic structure calculations is the expansion method introduced by Roothan.[5] Since this method does not require numerical integration, it is capable of a very high *intrinsic* accuracy. However, the choice of a basis set for expansion becomes prohibitively difficult for atoms of many electrons. Furthermore, the selection of a basis set introduces a bias into the solution which may conceal important, although small, orbital variations. It is therefore desirable to increase the intrinsic accuracy of the numerical atomic structure calculations.

With the wide variety of multiprecision packages available with computers today, it is not difficult, in principle, to increase accuracy. It is obvious, from Eq. (2), that simply by decreasing the mesh size, h, the order of error is reduced. By reducing h we require more points in our integration mesh. While this represents no large problem from the point of view of storage for present-day computers with their large banks of virtual storage, it does present a problem in the time required to perform the numerous iterations required in the solution.

Alternatively, the accuracy of the numerical calculations can be improved by extending the Numerov formula to include more points, thereby raising the order of and reducing the error term. The results of a 5-point fit have been given by Roothaan and Soukup.[6] In this paper we present the derivation of the extended Numerov method and present the formulae for a $(2N+1)$-point fit for $N = 2,3,4$.

II. DERIVATION

Combine the even, central-point difference formula,[7]

$$\delta^{2n} f_i^{(2p)} = \sum_{j=0}^{2n} (-1)^j \binom{2n}{j} f_{i+n-j}^{(2p)}, \tag{3}$$

with a Taylor series expansion about $x = x_i$:

$$f_{i+n-j}^{(2p)} = \sum_{k=0}^{\infty} \frac{h^k (n-j)^k f_i^{(2p+k)}}{k!}. \tag{4}$$

Thus

$$\delta^{2n} f_i^{(2p)} = \sum_{j=0}^{2n} \sum_{k=0}^{\infty} (-1)^j \binom{2n}{j} \frac{h^k (n-j)^k f_i^{(2p+k)}}{k!}. \tag{5}$$

or

$$\delta^{2n} f_i^{(2p)} = \sum_{k=0}^{\infty} \frac{h^k f_i^{(2p+k)}}{k!} \sum_{j=0}^{2n} (-1)^j \binom{2n}{j} (n-j)^k. \tag{6}$$

The following conditions on k will now be proved:

(1) k even only,
(2) $k \neq 0$,
(3) $k \geqslant n$.

Let

$$T_{n,k} = \sum_{j=0}^{2n} (-1)^j \binom{2n}{j} (n-j)^k. \tag{7}$$

Let $n - j \to j$, then

$$T_{n,k} = \sum_{j=-n}^{n} (-1)^{n-j} \binom{2n}{n-j} j^k. \tag{8}$$

Consider the sum, t_l, of the pair of terms, $j = \pm l$:

$$t_l = (-1)^{n-l} \binom{2n}{n-l} l^k + (-1)^{n+l} \binom{2n}{n+l} (-l)^k. \tag{9}$$

Now,

$$(-1)^{n-l} \equiv (-1)^{n+l}, \quad \binom{2n}{n-l} \equiv \binom{2n}{n+l},$$

and $\quad (-l)^k = (-1)^k l^k,$

 0022-2488/80/061366-04$1.00

hence

$$t_l = (-1)^{n-l}\binom{2n}{n-l} l^k[1+(-1)^k] = 0 \quad \text{if } k \text{ odd.}$$

Consider $k=0$:

$$T_{n,0} = \sum_{j=0}^{2n} (-1)^j \binom{2n}{j} \equiv (1-l)^{2n} = 0, \quad n \neq 0. \tag{10}$$

Let $S_k^n = T_{n,2k}$ in Eq. (7), giving

$$S_k^n = \sum_{j=0}^{2n} (-1)^j \binom{2n}{j} (n-j)^{2k}. \tag{11}$$

Now,

$$(n-j)^{2k} = \sum_{l=0}^{2k} \binom{2k}{l} n^{2k-l}(-1)^l j^l, \tag{12}$$

hence

$$S_k^n = \sum_{j=0}^{2n}\sum_{l=0}^{2k} (-1)^j \binom{2n}{j}\binom{2k}{l} n^{2k-l}(-1)^l j^l \tag{13}$$

or

$$S_k^n = \sum_{l=0}^{2k} \binom{2k}{l} n^{2k-l}(-1)^l \sum_{j=0}^{2n} (-1)^j \binom{2n}{j} j^l. \tag{14}$$

The Stirling numbers of the second kind[8] are

$$\mathcal{S}_n^m = \frac{1}{m!}\sum_{k=0}^{m} (-1)^{m-k}\binom{m}{k} k^n, \tag{15}$$

with the properties of importance here that $\mathcal{S}_n^m = 0$ for $m > n$, and $\mathcal{S}_n^n = 1$.

Hence

$$S_k^n = \sum_{l=0}^{2k} \binom{2k}{l} n^{2k-l}(-1)^l (2n)! \mathcal{S}_l^{2n} \tag{16}$$

We require $2n \leqslant l$, but $l \leqslant 2k \therefore 2n \leqslant 2k$ or $n \leqslant k$ Q.E.D.

Since $l \geqslant 2n$, Eq. (16) may be written as:

$$S_k^n = \sum_{l=2n}^{2k} \binom{2k}{l} n^{2k-l}(-1)^l (2n)! \mathcal{S}_l^{2n}, \tag{17}$$

hence

$$S_n^n = \binom{2n}{2n} n^{2n-2n}(-1)^{2n}(2n)! \mathcal{S}_{2n}^{2n} \equiv (2n)! \tag{18}$$

Incorporating conditions on k in Eq. (7), with the difinition $S_k^n = T_{n,2k}$ enables us to write

$$S_k^n = 2\sum_{j=1}^{n} (-1)^{n-j}\binom{2n}{n-j} j^{2k}. \tag{19}$$

Substituting Eq. (19) and conditions on k into Eq. (6) gives:

$$\delta^{2n} f_i^{(2p)} = \sum_{k=n}^{\infty} \frac{h^{2k} f_i^{[2(p+k)]}}{(2k)!} S_k^n. \tag{20}$$

Multiply Eq. (20) by h^{2p}, let $F_i^{(2p)} = h^{2p} f_i^{(2p)}$ and truncate the infinite series to form the $2N$ equations:

$$\delta^{2n} F_i^{(2p)} = \sum_{k=n}^{2N-p} \frac{F_i^{[2(p+k)]}}{(2k)!} S_k^n, \qquad n = 1,2,\ldots,N, \quad p = 0,1. \tag{21}$$

We now form a sum of equations (21) such that derivatives of fourth and higher order vanish.

Let

$$V = \sum_{n=1}^{N}\sum_{p=0}^{1} a_{np}\delta^{2n} F_i^{(2p)}, \tag{22a}$$

or

$$V = \sum_{n=1}^{N}\sum_{p=0}^{1}\sum_{k=n}^{2n-p} \frac{a_{np} S_k^n}{(2k)!} F_i^{[2(p+k)]}. \tag{22b}$$

Since $S_k^n = 0$ if $k < n$, we can equally well begin the sum over k at 0. Also perform the sum over p in (21) to get

$$V = \sum_{n=1}^{N}\left(a_{n0}\sum_{k=0}^{2N} \frac{S_k^n F_i^{(2k)}}{(2k)!} + a_{n1}\sum_{k=0}^{2N-1} \frac{S_k^n F_i^{(2k+2)}}{(2k)!}\right). \tag{23}$$

In the second sum let $k \rightarrow k-1$ so that

$$V = \sum_{n=1}^{N}\left(a_{n0}\sum_{k=0}^{2N} \frac{S_k^n F_i^{(2k)}}{(2k)!} + a_{n1}\sum_{k=1}^{2N} \frac{S_{k-1}^n F_i^{(2k)}}{(2k-2)!}\right). \tag{24}$$

Since the lowest value of n is 1, then the lowest value of k is 1, hence the first sum can equally well begin at $k=1$, allowing us to remove the k summation outside the n summation:

$$V = \sum_{k=1}^{2N} F_i^{(2k)}\left[\sum_{n=1}^{N}\left(\frac{a_{n0}S_k^n}{(2k)!} + \frac{a_{n1}S_{k-1}^n}{(2k-2)!}\right)\right]. \tag{25}$$

Replace the p summation in the large parentheses of (25):

$$V = \sum_{k=1}^{2N} F_i^{(2k)}\left(\sum_{n=1}^{N}\sum_{p=0}^{1} \frac{a_{np}S_{k-p}^n}{[2(k-p)]!}\right), \quad k-p \geqslant n \tag{26}$$

or

$$V = F_i'' \sum_{n=1}^{N}\sum_{p=0}^{1} \frac{a_{np}S_{1-p}^n}{[2(1-p)]!} + \sum_{k=2}^{2n} F_i^{(2h)} \times\left(\sum_{n=1}^{N}\sum_{p=0}^{1} \frac{a_{np}S_{k-p}^n}{[2(k-p)]!}\right), \quad k-p \geqslant n. \tag{27}$$

The condition, $1-p > n$, in the first term allows for the single possibility $n=1, p=0$. *Since* $S_1^1 = 2$, we get

$$V = a_{10}F_i'' + \sum_{k=2}^{2N} F_i^{(2k)}\left(\sum_{n=1}^{N}\sum_{p=0}^{1} \frac{a_{np}S_{k-p}^n}{[2(k-p)]!}\right), \qquad k-p \geqslant n. \tag{28}$$

Let $U = V/a_{10}$, $b_{np} = a_{np}/a_{10}$. Then

$$U = F_i'' + \sum_{k=2}^{2N} F_i^{(2k)} \sum_{n=1}^{N}\sum_{p=0}^{1} \frac{b_{np}S_{k-p}^n}{[2(k-p)]!}, \qquad k-p \geqslant n. \tag{29}$$

We now choose b_{np} such that the coefficient of each $F_i^{(2k)}$ in the sum over k is zero. Hence the conditions on the b_{np}'s are

$$\sum_{n=1}^{N}\sum_{p=0}^{1} \frac{b_{np}S_{k-p}^n}{[2(k-p)]!} = 0, \quad k = 2,3,\ldots,2N. \tag{30}$$

By definition $b_{10} = 1$ so there are $2N-1$ unknown b_{np}'s and Eqs. (29) are $2N-1$ equations so that the b_{np}'s are uniquely determined. Furthermore, from Eq. (29) this choice yields $U = F_i''$.

By the choice, $a_{10}=1$, $b_{np}=a_{np}$, $U=V=F_i''$. Combine Eqs. (22a), (22b), and the definition $F_i^{(2p)}=h^{2p}f_i^{(2p)}$ to get

$$F_i''=\sum_{n=1}^{N}\sum_{p=0}^{1}\sum_{j=0}^{2N}a_{np}(-1)^j\binom{2n}{j}F_{i+n-j}^{(2p)}. \qquad (31)$$

Let $n-j=l$, then have

$$F_i''=\sum_{n=1}^{N}\sum_{p=0}^{1}\sum_{l=-n}^{n}a_{np}(-1)^{n-l}\binom{2n}{n-l}F_{i+l}^{(2p)}. \qquad (32)$$

Now the binomial coefficient is zero if $n\pm l<0$, i.e., if $n<l<-n$. Hence one can equally well sum over l between $\pm N$ and take l summation outside,

$$F_i''=\sum_{l=-N}^{N}\sum_{p=0}^{1}F_{i+l}^{(2p)}\left[\sum_{n=1}^{N}a_{np}(-1)^{n-l}\binom{2n}{n-l}\right]. \qquad (33)$$

The term in square brackets is the coefficient of the $F_{i+l}^{(2p)}$ term in the Numerov procedure.

A. Error term

Let

$$c_{lp}=\sum_{n=1}^{N}a_{np}(-1)^{n-l}\binom{2n}{n-l},$$
$$l=-N,-N+1,\dots,N. \qquad (34)$$

Then rewrite Eq. (33) as

$$\sum_{p=0}^{1}\sum_{l=-N}^{N}c_{lp}F_{i+l}^{(2p)}-F_i''=\text{error term}$$

and by Eq. (22b) we find

$$\text{error term}=\sum_{n=1}^{N}\sum_{p=0}^{1}\frac{a_{np}S_{2N-p+1}^{n}}{[2(2N-p+1)]!}F_i^{(4N+2)}. \qquad (35)$$

Since $F_i^{(4N+2)}=h^{4N+2}f_i^{(4N+2)}$ then the error is calculated by taking the $4N+2$ derivative of $f(x)$ with respect to x and evaluate it at $x=x_i$ and not some arbitrary value of x within the range.

Since the next missing order is F_i^{4N+4} then, if the error correction is applied in a given calculation, the error is $O(h^{4N+4})$.

B. Numerical methods

Consider Eq. (30) with $b_{n,p}$ replaced by a_{np} and multiplied by $(4N)!$

$$\sum_{n=1}^{N}\sum_{p=0}^{1}\frac{4N!}{[2(k-p)]!}a_{np}S_{k-p}^{n}=0,\quad k=2,3,\dots,2N. \qquad (36)$$

Now the coefficients of the a_{np}'s are all integers. We may also write, for computational purposes, Eq. (36) as

$$\sum_{n=1}^{N}\sum_{p=0}^{1}\binom{4N}{2(k-p)}[4N-2(k-p)]!S_{k-p}^{n}a_{np}=0,$$
$$k=2,3,\dots,2N. \qquad (37)$$

The a_{np}'s may now be obtained as rational fractions. Let M be the least common denominator. Let $A_{np}=Ma_{np}$, then the A_{np}'s are all integers.

Consider Eq. (34). We note that $c_{lp}\equiv c_{-lp}$ so that we may rewrite Eq. (33) as

$$\sum_{p=0}^{1}\sum_{l=1}^{N}c_{lp}(F_{i+l}^{(2p)}+F_{i-l}^{(2p)})+\sum_{p=0}^{1}c_{0p}F_i^{(2p)}-F_i''$$
$$=\text{error term}. \qquad (38)$$

Replacing a_{np} by A_{np}/M in (34) gives

$$c_{lp}=\frac{1}{M}\sum_{n=1}^{N}A_{np}(-1)^{n-l}\binom{2n}{n-l}. \qquad (39)$$

Now let $c_{lp}=Mc_{lp}$ and put into (38):

$$\frac{1}{M}\sum_{p=0}^{1}\sum_{l=1}^{N}C_{lp}(F_{i+l}^{(2p)}+F_{i-l}^{(2p)})+\frac{1}{M}\sum_{p=0}^{1}C_{0p}F_i^{(2p)}-F_i''$$
$$=\text{error term}. \qquad (40)$$

Multiplying through by M leaves integer coefficients. Reorder terms, separating out second derivative terms, and include error term Eq. (35) for the final formula:

$$\sum_{l=1}^{N}C_{l0}(f_{i+l}+f_{i-l})+C_{00}f_i$$
$$-h^2\left(\sum_{l=1}^{N}-C_{l1}(f_{i+l}''+f_{i-l}'')+(M-C_{01})f_i''\right)$$
$$=\left(-\sum_{n=1}^{N}\sum_{p=0}^{1}\frac{A_{np}S_{2N-p+1}^{n}}{[2(2N-p+1)]!}\right)$$
$$\times h^{4N+2}f_i^{(4N+2)}. \qquad (41)$$

III. RESULTS

$N=1$, 3-point fit:

$$-24f_i+12(f_{i+1}+f_{i-1})-h^2[10f_i''+f_{i+1}+f_{i-1}]$$
$$=(-1/20)h^6f_i^{\mathrm{VI}}.$$

$N=2$, 5-point fit:

$$-4770f_i+1920(f_{i+1}+f_{i-1})+465(f_{i+2}+f_{i-2})$$
$$-h^2[2538f_i''+688(f_{i+1}''+f_{i-1}'')$$
$$+23(f_{i+2}''+f_{i-2}'')]$$
$$=(-79/1260)h^{10}f_i^{x}. \qquad (43)$$

$N=3$, 7-point fit:

$$-3252620f_i+790965(f_{i+1}+f_{i-1})$$
$$+785862(f_{i+2}+f_{i-2})$$
$$+49483(f_{i+3}+f_{i-3})-h^2[2175924f_i''$$
$$+989739(f_{i+1}''+f_{i-1}'')$$
$$+110322(f_{i+2}''+f_{i-2}'')+1857(f_{i+3}''+f_{i-3}'')]$$
$$=(-114669/400400)h^{14}f_i^{\mathrm{XIV}}. \qquad (44)$$

$N=4$, 9-point fit:

$$-64426815960f_i+4610152960(f_{i+1}+f_{i-1})$$
$$+23237683840(f_{i+2}+f_{i-2})$$
$$+4238517760(f_{i+3}+f_{i-3})+127053415(f_{i+4}+f_{i-4})$$
$$-h^2[57889160040f_i''$$
$$+33004678656(f_{i+1}''+f_{i-1}'')$$
$$+656546752(f_{i+2}''+f_{i-2}'')+390425088(f_{i+3}''+f_{i-3}'')$$
$$+3970884(f_{i+4}''+f_{i-4}'')]$$
$$=(-15048903/425425)h^{18}f_i^{\mathrm{XVIII}}. \qquad (45)$$

ACKNOWLEDGMENT

The author is deeply indebted to Professor Roothaan who innovated the idea of using an extended Numerov method, and encouraged and cooperated with the author in the preparation of this paper.

[1]D.R. Hartree, *The Calculation of Atomic Structures* (Wiley, New York, 1957).

[2]V. Fock, Z.f. Physik **61**, 126 (1930); **62**, 795 (1930).

[3]Charlotte Froese, Can. J. Phys. **41**, 1895 (1963).

[4]B. Numerov, Publ. de L'Observ. Astrophysique Central de Russie **2**, 188 (1933).

[5]C.C.J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951); **32**, 179 (1960).

[6]C.C.J. Roothaan and G.C. Soukup, Int. J. Quantum Chem. **15**, 449 (1979).

[7]D.R. Hartree, *Numerical Analysis* (Oxford University, London, 1958).

[8]C. Jordan, *Calculus of Finite Differences* (Chelsea, New York, 1950).

An application of Ray–Reid invariants

M. Lutzky
10111 Quinby St, Silver Spring, Maryland 20901

(Received 15 January 1980; accepted for publication 22 February 1980)

The formalism of Ray–Reid invariants is applied to a certain nonlinear system, first treated by Reid using a different approach. In addition, a generalization of the Ray–Reid invariant is given.

It has recently been shown by Ray and Reid[1] that invariants similar to the Lewis invariant for the time-dependent harmonic oscillator[2] exist also for more general systems; these new invariants may be written

$$I = \tfrac{1}{2}(x\dot{y} - y\dot{x})^2 + \int^{x/y} f(\lambda)\, d\lambda + \int^{y/x} g(\lambda)\, d\lambda, \qquad (1)$$

and satisfy $\dot{I} = 0$ if $x(t)$ and $y(t)$ satisfy

$$\ddot{y} + \omega^2(t)y = f(x/y)/(y^2x), \qquad (2a)$$

$$\ddot{x} + \omega^2(t)x = g(y/x)/(x^2y). \qquad (2b)$$

Thus I constitutes a conserved quantity for the dynamical system described by the coupled equations (2). When f and g are such that Eqs. (2) are uncoupled, algebraic considerations show that the invariant (1) yields a general solution of one equation in terms of particular solutions of the other[2,3]; however, this procedure is not effective if the equations are coupled. In this note we present an alternative approach, valid for partially coupled systems, and use it to show that a certain class of nonlinear dynamical systems may be treated by means of the invariant (1), thus verifying a recent conjecture of Ray and Reid.[4] In addition, we will exhibit an extension of the invariant (1) which provides a constant of the motion for a natural generalization of (2).

The dynamical system represented by the nonlinear equation

$$\ddot{y} + \omega^2(t)y = C_1 y^{1-m}(uv)^{(m-4)/2} + C_2 y^{1-2m}(uv)^{m-2} \qquad (3)$$

[$u(t)$ and $v(t)$ are any known, linearly independent solutions of the time-dependent harmonic oscillator $\ddot{u} + \omega^2(t)u = 0$, and C_1, C_2, and m are constants], has been treated by Reid,[5] who found the one-parameter family of solutions

$$y = [(A/k)u^m + B(uv)^{m/2} + (Ak)v^m]^{1/m}. \qquad (4)$$

In Eq. (4) k is an arbitrary constant, $B = 4C_1/(m-2)W^2$,

$$W = u\dot{v} - v\dot{u}, \text{ and } A = \{[C_2/(m-1)W^2] + \tfrac{1}{4}B^2\}^{1/2}.$$

The quantity W is the Wronskian of u and v, and is easily shown to be constant; therefore A and B are also constants.

On the basis of certain formal properties of the solution (4), Ray and Reid[4] conjectured that (3) could be analyzed using the formulation (2). We will demonstrate that this is indeed the case.

To derive the solution (4) using (1) and (2) we first set

$$f(\lambda) = d_1\lambda^{m-3} + d_2\lambda^{2m-3}$$

and

$$g(\lambda) = -\lambda$$

in Eqs. (1) and (2); we then find that the resulting partially coupled system

$$\ddot{y} + \omega^2(t)y = d_1x^{m-4}y^{1-m} + d_2x^{2m-4}y^{1-2m}, \qquad (5a)$$

$$\ddot{x} + \omega^2(t)x = -1/x^3, \qquad (5b)$$

possesses the invariant

$$I = \tfrac{1}{2}[\alpha(x/y)^{m-2} + \beta(x/y)^{2m-2} - (y/x)^2 + (x\dot{y} - y\dot{x})^2], \qquad (6)$$

where we have defined the constants $\alpha = 2d_1/(m-2)$, $\beta = d_2/(m-1)$. If we let $\xi = y/x$, so that $\dot{\xi} = (x\dot{y} - y\dot{x})/x^2$, then (6) yields the following differential equation for $\xi(t)$:

$$\frac{d\xi}{dt} = \frac{1}{x^2}(2I - \alpha\xi^{2-m} - \beta\xi^{2-2m} + \xi^2)^{1/2}, \qquad (7)$$

in which $x(t)$ is considered a known function of t. It turns out that for the particular choice $I = 0$ the integral arising from (7) may be expressed in closed form; in particular, setting $I = 0$ in (7) leads to

$$\int^{(y/x)} \frac{\xi^{m-1}d\xi}{(\xi^{2m} - \alpha\xi^m - \beta)^{1/2}} = \int^t \frac{dt}{x^2},$$

and evaluating the integral on the left gives

$$\frac{1}{m}\ln\left\{\left[\left(\frac{y}{x}\right)^{2m} - \alpha\left(\frac{y}{x}\right)^m - \beta\right]^{1/2} + \left[\left(\frac{y}{x}\right)^{2m} - \alpha\left(\frac{y}{x}\right)^m + \frac{\alpha^2}{4}\right]^{1/2}\right\} = \int^t \frac{dt}{x^2}. \qquad (8)$$

It has been shown by Ray and Reid[4] that a solution of (5b) may be represented in the form $x = (2uv/W)^{1/2}$, where u and v are linearly independent solutions of $\ddot{u} + \omega^2(t)u = 0$, and $W = u\dot{v} - v\dot{u}$ is the Wronskian of u and v. We may therefore write $\int^t dt/x^2 = \ln(Cv/u)^{1/2}$, where C is a constant of integration. Using this result on the right-hand side of (8), solving (8) for $(y/x)^m$, and again using $x = (2uv/W)^{1/2}$, leads to

$$y(t) = \left(\frac{2}{W}\right)^{1/2}\left[\frac{1}{2K}\left(\beta + \frac{\alpha^2}{4}\right)u^m + \frac{\alpha}{2}(uv)^{m/2} + \frac{K}{2}v^m\right]^{1/m}, \qquad (9)$$

where K is an arbitrary constant. Putting $x = (2uv/W)^{1/2}$ in (5a) then shows that (9) provides a one-parameter family of solutions of

$$\ddot{y} + \omega^2(t)y = d_1 y^{1-m}\left(\frac{2uv}{W}\right)^{(m-4)/2} + d_2 y^{1-2m}\left(\frac{2uv}{W}\right)^{m-2}. \qquad (10)$$

 0022-2488/80/061370-02$1.00

We may define new constants by

$$C_1 = d_1\left(\frac{2}{W}\right)^{(m-4)/2}, \quad C_2 = d_2\left(\frac{2}{W}\right)^{m-2},$$

$$k = \left(\frac{2}{W}\right)^{(m-2)/2} K\Big/\left\{\frac{C_2}{m-1} + \left[\frac{2C_1}{(m-2)W}\right]^2\right\}^{1/2};$$

if we express (9) and (10) in terms of these new constants, we obtain (3) and (4), and the proof is complete.

The demonstration that (4) is a solution of (3) by the above technique furnishes another example of the usefulness of the formulations (1) and (2) for time-dependent invariants of coupled systems. In view of the effectiveness of the formulation, we have throught it worthwhile to present a generalization of it to the following system:

$$\ddot{y} + \dot{\sigma}\dot{y} + \omega^2(t)y = \frac{e^{-2\sigma(t)}}{xy^2} f\left(\frac{x}{y}\right), \tag{11a}$$

$$\ddot{x} + \dot{\sigma}\dot{x} + \omega^2(t)x = \frac{e^{-2\sigma(t)}}{yx^2} g\left(\frac{y}{x}\right), \tag{11b}$$

where $g(y/x), \sigma(t), \omega(t)$, and $f(x/y)$ are arbitrary functions of their arguments. To obtain the invariant for this system multiply (11a) by $e^{2\sigma}(x\dot{y} - y\dot{x})x$, multiply (11b) by $e^{2\sigma} \times (x\dot{y} - y\dot{x})y$, and subtract the results. The expression so obtained may be written as a total time derivative, from which it immediately follows that the quantity

$$I = \tfrac{1}{2}e^{2\sigma}(x\dot{y} - y\dot{x})^2 + \int^{x/y} f(\lambda)\,d\lambda + \int^{y/x} g(\lambda)\,d\lambda \tag{12}$$

is a constant of the motion for (11). If we set $g(\lambda) = 0$ and $f(\lambda) = \lambda$ in (11) and (12) we obtain an invariant for the damped time-dependent oscillator, previously found in other ways by Eliezer and Gray,[6] Leach,[7] and Korsch.[8] As these authors have shown, an oscillator invariant can be specified by knowledge of any particular solution of a certain auxiliary equation; in our treatment (11b) yields the oscillator equation and (11a) is the auxiliary equation.

Note added in proof: The generalization (11) has also been obtained, independently and in another manner, by Ray and Reid.[9]

[1]J.R. Ray and J.L. Reid, Phys. Lett. A **71**, 317 (1979).
[2]H.R. Lewis, Jr., Phys. Rev. Lett. **18**, 510 (1967).
[3]M. Lutzky, Phys. Lett. A **68**, 3 (1978).
[4]J.R. Ray and J.L. Reid, J. Math. Phys. **20**, 2054 (1979).
[5]J.L. Reid, Proc. Am. Math. Soc. **38**, 532 (1973).
[6]C.J. Eliezer and A. Gray, SIAM J. Appl. Math. **30**, 463 (1976).
[7]P.G. Leach, SIAM J. Appl. Math. **34**, 496 (1978).
[8]H.J. Korsch, Phys. Lett. A **74**, 294 (1979).
[9]J. R. Ray and J. L. Reid (to be published).

Gauge theory in Hamiltonian classical mechanics: The long range fields

B. Boisseau
Department de Physique, Faculté des Sciences de Tours, Parc de Grandmont, Tours 37, France

(Received 21 November 1979; accepted for publication 11 January 1980)

The generalization of a previous work leads us to a gauge theory in Hamiltonian classical mechanics whose elements are: First, the infinitesimal canonical transformations considered as gauge transformations and, second, an infinite sequence of gauge potentials. A hierarchy of approximation orders allows the physical interpretation of this formalism. At the zero order we obtain the electromagnetic field, at the first order the electromagnetic and gravitational fields. At the second order a new field is added to the former ones. We have studied the new physical features involved by this hypothetical field, in the motion of a classical particle. In the approximation of a Keplerian motion, we have given the period. This last result could be eventually used to test this theory.

I. INTRODUCTION

In a recent paper[1] the electromagnetic and gravitational fields were obtained as gauge fields associated with a special class of canonical transformations continuously connected to the identity. We propose in this work, to generalize the process to all the canonical transformations continuously connected to the identity. We shall use here, for practical reasons, the infinitesimal transformations. The Hamiltonian of the free particle is not form invariant under such transformations. Invariance can be obtained by the introduction of gauge potentials according to a minimal coupling principle. The variance of these potentials is determined by requiring that the Hamiltonian should be form invariant.

Section II, after a transposition a relativistic picture of Hamiltonian mechanics, carries the scheme described above into effect. Thus we introduce the gauge functions $K_\alpha(p,x)$. By expanding the power series in p of K_α, one defines an infinite sequence of gauge potentials. Additional assumptions of physical origin on the convergence of series allow one, according to the selected approximation, to limit the development to a given order and so have a physical interpretation of this formalism. In Sec. III we study the first-order approximation. The latter yields the only two long range fields that we know: the electromagnetic and gravitational fields. (This result was obtained in Ref. 1.) Then we are particularly interested by the immediately upper approximation (second order) which will permit us to see what new effects it is possible to expect of this theory. We work out this study in Sec. IV. Thus one introduces a new long range field that one specifies by its variance and its influence on the motion of the particle. Finally we perform the Newtonian limit and in the case of a Keplerian motion we give the modification of Kepler's third law.

II. GAUGE THEORY IN HAMILTONIAN CLASSICAL MECHANICS

A. General formalism ($c = 1$)

Let us consider a Minkowski space (x^α) to which one associates an eight-dimensional phase space (x^α, p_α). The equations of motion of a free particle are Hamilton's equations

$$\frac{d}{d\tau}p_\alpha = -\frac{\partial H_0}{\partial x^\alpha}, \quad \frac{dx^\alpha}{d\tau} = \frac{\partial H_0}{\partial p_\alpha}, \tag{2.1}$$

where the Hamiltonian[2]

$$H_0 = (1/2m)\, p_\alpha\, p_\beta \eta^{\alpha\beta} \quad (\eta^{\alpha\beta} = -1,1,1,1), \tag{2.2}$$

is a scalar quantity and τ designates the proper time

$$d\tau = (-dx^\alpha\, dx^\beta\, \eta_{\alpha\beta})^{1/2}. \tag{2.3}$$

In order to consider canonical transformations in that eight-dimensional space, we suppose τ is an evolution parameter external to the space–time. We shall give it back its meaning of proper time at the last moment to solve the equations. Therefore the theory of canonical transformations is identical with the usual theory. Let us consider the canonical transformations in this eight-dimensional phase space

$$x'^\mu = x'^\mu(x^\alpha, p_\alpha), \qquad p'_\mu = p'_\mu(x^\alpha, p_\alpha). \tag{2.4}$$

In the new system, the equations of motion become:

$$\frac{d}{d\tau}\, p'_\mu = -\frac{\partial H'_0}{\partial x'^\mu}, \quad \frac{d}{d\tau}x'^\mu = \frac{\partial H'_0}{\partial p'_\mu}, \tag{2.5}$$

where

$$H'_0(x'^\mu, p'_\mu) = H_0(x^\alpha, p_\alpha), \tag{2.6}$$

since τ does not occur in the transformations (2.4).

All the infinitesimal canonical transformations can be generated by functions of type[3]

$$F_3(p,x') = -\, p_\alpha x'^\alpha - \epsilon G(p,x'), \tag{2.7}$$

where ϵ is an infinitesimal parameter and $G(p,x)$ is an arbitrary function that will be named a generating function. The canonical transformations (2.4) generated by (2.7) are given, to first order in ϵ, by[4]

$$x'^\mu = x^\mu - \epsilon(\partial G/\partial p_\mu)(p,x), \qquad p'_\mu = p_\mu + \epsilon(\partial G/\partial x^\mu)(p,x). \tag{2.8}$$

 0022-2488/80/061372-05$1.00

In the same way we can write the inverse relations:

$$x^\mu = x'^\mu + \epsilon(\partial G/\partial p'_\mu)(p',x'),$$
$$p_\mu = p'_\mu - \epsilon(\partial G/\partial x'^\mu)(p',x'). \tag{2.9}$$

By substituting (2.9) in (2.2), Eq. (2.6) gives

$$H'_0 = (1/2m)(p'_\alpha - \epsilon(\partial G/\partial x'^\alpha)(p',x')) \times (p'_\beta - \epsilon(\partial G/\partial x'^\beta)(p',x'))\eta^{\alpha\beta}. \tag{2.10}$$

By comparing (2.2) and (2.10) it is obvious that the Hamiltonian is not form invariant.

If one considers these transformations as gauge transformations, invariance will be obtained by a minimal coupling principle[5]: In the Hamiltonian H'_0 one substitutes the functions $K_\alpha(p',x')$ for the derivatives $\epsilon\partial G/\partial x'^\alpha(p',x')$ coming from the gauge transformations. The variance of these functions will enable us to obtain the form invariance of the new Hamiltonian so obtained. Therefore let us perform, in the Hamiltonian (2.10), the substitution

$$\epsilon(\partial G/\partial x'^\alpha)(p',x') \to K_\alpha(p',x'). \tag{2.11}$$

Thus one obtains, after dropping the primes, the new Hamiltonian

$$H = (1/2m)[p_\alpha - K_\alpha(p,x)][p_\beta - K_\beta(p,x)]\eta^{\alpha\beta}. \tag{2.12}$$

It will be convenient to introduce the quantity

$$h_\alpha = p_\alpha - K_\alpha(p,x). \tag{2.13}$$

The Hamiltonian (2.12) is then written

$$H = (1/2m)h_\alpha h_\beta \eta^{\alpha\beta}. \tag{2.14}$$

Let us examine the variance of this Hamiltonian under infinitesimal canonical transformation. It is sufficient to deal with h_α. Introduction of (2.9) into (2.13) gives

$$h'_\alpha(p',x') = p'_\alpha - \epsilon(\partial G/\partial x'^\alpha)(p',x') - K_\alpha(p' - \epsilon(\partial G/\partial x'), x' + \epsilon(\partial G/\partial p')). \tag{2.15}$$

If we define

$$K'_\alpha(p',x') = \epsilon(\partial G/\partial x'^\alpha)(p',x') + K_\alpha(p' - \epsilon(\partial G/\partial x'), x' + \epsilon(\partial G/\partial p')), \tag{2.16}$$

Eq. (2.15) can be written

$$h'_\alpha = p'_\alpha - K'_\alpha(p',x'), \tag{2.17}$$

and the Hamiltonian

$$H' = (1/2m)[p'_\alpha - K'_\alpha(p',x')][p'_\beta - K'_\beta(p',x')]\eta^{\alpha\beta}. \tag{2.18}$$

The comparison of Eqs. (2.12) and (2.18) shows clearly the formal invariance of the Hamiltonian on condition that the functions K_α should be transformed jointly according to (2.16).

Let us examine the transformation (2.16). By expanding the last term of the second member to first order in ϵ

$$K_\alpha(p' - \epsilon(\partial G/\partial x'), x' + \epsilon(\partial G/\partial p')) = K_\alpha(p',x') - \epsilon(\partial G/\partial x'^\mu)(\partial K_\alpha/\partial p'_\mu) + \epsilon(\partial G/\partial p'_\mu)(\partial K_\alpha/\partial x'^\mu), \tag{2.19}$$

the transformation (2.16) becomes

$$K'_\alpha(p',x') = K_\alpha(p',x') + \epsilon(\partial G/\partial x'^\alpha)(p',x') + \epsilon[K_\alpha, G]_{x',p'}, \tag{2.20}$$

where

$$[K_\alpha, G]_{x',p'} = (\partial K_\alpha/\partial x'^\mu)(\partial G/\partial p'_\mu) - (\partial K_\alpha/\partial p'_\mu)(\partial G/\partial x'^\mu), \tag{2.21}$$

is a Poisson bracket.

Let us remark that the transformation (2.20) is perfectly similar to those of gauge potentials in the usual theory of Yang and Mills. One simply substitutes a Poisson bracket for a commutator of matrices. Nevertheless we are going to use, later on, the transformation (2.20) in a slightly different form:

$$K'_\alpha(p',x') = K_\alpha(p',x) + \epsilon(\partial G/\partial x^\alpha)(p',x) - \epsilon(\partial K_\alpha/\partial p'_\mu)(p',x)(\partial G/\partial x^\mu)(p',x). \tag{2.22}$$

This can be obtained for example, from (2.20) by expanding $K_\alpha(p',x')$ to first order in ϵ

$$K_\alpha(p',x') = K_\alpha(p',x) - \epsilon(\partial G/\partial p'_\mu)(p',x') \times(\partial K_\alpha/\partial x^\mu)(p',x), \tag{2.23}$$

and by remarking that we may indifferently use the primed or unprimed variables in the terms enclosing the factor ϵ.

B. The long range fields.

Now we must interpret this general formalism and give a physical meaning to the gauge functions $K_\alpha(p,x)$. In physics, it is not a strong limitation to suppose that the functions are analytic. Therefore, let us express $G(p,x)$ and $K_\alpha(p,x)$ by a power series in p:

$$G = g^{(0)} + g^{(1)} + \cdots + g^{(i)} + \cdots, \tag{2.24}$$
$$K_\alpha = k_\alpha^{(0)} + k_\alpha^{(1)} + \cdots + k_\alpha^{(i)} + \cdots, \tag{2.25}$$

with

$$g^{(0)} = G(0,x) = G^{(0)}(x), \quad g^{(i)} = \frac{1}{i!} G^{(i)\mu_1\cdots\mu_i}(x)\, p_{\mu_1}\cdots p_{\mu_i}, \tag{2.26}$$

$$k_\alpha^{(0)} = K_\alpha(0,x) = K_\alpha^{(0)}(x), \quad k_\alpha^{(i)} = \frac{1}{i!} K_\alpha^{(i)\mu_1\cdots\mu_i}(x)\, p_{\mu_1}\cdots p_{\mu_i}, \tag{2.27}$$

[$G^{(i)\mu_1\cdots\mu_i}$ and $K_\alpha^{(i)\mu_1\cdots\mu_i}$ are symmetrical].

Let us add an arbitrary hypothesis whose physical meaning will be understood later on. We suppose that the series (2.25) converges rapidly and more precisely:

$$|K_\alpha^{(i+1)}(x)| \ll |K_\alpha^{(i)}(x)|, \quad |K_\alpha^{(i)}(x)|\cdot|K_\alpha^{(j)}(x)| \sim |K_\alpha^{(i+j)}(x)|. \tag{2.28}$$

The diverse quantities $K_\alpha^{(i)}(x)$ will be interpreted as the gauge potentials of the long range fields, $K_\alpha^{(0)}$ for the electromagnetic field, $K_\alpha^{(1)\mu}$ for the gravitational field, etc. Expressing $K_\alpha(p,x)$ by a power series expansion in p allows us to carry the transformation (2.22) onto the gauge potentials $K_\alpha^{(i)}(x)$. For convenience we denote

$$(\partial g^{(i)}(p',x)/\partial x^\mu) = g_{,\mu}^{(i)}(p',x), \quad (\partial k_\alpha^{(i)}(p',x)/\partial p'_\mu) = k_\alpha^{(i),\mu}(p',x). \tag{2.29}$$

By using Eqs. (2.24), (2.25), and (2.29) we may write Eq. (2.22) in the form

$$\sum_i k'^{(i)}_\alpha(p',x') = \sum_i \left(k^{(i)}_\alpha(p',x) + \epsilon g^{(i)}_{,\alpha}(p',x)\right) - \epsilon\Big(\sum_i k^{(i),\mu}_\alpha(p',x)\Big)\Big(\sum_j g^{(j)}_{,\mu}(p',x)\Big). \tag{2.30}$$

Remarking that $k^{(i)}_\alpha$ and $g^{(i)}_{,\alpha}$ are ith power of p' and

$$k^{(i),\mu}_\alpha = \frac{1}{(i-1)!} K^{(i)\mu_1\cdots\mu_{i-1}\mu}_\alpha \, p'_{\mu_1}\cdots p'_{\mu_{i-1}},$$

is a $(i-1)$th power of p', we can arrange (2.30) in ascending powers of p':

$$\sum_i k'^{(i)}_\alpha = \sum_i \{k^{(i)}_\alpha + \epsilon g^{(i)}_{,\alpha} - \epsilon[k^{(1),\mu}_\alpha g^{(i)}_{,\mu} + \cdots + k^{(i),\mu}_\alpha g^{(1)}_{,\mu} + k^{(i+1),\mu}_\alpha g^{(0)}_{,\mu}]\}. \tag{2.31}$$

From which, identifying the same powers of p' in the two members of (2.31), one obtains

$$K'^{0}_\alpha(x') = K^{(0)}_\alpha(x) + \epsilon G^{(0)}_{,\alpha}(x) - \epsilon K^{(1)\mu}_\alpha(x) G^{(0)}_{,\mu}(x),$$

$$\begin{aligned} K'^{(i)\mu_1\cdots\mu_i}_\alpha(x') &= K^{(i)\mu_1\cdots\mu_i}_\alpha(x) + \epsilon G^{(i)\mu_1\cdots\mu_i}_{,\alpha}(x) - \epsilon\big[K^{(1)\mu}_\alpha(x) G^{(i)\mu_1\cdots\mu_i}_{,\mu}(x) \\ &\quad + C^1_i K^{(2)\mu_1\mu}_\alpha(x) G^{(i-1)\mu_2\cdots\mu_i}_{,\mu}(x) + C^2_i K^{(3)\mu_1\mu_2\mu}_\alpha(x) \\ &\quad \times G^{(i-2)\mu_3\cdots\mu_i}_{,\mu}(x) + \dots + C^{(i-1)}_i K^{(i)\mu_1\cdots\mu_{i-1}\mu}_\alpha(x) G^{(1)\mu_i}_{,\mu}(x) \\ &\quad + K^{(i+1)\mu_1\cdots\mu_i\mu}_\alpha(x) G^{(0)}_{,\mu}(x)\big]. \end{aligned} \tag{2.32}$$

Equations (2.32) describe the variance of gauge potentials.

C. Approximation of nth order

Let us make use of hypothesis (2.28). We say that we are in the approximation of nth order if we neglect the potentials $K^{(i)}_\alpha$ with $i > n$. In this approximation, the ith equation (2.32) (with $i \leqslant n-1$) is not altered, on the contrary we neglect the last term of the second member of the nth equation:

$$K^{(n+1)\mu_1\cdots\mu_n\mu}_\alpha(x) G^{(0)}_{,\mu}(x) = 0. \tag{2.33}$$

Let us define:

$$h^{(n)}_\alpha = p_\alpha - K^{(0)}_\alpha - K^{(1)\mu}_\alpha(x) p_\mu - \cdots - \frac{1}{n!} K^{(n)\mu_1\cdots\mu_n}_\alpha(x) p_{\mu_1}\cdots p_{\mu_n}. \tag{2.34}$$

The Hamiltonian is written

$$H = (1/2m) h^{(n)}_\alpha h^{(n)}_\beta \eta^{\alpha\beta}, \tag{2.35}$$

where one eliminates the products such as $K^{(i)}_\alpha K^{(j)}_\beta$ with $i + j > n$.

Let us introduce now new notations in order to make the physical interpretation of gauge potentials easier. In Eq. (2.34) it is possible to group the terms in p

$$p_\alpha - K^{(1)\mu}_\alpha p_\mu = (\delta^\mu_\alpha - K^{(1)\mu}_\alpha) p_\mu.$$

Let us define

$$V_\alpha{}^\mu = \delta^\mu_\alpha - K^{(1)\mu}_\alpha, \tag{2.36}$$

and $V^\alpha{}_\mu$ the inverse matrix (one suppose it exists).

Let us introduce

$$B_\mu = V^\alpha{}_\mu K^{(0)}_\alpha, \quad K^{(0)}_\alpha = B_\mu V_\alpha{}^\mu, \tag{2.37}$$

$$N^{(i)\mu_1\cdots\mu_i}_\mu = V^\alpha{}_\mu K^{(i)\mu_1\cdots\mu_i}_\alpha, \quad K^{(i)\mu_1\cdots\mu_i}_\alpha = N^{(i)\mu_1\cdots\mu_i}_\mu V_\alpha{}^\mu \quad (i \geqslant 2). \tag{2.38}$$

Therefore Eq. (2.34) is written

$$\begin{aligned} h^{(n)}_\alpha &= h'^{(n)}_\mu V_\alpha{}^\mu \\ &= \Big(-B_\mu + p_\mu - \tfrac{1}{2} N^{(2)\mu_1\mu_2}_\mu p_{\mu_1} p_{\mu_2} - \cdots - \frac{1}{n!} N^{(n)\mu_1\cdots\mu_n}_\mu p_{\mu_1}\cdots p_{\mu_n}\Big) V_\alpha{}^\mu, \end{aligned} \tag{2.39}$$

and the Hamiltonian

$$H = (1/2m) h'^{(n)}_\mu h'^{(n)}_\nu g^{\mu\nu}, \tag{2.40}$$

with

$$g^{\mu\nu} = V_\alpha{}^\mu V_\beta{}^\nu \eta^{\alpha\beta}. \tag{2.41}$$

In Eq. (2.40) one eliminates the products such as $N^{(i)}_\mu \cdot N^{(j)}_\nu$ with $i + j > n$.

We shall now recognize the nature of gauge potentials that we have introduced, on the one hand by their variances [Eqs. (2.32) and (2.33)], on the other hand by the equations of motion that one obtains from the Hamiltonian [(2.35) and (2.40)]. To do this, we envisage two cases: First, the approximation $n = 1$ which generates all the known long range fields (electromagnetic and gravitational fields), second, the immediately upper approximation $n = 2$ which will permit us to see what new effects it is possible to expect in the frame of this theory. Let us remark that the form of the Hamiltonian (2.40) places in a prominent position a breaking between $n \leqslant 1$ and $n > 1$. The fact of restricting the investigations to $n = 2$, if it simplifies computations, would not be likely to singularize essentially the problem $n > 1$.

III. APPROXIMATION $n = 1$: THE ELECTROMAGNETIC AND GRAVITATIONAL FIELDS

The Hamiltonian is reduced to

$$H = (1/2m)(p_\mu - B_\mu)(p_\nu - B_\nu) g^{\mu\nu}. \tag{3.1}$$

The variance of potentials [Eqs. (2.32) and (2.33)] is the following:

$$K'^{(0)}_\alpha(x') = K^{(0)}_\alpha(x) + \epsilon G^{(0)}_{,\alpha}(x) - \epsilon K^{(1)\mu}_\alpha(x) G^{(0)}_{,\mu}(x), \tag{3.2}$$

$$K'^{(1)\mu}_\alpha(x') = K^{(1)\mu}_\alpha(x) + \epsilon G^{(1)\mu}_{,\alpha}(x) - \epsilon K^{(1)\rho}_\alpha(x) G^{(1)\mu}_{,\rho}(x). \tag{3.3}$$

By using Eqs. (2.36) and (2.37), the relations (3.2) and (3.3) become:

$$B'_\mu(x') = [\delta^\nu_\mu + \epsilon G^{(1)\nu}_{,\mu}(x)][B_\nu(x) + \epsilon G^{(0)}_{,\nu}(x)], \tag{3.4}$$

$$V'^\mu_\alpha(x') = [\delta^\mu_\nu - \epsilon G^{(1)\mu}_{,\nu}(x)] V_\alpha{}^\nu(x). \tag{3.5}$$

Interpretation of this result is simple. If $G(p,x)$ is reduced to $G^{(0)}(x) + G^{(1)\mu}(x) p_\mu$, we have:

$$\delta^\mu_\nu - \epsilon G^{(1)\mu}_{,\nu} = (\partial x'^\mu / \partial x^\nu), \quad \delta^\nu_\mu + \epsilon G^{(1)\nu}_{,\mu} = (\partial x^\nu / \partial x'^\mu). \tag{3.6}$$

Therefore it suffices to go back to Ref. 1 in which Eqs. (2.41), (3.1), (3.4), and (3.5) were obtained.

Let us be reminded here of the results. The potential B_μ has the variance of an electromagnetic potential [Eq. (3.4)], $g^{\mu\nu}$ is a Riemannian metric associated with tetrad[6] $V_\alpha{}^\mu$ [Eqs. (2.41) and (3.5)], the proper time is

$$d\tau^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu. \tag{3.7}$$

Hamilton's equations yield

$$m\left((dU^{\lambda}/d\tau) + \begin{Bmatrix}\lambda\\ \mu\nu\end{Bmatrix} U^{\mu}U^{\nu}\right) = g^{\lambda\rho}eF_{\rho\sigma}U^{\sigma}, \qquad (3.8)$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $B_\mu = eA_\mu$ (e is an electric charge), $U^\lambda = dx^\lambda/d\tau$ is the four velocity and $\{^{\lambda}_{\mu\nu}\}$ designates Christoffel's symbols. The Eq. (3.8) describes the motion of a charged particle in electromagnetic and gravitational fields. We deduce from this that A_μ and $g_{\mu\nu}$ [or $K_\alpha^{(0)}$ and $K_\alpha^{(1)\mu}$] are the gauge potentials of the electromagnetic and gravitational fields. Finally, let us notice that the approximation $n = 0$ gives us only the electromagnetic potential A_μ [or $K_\alpha^{(0)}$].

IV. APPROXIMATION $n = 2$

In this approximation the Hamiltonian is written

$$H = (1/2m)(p_\mu - B_\mu)(p_\nu - B_\nu)g^{\mu\nu} - (1/2m)g^{\mu\nu} \times (p_\mu - B_\mu)N_\nu^{(2)\nu_1\nu_2} p_{\nu_1} p_{\nu_2}. \qquad (4.1)$$

The variance of gauge potentials

$$K_\alpha^{\prime(0)}(x') = K_\alpha^{(0)}(x) + \epsilon G^{(0)}_{,\alpha}(x) - K_\alpha^{(1)\mu}(x)G^{(0)}_{,\mu}(x),$$
$$K_\alpha^{\prime(1)\mu}(x') = K_\alpha^{(1)\mu}(x) + \epsilon G^{(1)\mu}_{,\alpha}(x) - \epsilon K_\alpha^{(1)\rho}(x)G^{(1)\mu}_{,\rho}(x) - \epsilon K_\alpha^{(2)\mu\rho}(x)G^{(0)}_{,\rho}(x), \qquad (4.2)$$
$$K_\alpha^{\prime(2)\mu_1\mu_2}(x') = K_\alpha^{(2)\mu_1\mu_2}(x) + \epsilon G^{(2)\mu_1\mu_2}_{,\alpha}(x) - \epsilon K_\alpha^{(1)\rho}(x) \times G^{2\mu_1\mu_2}_{,\rho}(x) - \epsilon 2K_\alpha^{(2)\mu_1\rho}(x)G^{(1)\mu_2}_{,\rho}(x),$$

is intricate and the understanding is not improved in the new notations. On the contrary, if one examines the variance of these potentials when $G(p,x)$ is reduced to $G^{(1)\mu}(x)p_\mu$, that is to say when the canonical transformation is reduced to a transformation of general relativity, we obtain

$$B'_\mu(x') = (\partial x^\nu/\partial x'^\mu)B_\nu(x),$$
$$V'^{\mu}_{\alpha}(x') = (\partial x'^\mu/\partial x^\nu)V_\alpha{}^\nu(x), \qquad (4.3)$$
$$K_\alpha^{\prime(2)\mu_1\mu_2}(x') = (\partial x'^{\mu_1}/\partial x^{\nu_1})(\partial x'^{\mu_2}/\partial x^{\nu_2})K_\alpha^{(2)\nu_1\nu_2}(x).$$

Accordingly $N_\mu^{(2)\mu_1\mu_2}$ is a three-tensor.

We propose, for a better understanding of the new field, to eliminate the variables p_μ from Hamilton's equations

$$(d/d\tau)x^\mu = \frac{\partial H}{\partial p_\mu}, \quad (d/d\tau)p_\mu = -\frac{\partial H}{\partial x^\mu}, \qquad (4.4)$$

and, in this way, to obtain the equations of motion of the particle in a form similar to Eq. (3.8).

A. Neutral particles

In order to simplify computation, we exclude the electromagnetic field (we suppose the particle is not charged). The Hamiltonian (4.1) becomes:

$$H = (1/2m)(g^{\mu\nu}p_\mu p_\nu - \tfrac{1}{3}\epsilon^{\mu_1\mu_2\mu_3}p_{\mu_1}p_{\mu_2}p_{\mu_3}), \qquad (4.5)$$

in which we have introduced the symmetric tensor

$$\epsilon^{\mu_1\mu_2\mu_3} = g^{\mu_1\nu}N_\nu^{(2)\mu_2\mu_3} + g^{\mu_2\nu}N_\nu^{(2)\mu_3\mu_1} + g^{\mu_3\nu}N_\nu^{(2)\mu_1\mu_2}. \qquad (4.6)$$

From the first Eq. (4.4) associated with the Hamiltonian (4.5) we deduce that

$$v^\mu = (dx^\mu/d\tau) = (1/2m)(g^{\mu\nu}p_\nu - \tfrac{1}{2}\epsilon^{\mu\nu\rho}p_\nu p_\rho), \qquad (4.7)$$

since $\epsilon^{\mu\nu\rho} \ll g^{\mu\nu}$ [hypothesis (2.28)], we are going to solve Eq. (4.7) by iterative approximation. To first order in $\epsilon^{\mu\nu\rho}$ (i.e., to second order), we obtain

$$p_\gamma = mg_{\gamma\nu}v^\nu + (m^2/2)\epsilon^{\mu\nu\rho}g_{\mu\gamma}g_{\nu\chi}g_{\rho\sigma}v^\chi v^\sigma. \qquad (4.8)$$

We must now introduce Eq. (4.8) in the second Eq. (4.4). First, let us compute

$$(\partial H/\partial x^\gamma) = (1/2m)(\partial_\gamma g^{\gamma_1\gamma_2}p_{\gamma_1}p_{\gamma_2} - \tfrac{1}{3}\partial_\gamma \epsilon^{\gamma_1\gamma_2\gamma_3}p_{\gamma_1}p_{\gamma_2}p_{\gamma_3}). \qquad (4.9)$$

By inserting Eq. (4.8) in (4.9) and limiting ourself to first order in $\epsilon^{\mu\nu\rho}$ we have, after some computations

$$(\partial H/\partial x^\gamma) = (m/2)\partial_\gamma g^{\gamma_1\gamma_2}g_{\gamma_1\nu_1}g_{\gamma_2\nu_2}v^{\nu_1}v^{\nu_2} - (m^2/6)\partial_\gamma \epsilon_{\nu_1\nu_2\nu_3}v^{\nu_1}v^{\nu_2}v^{\nu_3}. \qquad (4.10)$$

Moreover

$$(d/d\tau)p_\gamma = (d/d\tau)(mg_{\gamma\nu}v^\nu) + m^2\epsilon_{\gamma\chi\sigma}v^\chi(dv^\sigma/d\tau) + (m^2/2)\partial_\nu\epsilon_{\gamma\chi\sigma}v^\nu v^\chi v^\sigma. \qquad (4.11)$$

So the second Eq. (4.4) is written

$$(d/d\tau)(mg_{\gamma\nu}v^\nu) + \frac{m}{2}\partial_\gamma g^{\gamma_1\gamma_2}g_{\gamma_1\nu_1}g_{\gamma_2\nu_2}v^{\nu_1}v^{\nu_2} + m^2\epsilon_{\gamma\chi\sigma}v^\chi(dv^\sigma/d\tau)$$
$$= (m^2/6)\partial_\gamma\epsilon_{\nu_1\nu_2\nu_3}v^{\nu_1}v^{\nu_2}v^{\nu_3} - (m^2/2)\partial_\nu\epsilon_{\gamma\chi\sigma}v^\nu v^\chi v^\sigma, \qquad (4.12)$$

or

$$mg_{\gamma\nu}\left((d/d\tau)v^\nu + \begin{Bmatrix}\nu\\ \nu_1\nu_2\end{Bmatrix}v^{\nu_1}v^{\nu_2}\right) + m^2\epsilon_{\gamma\chi\sigma}v^\chi\frac{dv^\sigma}{d\tau}$$
$$= -m^2g_{\gamma\nu}\begin{Bmatrix}\nu\\ \nu_1\nu_2\nu_3\end{Bmatrix}v^{\nu_1}v^{\nu_2}v^{\nu_3}. \qquad (4.13)$$

Where we have used the Christoffel's symbols

$$\begin{Bmatrix}\nu\\ \nu_1\nu_2\end{Bmatrix} = \tfrac{1}{2}g^{\nu\gamma}(\partial_{\nu_1}g_{\gamma\nu_2} + \partial_{\nu_2}g_{\gamma\nu_1} - \partial_\gamma g_{\nu_1\nu_2}), \qquad (4.14)$$

and symbols constructed in a similar way

$$\begin{Bmatrix}\nu\\ \nu_1\nu_2\nu_3\end{Bmatrix} = (1/3!)g^{\nu\gamma}(\partial_{\nu_1}\epsilon_{\gamma\nu_2\nu_3} + \partial_{\nu_2}\epsilon_{\gamma\nu_3\nu_1} + \partial_{\nu_3}\epsilon_{\gamma\nu_1\nu_2} - \partial_\gamma\epsilon_{\nu_1\nu_2\nu_3}). \qquad (4.15)$$

Equation (4.13) can be written

$$\frac{d}{d\tau}v^\sigma(\delta_\sigma^\nu + mg^{\gamma\nu}\epsilon_{\gamma\chi\sigma}v^\chi) + \begin{Bmatrix}\nu\\ \nu_1\nu_2\end{Bmatrix}v^{\nu_1}v^{\nu_2} + m\begin{Bmatrix}\nu\\ \nu_1\nu_2\nu_3\end{Bmatrix}v^{\nu_1}v^{\nu_2}v^{\nu_3} = 0. \qquad (4.16)$$

One multiplies two members of (4.16) by $\delta_\nu^\mu - mg^{\gamma'\mu}\epsilon_{\gamma'\chi'\nu}v^{\chi'}$, by remembering that $\{^{\nu}_{\nu_1\nu_2}\}$ is a first-order term ($g_{\mu\nu} = \eta_{\mu\nu}$ + first order) one obtains finally to second order:

$$\frac{dv^\mu}{d\tau} + \begin{Bmatrix}\mu\\ \nu_1\nu_2\end{Bmatrix}v^{\nu_1}v^{\nu_2} + m\begin{Bmatrix}\mu\\ \nu_1\nu_2\nu_3\end{Bmatrix}v^{\nu_1}v^{\nu_2}v^{\nu_3} = 0. \qquad (4.17)$$

Thus we came to the equation of a particle of mass m in the first-order field (graviatation) and in the hypothetical second-order field. We observe that this second-order field will have an influence proportional to the mass of the particle. This leads to search for an eventual illustration of this field among the celestial objects.

B. Newtonian approximation, Keplerian motion

Let us consider a particle with a large mass m moving slowly. It is possible to neglect $dx/d\tau$ with respect to $dt/d\tau$, Eq. (4.17) becomes

$$(d^2x^\nu/d\tau^2) + \begin{Bmatrix}\nu\\00\end{Bmatrix}(dt/d\tau)^2 + m\begin{Bmatrix}\nu\\000\end{Bmatrix}(dt/d\tau)^3 = 0. \tag{4.18}$$

We suppose that the fields are stationary, Eqs. (4.14) and (4.15) give

$$\begin{Bmatrix}\nu\\00\end{Bmatrix} = -\tfrac{1}{2}g^{\nu\gamma}\partial_\gamma g_{00}, \tag{4.19}$$

$$\begin{Bmatrix}\nu\\000\end{Bmatrix} = -(1/3!)g^{\nu\gamma}\partial_\gamma \epsilon_{000}. \tag{4.20}$$

We are specially interested by the second-order field, so we are going to suppose the gravitational field is weak:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1 \quad (\text{or } |K_\alpha^{(1)\mu}| \ll 1), \tag{4.21}$$

from which

$$\begin{Bmatrix}\nu\\00\end{Bmatrix} = -\tfrac{1}{2}\eta^{\nu\gamma}\partial_\gamma h_{00} = \eta^{\nu\gamma}\partial_\gamma K_0^{(1)0}, \tag{4.22}$$

$$\begin{Bmatrix}\nu\\000\end{Bmatrix} = -\tfrac{1}{6}\eta^{\nu\gamma}\partial_\gamma \epsilon_{000} = -\tfrac{1}{2}\eta^{\nu\gamma}\partial_\gamma N_0^{(2)00} = -\tfrac{1}{2}\eta^{\nu\gamma}\partial_\gamma K_0^{(2)00}. \tag{4.23}$$

Equations (4.18) give

$$\frac{d^2x^i}{d\tau^2} + \partial_i K_0^{(1)0}(dt/d\tau)^2 - \tfrac{1}{2}m\partial_i K_0^{(2)00}(dt/d\tau)^3 = 0, \tag{4.24}$$

$$(d^2t/d\tau^2) = 0 \Rightarrow (dt/d\tau) = \text{const}, \tag{4.25}$$

from which

$$(d^2x^i/dt^2) = -\partial_i K_0^{(1)0} + \tfrac{1}{2}m(dt/d\tau)\partial_i K_0^{(2)00}. \tag{4.26}$$

This equation is given with ($c = 1$) units, in the usual units we obtain

$$(d^2x^i/dt^2) = -\partial_i K_0^{(1)0}c^2 + \tfrac{1}{2}mc^3\partial_i K_0^{(2)00} \quad ((dt/d\tau) = 1). \tag{4.27}$$

Finally, let us introduce the nonrelativistic potentials

$$\varphi = c^2 K_0^{(1)0}, \quad \psi = -\tfrac{1}{2}c^3 K_0^{(2)00}. \tag{4.28}$$

Therefore the Newtonian equation of the particle is

$$m\frac{d^2\mathbf{x}}{dt^2} = -m\nabla\varphi - m^2\nabla\psi. \tag{4.29}$$

If the gravitational potential is owed to a spherical mass M at a distance r from the particle, it takes the form:

$$\varphi = -GM/r, \tag{4.30}$$

which give the force

$$\mathbf{F}^{(1)} = -(GmM/r^2)(\mathbf{r}/r). \tag{4.31}$$

On the contrary we know scarcely anything on the potential ψ. We just know by Eq. (4.29) that the force is proportional to the square of the particle mass. We may imagine in order to respect the symmetry between the action and reaction that ψ is proportional to M^2. Moreover the gravitostatic and electrostatic potentials in the considered approximation are in $1/r$, it is reasonable to induce from this, that ψ is also a potential in $1/r$. So we guess:

$$\psi = \pm EM^2/r, \tag{4.32}$$

whence

$$\mathbf{F}^{(2)} = \pm E(m^2M^2/r)(\mathbf{r}/r). \tag{4.33}$$

We cannot say anything either on the sign or on the value of the constant E except that it is very weak ($E/G \ll 1$). Equation (4.29) is written

$$\frac{d^2\mathbf{x}}{dt^2} = -GM\left(1 \pm \frac{E}{G}mM\right)\frac{\mathbf{r}}{r^3}. \tag{4.34}$$

Equation (4.34) gives a Keplerian motion whose period T is connected to the half axis a of the ellipse by

$$\frac{a^3}{T^2} = \frac{GM}{4\pi^2}\left(1 \pm \frac{E}{G}mM\right)\left(1 + \frac{m}{M}\right). \tag{4.35}$$

or with $m/M \ll 1$:

$$\frac{a^3}{T^2} = \frac{GM}{4\pi^2}\left(1 \pm \frac{E}{G}mM\right). \tag{4.36}$$

This last formula could be eventually used to test this theory.

V. CONCLUSION

This work constitutes an extension of the covariance of the particle equations from the group of general relativity to a larger one: that of the infinitesimal canonical transformations. The two essential results of this theory are on the one hand the identification of infinitesimal canonical transformations with gauge transformations, and, on the other hand the introduction of the long range fields in a unitary way.

The *ad hoc* hypothesis of decreasing gauge potentials [Relations (2.28)] is consistent with the present physics; the coupling constant of the electromagnetic field ($e^2/4\pi\hbar c = 1/137$) is very large compared with the one of the gravitational field ($Gm^2/\hbar c = 10^{-39}$; m = proton mass). This hypothesis explains also why we do not see, at the present time, the other fields. We may nevertheless hope to put them in evidence, if they exist, in the motion of large masses, which will probably lead us to extend these investigations to the scope of the astronomy. The hypothesis (2.28) replaced the hypothesis of the large unit of mass which was used in an earlier paper. Indeed this hypothesis amounted to supposing the existence of a large fundamental unit of mass playing a part analogous to the one of the velocity of light. But in the present state of physics we have not any information at all about the existence of such a unit.

This theory does not give the way of deducing the equations of evolution of fields. But, besides, we know the equations of electromagnetic and gravitational fields. The formal analogy between the new field analyzed in Sec. IV and the gravitational field [Eqs. (4.14), (4.15), and (4.17)] allows us to hope constructing equations for this new field analogous with those of Einstein.

[1]B. Boisseau and C. Barrabès, J. Math. Phys. **20**, 2058 (1979).
[2]J.W. Leech, *Classical Mechanics* (Methuen, London, 1957), Chap. 10.
[3]See Ref. 1, Sec. 3.
[4]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1950), Chap. 8, Sec. 6.
[5]For justification of such a denomination see Ref. 1, Sec. 2.
[6]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), pp. 365–373.

Coherent states and projective representation of the linear canonical transformations [a]

Ingrid Daubechies [b]
Theoretische Natuurkunde, Vrije Universiteit Brussel, Pleinlaan 2, B 1050 Brussels, Belgium

(Received 26 November 1979; accepted for publication 25 January 1980)

Using a family of coherent state representations we obtain in a natural and coordinate-independent way an explicit realization of a projective unitary representation of the symplectic group. Dequantization of these operators gives us the corresponding classical functions.

1. INTRODUCTION

Canonical transformations and their relations to quantum mechanics have been studied extensively and in many different settings.[1-10] See, for instance Refs. 2 and 3 for a representation in terms of coherent states, Ref. 4 for applications of this treatment of the homogeneous linear canonical transformations, Ref. 5 for an application of the inhomogeneous linear canonical transformations, and Ref. 6 for a relation with Bogoliubov transformations and quasi-free states on the CCR algebra. In Ref. 7 it was advocated that the most natural way to study canonical transformations (we are only concerned with the linear ones here, even if we don't specify so further on) is (1) to work in a phase space realization, and (2) to consider a suitable family of closed subspaces of $L^2(E;dv)$, the square integrable functions on phase space, instead of only one Hilbert space as the basic setting. We follow this point of view here, and use it to derive a simple and natural expression for the operators of the symplectic group, the so-called metaplectic representation. This metaplectic representation was constructed already some ten years ago by Bargmann[2] and Itzykson[3] independently, who both used a holomorphic representation of the canonical commutation relations. Another approach can be found in Ref. 4. In this latter treatment, however, a certain class of linear transformations cannot be treated by the direct formula, and can only be recovered by taking products of linear transformations outside this class; this is not the case in either Refs. 2, 3, or the present paper. Our treatment differs from the ones given in Refs. 2 and 3 in that we obtain the representation almost automatically from the structure of the family of closed subspaces of $L^2(E;dv)$ mentioned above. In fact, for any state ψ with wave function ϕ_ψ in the coherent state representation, we obtain the image $W_S\phi_\psi$ of ϕ_ψ under a canonical transformation S simply by a substitution $(U_S\phi_\psi)(v) = \phi_\psi(S^{-1}v)$, followed by a projection. This projection has to be introduced because the naive substitution above does not always leave invariant the Hilbert space of coherent states. It turns out that this succession of two simple operations (a naive substitution, and a projection back onto the right space when things threaten to go wrong because the substitution has taken us out of it) is, up to some constant factor, a unitary operator. The family of these operators gives us our projective representation. We work with intrinsic and coordinate-free notations differing from the notations used in Refs. 2, 3, or 4. At the end of the paper we rewrite some of the results in the more familiar x–p notations.

Following the prescription given in Ref. 11 for the dequantization of these operators, we proceed then to compute the classical functions corresponding to the symplectic transformations. This calculation of classical functions for symplectic transformations has been done for one-parameter subgroups of the symplectic group.[8,9] One then only catches a small part of the symplectic group at a time; moreover, since the group is not exponential, not every symplectic transformation can be considered as an element of such a one-parameter subgroup. In Ref. 10 a general formula for the classical functions corresponding to symplectic transformations is given, valid whenever the symplectic transformation S is nonexceptional, i.e., whenever $\det(\mathbb{1} + S) \neq 0$. The case of an exceptional S is also tackled in Ref. 10 but in an indirect way. In this paper we derive an explicit expression (7.1) or (8.1) which holds for all cases, whether S is exceptional or not. Of course, if we assume S to be nonexceptional, our result simplifies, and we fall back on Huguenin's result [see Eq. (7.2)].

The paper is organized as follows: In Sec. 2 we introduce some definitions and notations, which are essentially those used in Refs. 7 and 11. We also state our results at the end of this section. In Secs. 3–6 we construct a unitary projective representation of the symplectic group using the family of Hilbert spaces mentioned above. In Sec. 7 we dequantize these operators to obtain the corresponding classical functions. Up to Sec. 7 everything is written in intrinsic and coordinate-free notations. In Sec. 8 we show in which way the results can be rewritten in the usual x–p notations. Section 9 contains some applications: calculation of the classical functions for some one-parameter subgroups of the symplectic group; a method for calculating any matrix element of the evolution operator associated to a quadratic Hamiltonian. We end with some remarks.

2. DEFINITIONS AND NOTATIONS

Note: Following A. Grossmann, we are borrowing most of the following notations from D. Kastler, who introduced them in a slightly different setting.[12]

[a] Part of this work was done at the CNRS Marseille.
[b] Research fellow at the Interuniversitair Instituut voor Kernwetenschappen (Research Project 21 EN).

We denote by E a real vector space of even dimension $2n < \infty$. On this vector space a symplectic form σ (i.e., a bilinear, antisymmetric map from $E \times E$ to $\mathbb{R}$) is defined, which we assume to be nondegenerate [i.e., $\sigma(u,v) = 0$, $\forall u \in E \Rightarrow v = 0$]. Using this symplectic form we can define an affine function φ on $E \times E \times E$ [10,13]:

$$\varphi(u,v,w) = 4(\sigma(u,w) + \sigma(w,v) + \sigma(v,u)),$$

which can be interpreted as the surface of the oriented triangle with vertices u,v,w, and which plays a role in the so-called twisted product (see for instance Ref. 11).

We normalize the invariant measure dv on E by requiring $F^2 = \mathbb{1}$, where F is the symplectic Fourier transform

$$(Ff)(v) = 2^{-n} \int dw\, e^{i\sigma(v,w)} f(w).$$

Let $\mathscr{H}$ be the Hilbert space $L^2(E;dv)$.

On $\mathscr{H}$ we define a family of unitary operators $\{W(a); a \in E\}$ by

$$(W(a)\psi)(v) = e^{i\sigma(a,v)}\psi(v-a).$$

These operators $W(a)$ satisfy the relation

$$W(a)\, W(b) = e^{i\sigma(a,b)} W(a+b);$$

hence, they form a representation of the Weyl commutation relations. This representation is not irreducible, but we can build a family of irreducible subrepresentations by introducing complex structures.

A linear map $J: E \to E$ is said to be a σ-allowed complex structure if

$$J^2 = -\mathbb{1},$$

$$\sigma(Jv,Jw) = \sigma(v,w), \quad \forall v,w, \in E,$$

$$\sigma(v,Jv) > 0, \quad \text{if } v \neq 0.$$

For any such σ-allowed complex structure we define the function

$$\Omega_J(v) = \exp\left[-\tfrac{1}{2}\sigma(v,Jv)\right].$$

These Ω_J are elements of $\mathscr{H}$. We define now the following subspaces of $\mathscr{H}$:

$$\mathscr{H}_J = \{\psi \cdot \Omega_J \mid \psi \text{ is holomorphic w.r.t. } J$$

$$(\text{i.e., } \nabla^{Ja}\psi = i\nabla^a\psi, \forall a \in E) \text{ and } \psi\cdot\Omega_J \in \mathscr{H}\}.$$

These $\mathscr{H}_J$ are closed subspaces of $\mathscr{H}$,[14] which are left invariant by the $W(v)$. Furthermore, the restrictions $W_J(v) = W(v)|_{\mathscr{H}_J}$ of the $W(v)$ to the spaces $\mathscr{H}_J$ form irreducible representations of the Weyl commutation relations.[14] (The notations used in Ref. 14 are different from the ones used here. The reader who would want to compare should make the obvious unitary transformation.)

In each of the $\mathscr{H}_J$ we can consider the elements

$$\Omega_J^a = W(a)\Omega_J;$$

they are in fact the coherent states with respect to the choice of complex structure (or equivalently of complex polarization) J. The closed span of the Ω_J^a is the Hilbert space $\mathscr{H}_J$; the Ω_J^a have moreover the following useful "reproducing property"[14,15]:

$$\forall \Psi \in \mathscr{H}_J : \Psi(a) = (\Omega_J^a, \Psi). \tag{2.1}$$

As a result of this any operator A_J on $\mathscr{H}_J$ can be represented by its matrix elements $A_J(a,b) = (\Omega_J^a, A_J \Omega_J^b)$:

$$\Psi \in \mathscr{H}_J \Rightarrow (A_J\Psi)(a) = \int db A_J(a,b)\Psi(b).$$

Because of this property we also call $A(\cdot,\cdot)$ the kernel of the operator A.

Whenever a function f on phase space is given, we can compute its quantal counterpart on the Hilbert space $\mathscr{H}_J$:

$$Q_J(f) = 2^{-n} \int dv (Ff)(v) W_J(-v/2); \tag{2.2}$$

this is the usual Weyl quantization procedure when an irreducible representation of the Weyl commutation relations is given. We can rewrite this expression as[15]

$$Q_J(f) = 2^n \int dv\, f(v) \Pi_J(v), \tag{2.3}$$

where $\Pi_J(v) = W_J(2v)\Pi$, and $(\Pi\,\Psi)(v) = \Psi(-v)$ for any Ψ in $\mathscr{H}$.

Note that both expressions (2.2) and (2.3) can be used to define $Q(f)$ as an operator on the big space $\mathscr{H}$ (at least for reasonable f) which, when restricted to the different $\mathscr{H}_J$, yields $Q_J(f)$ again.[7] The correspondence $f \to Q_J(f)$ can be inverted, i.e., an operator A_J on $\mathscr{H}_J$ can be "dequantized" as follows[11]:

$$f_A(v) = \iint da\, db A_J(a,b) \{b,a|v\}_J, \tag{2.4}$$

with

$$\{b,a|v\}_J = 2^n (\Omega_J^b, \Pi_J(v)\Omega_J^a). \tag{2.5}$$

It is easy to check that the dequantized function of $Q_J(f)$ is always f, regardless of the chosen J.

In these notations our results can be stated as follows: For any symplectic transformation S (i.e., any linear map on E leaving σ invariant; see Sec.4) we have a classical function w_S given by

$$w_{S,J}(v) = (\det[(\mathbb{1} - iJ) + S(\mathbb{1} + iJ)])^{1/2}$$
$$\times \int db \Omega_J(b + Sb - 2v) e^{i\varphi[(b/2),v,(Sb/2)]}$$

[see Eq. (7.1); we have chosen one fixed complex structure J]. Here one can choose either of the two square roots of the determinant. If there is no good reason to do otherwise, we choose the one with argument in $]-\pi/2, \pi/2]$. If $\det(\mathbb{1} + S) \neq 0$, this simplifies to give [see Eq. (7.2)]

$$w_{S,J}(v) = \frac{2^n}{\sqrt{\det(\mathbb{1} + S)}} e^{4i\sigma(v,(\mathbb{1}+S)^{-1}v)},$$

which is the result obtained in Ref. 10.

The operators $W_J(S)$ in $\mathscr{H}_J$ which are quantizations of these functions are given by

$$W_J(S) = 2^{-n}(\det[(\mathbb{1} - iJ) + S(\mathbb{1} + iJ)])^{1/2}$$
$$\times \int db\, |\Omega_J^{Sb})(\Omega_J^b|$$

[see Eq. (7.4)]. Another form of this operator can be found in Sec. 6. These operators form a unitary projective representation of the symplectic group:

$$W_J(S_1)W_J(S_2) = \rho_J(S_1,S_2)W_J(S_1,S_2).$$

The multiplier ρ takes only the values ± 1; given S_1,S_2 it is possible to determine the sign of $\rho(S_1,S_2)$ once one has fixed one's choice of the square roots of the corresponding determinants $\det[(1-iJ)+S(1+iJ)]$ (see Sec. 6).

Moreover, the operators are the representation on the quantum level of the linear canonical transformations on phase space. We have indeed for any symplectic transformation S (the symplectic transformations are in fact just the linear canonical transformations) and for any function f on E:

$$W_J(S)Q_J(f)W_J^*(S) = Q_J(\mathbf{S}f) \quad \text{with } \mathbf{S}f(v) = f(S^{-1}v)$$

[see Eq. (6.4)]. Analogous relations hold for the $w_{S,J}$:

$$w_{S_1,J} \circ w_{S_2,J} = \rho_J(S_1,S_2) w_{S_1,S_2,J},$$

$$w_{S,J} \circ f \circ w^*_{S,J} = \mathbf{S}f,$$

where $\circ$ denotes the twisted product (see, for instance, Ref. 11). These formulas depend on the choice of J. The relation between the $w_{S,J}$ and the $w_{S,J'}$ are given in Sec. 9.

3. PROJECTION OPERATORS ON THE $\mathscr{H}_J$

Since the $\mathscr{H}_J$ are closed subspaces of $\mathscr{H}$, there exist orthogonal projection operators P_J mapping $\mathscr{H}$ to $\mathscr{H}_J$. With the help of the Ω^a_J, these projection operators can be explicitly constructed.

Indeed, since the Ω^a_J span the subspace $\mathscr{H}_J$, we have

$$P_J\Psi = 0 \Longleftrightarrow (\Omega^a_J,\Psi) = 0.$$

On the other hand, we have also Eq. (2.1):

$$P_J\Psi = \Psi \Longleftrightarrow \Psi(a) = (\Omega^a_J,\Psi).$$

It is now obvious that P_J is given by

$$(P_J\Psi)(a) = (\Omega^a_J,\Psi). \tag{3.1}$$

Written more explicitly this means that the projection $P_J\psi$ of any square integrable function ψ on $\mathscr{H}_J$ is given by

$$(P_J\psi)(a) = \int dv\, \overline{\Omega^a(v)}\, \psi(v).$$

This function $P_J\psi$ has automatically the right holomorphy properties.

This can also be written as (in Dirac's bra–ket notation)

$$P_J = \int |\Omega^a_J)da(\Omega^a_J|.$$

Since on the other hand the $\mathscr{H}_J$ are invariant under the $W(v)$, we have

$$P_J W(v) = W_J(v)P_J, \quad \forall v\in E. \tag{3.2}$$

4. THE SYMPLECTIC GROUP AND ITS NATURAL REPRESENTATION IN $L^2(E;dv)$

The symplectic group $\mathrm{Sp}(E,\sigma)$ is defined as the set of real linear maps from E to E which leave σ invariant:

$$S\in\mathrm{Sp}(E,\sigma) \Longleftrightarrow \sigma(Sv,Sw) = \sigma(v,w), \quad \forall v,w\in E.$$

Note that for any given complex structure J, and for any $S\in\mathrm{Sp}(E,\sigma)$, the map SJS^{-1} is again a complex structure. The converse is also true: Whenever two complex structures J, J' are given, there exists a symplectic transformation S in $\mathrm{Sp}(E,\sigma)$ such that $J' = SJS^{-1}$ [For any J one can construct a J-symplectic basis of E, i.e., a basis $\{e_1,...,e_n,f_1,...,f_n\}$ in E such that $\sigma(e_i,e_j) = 0 = \sigma(f_i,f_j)$, $\sigma(e_i,f_j) = \delta_{ij}$ and $f_j = Je_j$. The map S mapping a J-symplectic basis to a J'-symplectic basis is in $\mathrm{Sp}(E,\sigma)$,[16] and satisfies $SJS^{-1} = J'$.]

A symplectic transformation always has determinant 1.[16] Since any complex structure J is obviously in $\mathrm{Sp}(E,\sigma)$, we have in particular $\det J = 1$. This will be used in calculations later on.

As in the case of the Galilei group or the Poincaré group we can define the inhomogeneous symplectic group $\mathrm{ISp}(E,\sigma)$ by taking the semidirect product of $\mathrm{Sp}(E,\sigma)$ with the translation group on E: the elements of $\mathrm{ISp}(E,\sigma)$ are pairs (S,a) with $S\in\mathrm{Sp}(E,\sigma)$, $a\in E$; the product of two such pairs is defined as

$$(S,a)(S',a') = (SS',Sa' + a).$$

The natural representation of $\mathrm{Sp}(E,\sigma)$ on $L^2(E;dv)$ is given by

$$(U_S\Psi)(v) = \Psi(S^{-1}v).$$

This is obviously a unitary representation of $\mathrm{Sp}(E,\sigma)$. Note that the $\mathscr{H}_J$ are not invariant under U_S unless $SJS^{-1} = J$. An easy calculation yields

$$U_S\Omega^a_J = \Omega^{Sa}_{SJS^{-1}}. \tag{4.1}$$

Taking into account the definition (3.1) of the orthogonal projection operators P_J, we see that this implies

$$\begin{aligned}(U_SP_{J'}\Omega^b_J)(a) &= (\Omega^{S^{-1}a}_{J'},\Omega^b_J)\\ &= (\Omega^a_{SJ'S^{-1}},\Omega^{Sb}_{SJS^{-1}})\\ &= (P_{SJ'S^{-1}}U_S\Omega^b_J)(a);\end{aligned}$$

hence,

$$U_S \circ P_{J'}\big|_{\mathscr{H}_J} = P_{SJ'S^{-1}} \circ U_S\big|_{\mathscr{H}_J}. \tag{4.2}$$

It is easy to see that

$$U_S W(v) = W(Sv)U_S. \tag{4.3}$$

Hence, we have also a unitary representation of $\mathrm{ISp}(E,\sigma)$ on $L^2(E;dv)$ given by

$$U_{S,a} = W(a)U_S.$$

5. INTERTWINING OPERATORS BETWEEN THE $\mathscr{H}_J$ (SEE ALSO REF. 7, AND IN A SOMEWHAT DIFFERENT CONTEXT REF. 17)

We will use the natural representation of $\mathrm{Sp}(E,\sigma)$ on $L^2(E;dv)$ to define a projective representation on each $\mathscr{H}_J$. Since the U_S map each $\mathscr{H}_J$ to $\mathscr{H}_{SJS^{-1}}$, we will need some device to map everything back from $\mathscr{H}_{SJS^{-1}}$ to $\mathscr{H}_J$. This device will be given by the maps intertwining the $W_J(v)$: moreover, we will be able to construct these intertwining maps explicitly.

Let any two J, J' be given. Since the $W_J(v)$ form an irreducible representation of the Weyl commutation relations on $\mathscr{H}_J$, and the same is true for the $W_{J'}(v)$ on $\mathscr{H}_{J'}$, von Neumann's theorem tells us there exists a unitary map $T_{J',J}$ from $\mathscr{H}_J$ to $\mathscr{H}_{J'}$ intertwining the $W_J(v)$ and $W_{J'}(v)$. Hence,

$$T_{J',J}W_J(v) = W_{J'}(v)T_{J',J}. \tag{5.1}$$

We proceed now to compute these $T_{J',J}$.

Combining $T_{J',J}^{-1}$ [Eq. (5.1)] $T_{J',J}^{-1}$ with Eq. (3.2), we see that

$$P_{J'}\,|_{\mathcal{H}_J} T_{J',J}^{-1} W_{J'}(v) = P_{J'}\,|_{\mathcal{H}_J} W_J(v) T_{J',J}^{-1} = W_{J'}(v) P_{J'}\,|_{\mathcal{H}_J} T_{J',J}^{-1}\,.$$

Hence, the operator $P_{J'}\,|_{\mathcal{H}_J} T_{J',J}^{-1} \in \mathscr{B}(\mathcal{H}_{J'})$ commutes with all the $W_{J'}(v)$, which implies that it is a multiple of $\mathbb{1}_{\mathcal{H}_{J'}}$, or

$$P_{J'}\,|_{\mathcal{H}_J} = \gamma_{J',J} T_{J',J}\,. \tag{5.2}$$

The constant $\gamma_{J',J}$ is always different from zero: If it were zero, we would have $\mathcal{H}_{J'} \perp \mathcal{H}_J$; hence, $(\Omega_{J'},\Omega_J) = 0$, which is impossible since this inner product is the integral of a strictly positive function. On the other hand, if $|\gamma_{J',J}| = 1$, then $\|P_{J'}\Psi\| = \|\Psi\|$; hence, $P_{J'}\Psi = \Psi$ for any Ψ in $\mathcal{H}_J$, or $\mathcal{H}_{J'} = \mathcal{H}_J$. From Eq. (2.1) we see that this implies that the $\mathcal{H}_J$ are all different ($J' \neq J \Rightarrow \mathcal{H}_{J'} \neq \mathcal{H}_J$), have trivial intersection (this is essentially Schur's lemma), but that no nontrivial vector in $\mathcal{H}_J$ can be orthogonal to all vectors $\mathcal{H}_{J'}$.

Note also that Eq. (5.2) implies that, up to some constant, $P_{J'}P_J$ is a partial isometry in $\mathcal{H}$ with initial subspace $\mathcal{H}_J$ and final subspace $\mathcal{H}_{J'}$ which, as a map from $\mathcal{H}_J$ to $\mathcal{H}_{J'}$, intertwines W_J with $W_{J'}$. From Eq. (5.2) we see that

$$|\gamma_{J',J}|^2 = \|P_{J'}\Omega_J\|^2 = (\Omega_J,P_{J'}\Omega_J) = \int da\,|(\Omega^a_{J'},\Omega_J)|^2\,. \tag{5.3}$$

For the time being, we choose $\gamma_{J',J} = |\gamma_{J',J}|$. This amounts to fixing the up to now undetermined phase factor in $T_{J',J}$.

Putting now $\beta_{J',J} = \gamma_{J',J}^{-1}$ (which we are allowed to do, since $\gamma_{J',J} \neq 0$) we have

$$T_{J',J} = \beta_{J',J} P_{J'}\,|_{\mathcal{H}_J}\,. \tag{5.4}$$

We can use Eq. (5.3) to compute $\beta_{J',J}$; after some calculation [see Eq. (A16)] we get

$$\beta_{J',J} = 2^{-n/2}[\det(J+J')]^{1/4}\,. \tag{5.5}$$

It is obvious from Eq. (5.5) that

$$\beta_{J',J} = \beta_{J',J}\,,$$

$$\beta_{SJ'S^{-1},SJS^{-1}} = \beta_{J',J}, \quad \forall S \in \mathrm{Sp}(E,\sigma)\,.$$

Moreover, if we consider three subspaces $\mathcal{H}_J, \mathcal{H}_{J'}, \mathcal{H}_{J''}$, then the map $T_{J'',J'} \circ T_{J',J}$ is a unitary map intertwining the $W_J(v)$ and the $W_{J''}(v)$. Owing to the irreducibility of the $W_J(v)$, this implies the existence of a phase factor $\alpha(J'',J',J)$ such that

$$T_{J'',J'} \circ T_{J',J} = \alpha(J'',J',J) T_{J'',J}\,. \tag{5.6}$$

With our choice for $\beta_{J',J}$, this α is given by

$$\alpha(J'',J',J) = \|P_{J'},\Omega_J\|^{-1}\|P_J\Omega_{J''}\|^{-1}P_{J''}\Omega_{J'}\|^{-1} \times (P_{J''}\Omega_J,P_{J'}\Omega_J) = \frac{(P_{J''}\Omega_J,P_{J'}\Omega_J)}{|(P_{J''}\Omega_J,P_{J'}\Omega_J)|}\,. \tag{5.7}$$

Since

$$(P_{J''}\Omega_J,P_{J'}\Omega_J) = \int da(\Omega_J,\Omega^a_{J''})(\Omega^a_{J'},\Omega_J) = \int da(\Omega_J{}^{-a},\Omega_{J''})(\Omega_{J'},\Omega_J{}^{-a}) = (\Omega_{J'},P_J\Omega_{J''})\,,$$

we can also write α as

$$\alpha(J'',J',J) = \frac{(\Omega_{J'},P_J\Omega_{J''})}{|(\Omega_{J'},P_J\Omega_{J''})|}\,. \tag{5.7'}$$

In particular, $\alpha(J'',J,J) = \alpha(J,J',J) = \alpha(J',J',J) = 1$.

Note, incidentally, that as a by-product of our reasoning above we have proved that

$$|(\Omega_{J'},P_J\Omega_{J''})| = |(P_{J''}\Omega_J,P_{J'}\Omega_J)| = \|P_{J'}\Omega_J\|\,\|P_J\Omega_{J''}\|\,\|P_{J''}\Omega_{J'}\|\,.$$

Since $\alpha(J,J',J) = 1$, we have

$$T^*_{J',J} = T^{-1}_{J',J} = T_{J,J'}\,. \tag{5.8}$$

Inverting Eq. (5.6) and using Eq. (5.8), we get

$$\alpha(J,J',J'') = \alpha^{-1}(J'',J',J) = \alpha^*(J'',J',J)\,.$$

Combining this with Eq. (5.7') one can easily show that

$$\alpha(J'',J',J) = \alpha(J',J,J'') = \alpha(J,J'',J')\,.$$

From Eq. (5.7) or (5.7') one sees again that

$$\alpha(SJ''S^{-1},SJ'S^{-1},SJS^{-1}) = \alpha(J'',J',J), \quad \forall S \in \mathrm{Sp}(E,\sigma).$$

We have of course also

$$\alpha(J''',J'',J')\,\alpha(J''',J',J) = \alpha(J''',J'',J)\,\alpha(J'',J',J)\,.$$

We can calculate α explicitly from Eq. (5.7') (see Appendix A). The result is

$$\alpha(J'',J',J) = \lim_{\xi \to 1} \left(\exp(i \arg\sqrt{\det(2J + J' + J'' - i\xi\mathbb{1} - i\xi J'J'')}\,)\right) \tag{5.9}$$

Here the argument of the square root of the determinant is determined by the requirement that it be continuous in ξ and equal to zero for $\xi = 0$ (see Appendix A).

6. A PROJECTIVE REPRESENTATION OF THE SYMPLECTIC GROUP ON THE $\mathcal{H}_J$

We have now a device to map from a $\mathcal{H}_J$ to a $\mathcal{H}_{J'}$: It is given by the orthogonal projection operator onto $\mathcal{H}_{J'}$, which, when restricted to $\mathcal{H}_J$, is a unitary map up to some constant we can compute. This device will now be used to define a family of maps $\{V_J(S);\, S \in \mathrm{Sp}(E,\sigma)\}$ which will be unitary maps from $\mathcal{H}_J$ to itself:

$$V_J(S) = T_{J,SJS^{-1}} \circ U_S\,|_{\mathcal{H}_J} = \beta_{J,SJS^{-1}} P_J \circ U_S\,|_{\mathcal{H}_J}\,. \tag{6.1}$$

Here β is given by Eq. (5.5):

$$\beta_{J,SJS^{-1}} = 2^{-n/2}[\det(SJ+JS)]^{1/4}\,.$$

Since both the U_S and the $T_{J',J}$ are unitary, and since $U_S\mathcal{H}_J = \mathcal{H}_{SJS^{-1}}$, the $V_J(S)$ are obviously unitary maps. In some sense they are even the most natural unitary maps in $\mathscr{B}(\mathcal{H}_J)$ representing the symplectic transformations: For any S, we simply apply U_S; since U_S does not leave $\mathcal{H}_J$ invariant in general, we project back onto $\mathcal{H}_J$, and we normalize.

The $V_J(S)$ form a projective representation of $\mathrm{Sp}(E,\sigma)$, and we can even give an expression for the multiplier. Indeed, we have

$$U_S \circ T_{J',J} = \beta_{J',J} U_S \circ P_{J'} \,|_{\mathscr{H}_J} = \beta_{SJ'S^{-1},SJS^{-1}} P_{SJ'S^{-1}} \circ U_S \,|_{\mathscr{H}_J} = T_{SJ'S^{-1},SJS^{-1}} \circ U_S \,|_{\mathscr{H}_J} .$$

Hence,

$$\begin{aligned} V_J(S_1) \circ V_J(S_2) &= T_{J,S_1JS_1^{-1}} \circ U_{S_1} \circ T_{J,S_2JS_2^{-1}} \circ U_{S_2} \,|_{\mathscr{H}_J} = T_{J,S_1JS_1^{-1}} \circ T_{S_1JS_1^{-1},S_1S_2JS_2^{-1}S_1^{-1}} \circ U_{S_1} \circ U_{S_2} \,|_{\mathscr{H}_J} \\ &= \alpha(J,S_1JS_1^{-1},S_1S_2JS_2^{-1}S_1^{-1}) T_{J,S_1S_2J(S_1,S_2)^{-1}} \circ U_{S_1S_2} \,|_{\mathscr{H}_J} \\ &= \alpha(S_1^{-1}JS_1,J,S_2JS_2^{-1}) V_J(S_1S_2) . \end{aligned}$$

So we have indeed a projective representation of $\mathrm{Sp}(E,\sigma)$, with multiplier $\tilde{\alpha}(S_1,S_2) = \alpha(S_1^{-1}JS_1,J,S_2JS_2^{-1})$, where the right-hand side is given by Eqs. (5.7′) and (5.9):

$$\begin{aligned} \tilde{\alpha}(S_1,S_2) = \alpha(S_1^{-1}JS_1,J,S_2JS_2^{-1}) &= \exp\big(i \arg(\Omega_{S_1^{-1}JS_1},P_J\Omega_{S_2JS_2^{-1}})\big) \\ &= \exp\Big(i \arg\Big(\int da(\Omega_J,U_{S_1}\Omega_J^a)(\Omega_J^a,U_{S_2}\Omega_J)\Big)\Big) \\ &= \lim_{\xi \to 1} \Big(\exp\Big(i \arg\sqrt{\det(2J + S_2JS_2^{-1} + S_1^{-1}JS_1 - i\xi \mathbb{1} - i\xi S_1^{-1}JS_1S_2JS_2^{-1})}\Big)\Big) . \end{aligned}$$

Here the argument of the square root of the determinant is determined by the same continuity requirement as at the end of the preceding section.

These $\tilde{\alpha}$ have the usual multiplier property

$$\tilde{\alpha}(S_1,S_2S_3)\tilde{\alpha}(S_2,S_3) = \tilde{\alpha}(S_1,S_2)\tilde{\alpha}(S_1S_2,S_3) .$$

The properties of the $\alpha(J'',J',J)$ listed at the end of the preceding section imply

$$\tilde{\alpha}(S,\mathbb{1}) = 1 \tag{6.2a}$$

or even

$$S_1JS_1^{-1} = J \Rightarrow \tilde{\alpha}(S_1,S_2) = 1 . \tag{6.2b}$$

Also

$$\tilde{\alpha}(S_1^{-1},S_2^{-1}) = \tilde{\alpha}^*(S_2,S_1) , \tag{6.2c}$$

$$\tilde{\alpha}(S,S^{-1}) = 1. \tag{6.2d}$$

The operators $V_J(S)$ thus form a projective representation of $\mathrm{Sp}(E,\sigma)$ which is, however, not the metaplectic representation. In this latter representation one deals in fact with a true representation R of a two-sheeted covering of $\mathrm{Sp}(E,\sigma)$ in which the representation images of the two lifts Σ_1, Σ_2 of the same symplectic operator S differ only by a sign: $R(\Sigma_1) = -R(\Sigma_2)$. This implies that the multiplier of the projective representation of $\mathrm{Sp}(E,\sigma)$ induced by the metaplectic representation takes only the values ± 1, which is not the case for our multiplier $\tilde{\alpha}$. We can, however, reduce our representation above to the metaplectic one. To do this, one should define

$$W_J(S) = \zeta_{J,S} V_J(S) ,$$

where $\zeta_{J,S}$ is a phase factor ($|\zeta_{J,S}| = 1$). These $W_J(S)$ form again a projective representation of $\mathrm{Sp}(E,\sigma)$ with a new multiplier:

$$\rho(S_1,S_2) = \zeta_{J,S_1}\zeta_{J,S_2}\zeta_{J,S_1S_2}^{-1}\tilde{\alpha}(S_1,S_2) .$$

We want this multiplier to take only the values ± 1; hence,

$$[\tilde{\alpha}(S_1,S_2)]^2 = \zeta_{J,S_1S_2}^2 \zeta_{J,S_1}^{*2} \zeta_{J,S_2}^{*2} .$$

So any decomposition of $\tilde{\alpha}^2$ in this form will give us a possibility to reduce our representation to the metaplectic one. However [see Eq. (B5)], one has

$$\begin{aligned} [\tilde{\alpha}(S_1,S_2)]^2 = \exp(i \arg(&\det(\mathbb{1} - iJ) + S_1S_2(\mathbb{1} + iJ) \\ &\cdot\det[(\mathbb{1} + iJ) + S_1(\mathbb{1} - iJ)] \\ &\cdot\det[(\mathbb{1} + iJ) + S_2(\mathbb{1} - iJ)] . \end{aligned}$$

This decomposition has exactly the right form. Moreover,

$$\begin{aligned} |\det[(\mathbb{1} - iJ) + S(\mathbb{1} + iJ)]| &= 2^n[\det(SJ + JS)]^{1/2} \\ &= 2^{2n}\beta_{J,SJS^{-1}}^2 \end{aligned}$$

[see Eq. (B6)].

Hence, we can define

$$\eta_{J,S} = 2^{-n}(\det[(\mathbb{1} - iJ) + S(\mathbb{1} + iJ)])^{1/2} \tag{6.3a}$$

(since $|\eta_{J,S}| = \beta_{J,SJS^{-1}}$, this is always different from zero) and

$$\begin{aligned} W_J(S) &= \exp(i \arg\eta_{J,S}) V_J(S) \\ &= \eta_{J,S} P_J \circ U_S \,|_{\mathscr{H}_J} . \end{aligned} \tag{6.3b}$$

In the definition of $\eta_{J,S}$ we choose the square root with argument in $]-\pi/2,\pi/2]$. [A continuity procedure to determine the phase of this square root would not be unambiguous for all S: There do exist S for which $\det(\mathbb{1} + S) = 0$.] In fact, there is absolutely no reason to prefer the root with positive real part to the one with negative real part. It is just a topological fact of life that it is impossible to choose the signs of the $\eta_{J,S}$ in such a way that the projective representation of $\mathrm{Sp}(E,\sigma)$ becomes a true one. Changing the sign of $\eta_{J,S}$ for a subfamily of $\mathrm{Sp}(E,\sigma)$ means only changing some signs of multipliers where elements of this subfamily occur. We will use this freedom in the choice of the sign of $\eta_{J,S}$ in the treatment of nonexceptional S later on.

Note that our constant $\eta_{J,S}$ leads to the same matrix elements as Bargmann's constant ν_g^2 (see Appendix B).

By construction the $W_J(S)$ form a projective representation of $\mathrm{Sp}(E,\sigma)$ with a multiplier which takes only the values ± 1:

$$W_J(S_1)W_J(S_2) = \rho_J(S_1,S_2)W_J(S_1S_2) ,$$

$$\rho J(S_1,S_2) = \tilde{\alpha}(S_1,S_2)\exp[i \arg(\eta_{J,S_1}\eta_{J,S_2}\eta_{J,S_1S_2}^{-1})] = \pm 1 .$$

Note that the constant $\eta_{J,S}$ depends explicitly on S and not only on SJS^{-1}. Indeed it may happen that $SJS^{-1} = J$, and hence $\beta_{J,SJS^{-1}} = 1$, yet $\eta_{J,S} \neq 1$. As a consequence of this ρ does not inherit $\tilde{\alpha}$'s nice property (6.2b). Properties (6.2c) and (6.2d) also fail to hold in general for ρ: One can find S such that $\eta_{J,S} = i$, and hence $\eta_{J,S^{-1}} = i$, which implies $\rho(S,S^{-1}) = -1$. So the only property of $\tilde{\alpha}$ which passes on to ρ is Eq. (6.2a).

Using Eqs. (4.3) and (5.1) we see that for any $S\in \mathrm{Sp}(E,\sigma)$,

$$W_J(S)W_J(v) = \eta_{J,S}\,\beta^{-1}_{J,SJS^{-1}}T_{J,SJS^{-1}}W_{SJS^{-1}}(Sv)U_S|_{\mathscr{H}_J} = W_J(Sv)W_J(S)$$

or

$$W_J(S)W_J(v)W_J(S)^{-1} = W_J(Sv)\,. \tag{6.4}$$

Combining this with Eq. (2.2) or (2.3), we see that

$$W_J(S)Q_J(f)W_J(S)^{-1} = Q_J(\mathbf{S}f), \tag{6.5}$$

where $\mathbf{S}f$ is the function defined by $(\mathbf{S}f)(v) = f(S^{-1}v)$.

Of course, we can extend all this to the inhomogeneous group $\mathrm{ISp}(E,\sigma)$. We have

$$W_J(S,a) = W_J(a)W_J(S)\,,$$

with

$$W_J(S_1,a_1)W_J(S_2,a_2) = e^{i\sigma(a_1,S_1a_2)}W_J(a_1+S_1a_2)\,\rho(S_1,S_2)W_J(S_1S_2) = e^{i\sigma(a_1,S_1a_2)}\rho(S_1,S_2)W_J((S_1,a_1)(S_2,a_2))\,.$$

Generalizing Eq. (6.4), we get

$$W_J(S,a)W_J(v)W_J(S,a)^{-1} = e^{2i\sigma(a,Sv)}W_J(Sv)$$

or

$$W_J(S,a)\Pi_J(v)W_J(S,a)^{-1} = \Pi_J(Sv+a);$$

hence,

$$W_J(S,a)Q_J(f)W_J(S,a)^{-1} = Q_J((\mathbf{S},\mathbf{a})f), \tag{6.6}$$

with

$$((\mathbf{S},\mathbf{a})f)(v) = f(S^{-1}v - S^{-1}a).$$

Note that, for n even, the operators $W(\pm\mathbb{1},a)$ are the $W(\pm;a)$ introduced in Ref. 13, and that, as was to be expected, this representation $\mathrm{ISp}(E,\sigma)$ is thus an extension of the Wigner–Weyl system as defined in Ref. 13. [For n odd a phase factor has to be introduced: in this case we have indeed $W_J = (-\mathbb{1},0) = i\Pi_J = i\,W(-;0)$.]

From Eqs. (6.4) and (6.5) we see that our operators $W_J(S)$ are exactly the quantal counterparts of the functions w in Ref. 10, up to some phase factor. Hence, we can apply the dequantization procedure given in Ref. 11 to calculate these functions. This will be done in the next section.

7. DEQUANTIZATION OF THE OPERATORS $W_J(S)$ AND $W_J(S,a)$

To apply the dequantization procedure sketched in Eqs. (2.4) and (2.5), we have to compute first the matrix elements of the operators $W_J(S)$ with respect to the coherent states:

$$W_J(S)(a,b) = (\Omega_J^a, W_J(S)\Omega_J^b) = \eta_{J,S}(\Omega_J^a,\Omega^{Sb}_{SJS^{-1}})\,.$$

We calculate now the corresponding function w_S:

$$w_S(v) = 2^n\eta_{J,S}\iint da\,db\,(\Omega_J^b,\Pi(v)\Omega_J^a)(\Omega_J^a,\Omega^{Sb}_{SJS^{-1}}) = 2^n\eta_{J,S}\int db\,(\Omega_J^b,\Pi(v)\Omega^{Sb}_{SJS^{-1}})\,.$$

A straightforward calculation (Appendix C), using $F\Omega_J = \Omega_J$ and formula (6.3) for η, yields

$$w_S(v) = \{\det[(\mathbb{1}-iJ)+S(\mathbb{1}+iJ)]\}^{1/2} \times\int db\Omega_J(b+Sb-2v)e^{i\varphi(b/2,v,Sb/2)}\,, \tag{7.1}$$

where φ is defined in Sec. 2.

Formula (7.1) is valid for any S in $\mathrm{Sp}(E,\sigma)$. If S is exceptional, i.e., if $\mathbb{1}+S$ is singular, we see that for some directions in E the Ω_J factor in the integrand of Eq. (7.1) plays no role, which leaves us with an integral of the phase factor $e^{i\varphi}$, and hence gives us δ functions in the final result. If, however, $\mathbb{1}+S$ is regular, we can always find $u=(\mathbb{1}+S)^{-1}v$ such that $v=(\mathbb{1}+S)u$; hence,

$$\begin{aligned}w_S(v) &= 2^n\eta_{J,S}\int db\Omega_J[(\mathbb{1}+S)(b-2u)]e^{i\varphi(b/2,u+Su,Sb/2)}\\ &= \Big(2^n\eta_{J,S}\int db\Omega_J[(\mathbb{1}+S)b]e^{i\sigma(b,Sb)}\Big)e^{4i\sigma(Su,u)}\\ &= K_S\exp[4i\sigma(v,(\mathbb{1}+S)^{-1}v)]\\ &= K_S\exp\Big[2i\sigma\Big(v,\frac{\mathbb{1}-S}{\mathbb{1}+S}v\Big)\Big]\,.\end{aligned} \tag{7.2}$$

Since S is nonexceptional, we can use our freedom in the choice of a sign for $\eta_{J,S}$ to redefine η as

$$\eta_{J,S} = 2^{-n}\times\lim_{\xi\to 1}\big(\exp\{i\arg\sqrt{\det[\mathbb{1}+S-i\xi J(\mathbb{1}-S)]}\}\big),$$

with again the assumptions that the root of the determinant is continuous in ξ and positive for $\xi=0$. With this choice for the sign of η, we have

$$K_S = 2^n/\sqrt{\det(\mathbb{1}+S)}\,. \tag{7.3}$$

The calculation is given in Appendix A.

Note that the result (7.2) and (7.3) is exacly what was obtained in Ref. 10 for the classical functions corresponding to nonexceptional S.

When S is exceptional, but $J\ker(\mathbb{1}+S)$, we can again simplify formula (7.1) to obtain something anaologous to Eq. (7.2). Indeed, in this case we can decouple the degrees of freedom associated with $\ker(\mathbb{1}+S)$, i.e., we can write E as a direct sum $E=E'\oplus E''$ [$E''=\ker(\mathbb{1}+S)$], such that $\sigma(E',E'')=0$, $JE'=E'$, $JE''=E''$; S can then be considered as a sum $S=S'+S''$, where S' is a non exceptional element of $\mathrm{Sp}(E',\sigma_{E'\times E'})$, and $S''=-\mathbb{1}_{E''}$. Formula (7.1) can then be simplified to give ($v=v'+v''$, with $v'\in E'$, $v''\in E''$)

$$w_S(v) = K_S\delta(v'')\exp[4i\sigma(v',(\mathbb{1}+S)^{-1}v')]\,,$$

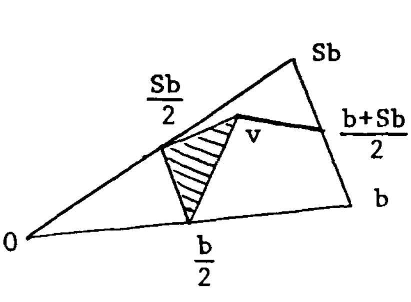

FIG. 1

with

$$K_S = \frac{2^{n'}2^{-n''}}{\sqrt{\det_{E'}(1+S)}}\left(\frac{1-(-1)^{n''}}{2} + i\,\frac{1+(-1)^{n''}}{2}\right).$$

Here $n' = \frac{1}{2}\dim E'$, $n'' = \frac{1}{2}\dim E''$.

The extra factor gives a coefficient 1 if n'' is even, and i if n'' is odd. In particular, we have

$$w_{-1}(v) = 2^{-n}\delta(v)\left(\frac{1-(-1)^n}{2} + i\,\frac{1+(-1)^n}{2}\right).$$

There exist, however, exceptional S for which no J can be found such that $J\ker(1+S) = \ker(1+S)$. For these S, we have to apply directly formula (7.1).

Note that the integrand in the general formula (7.1) has the following nice geometric interpretation: Take the triangle with vertices $0,b,Sb$. The midpoints of the sides of this triangle are $b/2, Sb/2$, and $b+Sb/2$. Then $\varphi(b/2,v,Sb/2)$ is exactly the surface of the oriented triangle $(b/2,v,Sb/2)$, while $\log\Omega_J(b-2v+Sb)$ is $-2\times$ the distance of v to the third midpoint $(b+Sb)/2$ ["distance" being defined with respect to the Euclidean form $s(u,w) = \sigma(u,Jw)$] (see Fig. 1).

We can of course also calculate the functions corresponding to the $W(S,a)$ for the inhomogeneous group; this gives

$$\begin{aligned} w_{S,a}(v) &= 2^n\eta_{J,S}e^{2i\sigma(a,v)}\int db\,\Omega_J(b-2v+a+Sb)e^{i\varphi(b/2,v-a/2,Sb/2)} \\ &= e^{2i\sigma(a,v)}w_S(v-a/2). \end{aligned}$$

As a special case we have the well-known result

$$w_a(v) = w_{1,a}(v) = e^{2i\sigma(a,v)}.$$

Requantization of the functions (7.1) along the procedure sketched in Eq. (2.3) yields (for the detailed calculation, see Appendix C)

$$\begin{aligned} W_J(S) &= 2^n\int dv\,w_S(v)\Pi_J(v) \\ &= 2^{2n}\eta_{J,S}\int dv\int db\,\Omega_J(b+Sb-2v)e^{i\varphi(b/2,v,Sb/2)} \\ &\quad\times\Pi_J(v) \\ &= \eta_{J,S}\int db\,|\Omega_J^{Sb})(\Omega_J^{Sb}| \\ &= 2^{-n}\{\det[(1-iJ)+S(1+iJ)]\}^{1/2} \\ &\quad\times\int db\,|\Omega_J^{Sb})(\Omega_J^{b}|. \end{aligned} \tag{7.4}$$

This is of course again the same operator as given by Eq. (6.1), as one can easily check by comparing the kernels corresponding to Eqs. (6.1) and (7.4).

8. THE TRANSLATION TO x–p NOTATIONS

The translation of our intrinsic notation system to any particular more explicit notation system is completely determined once one has given explicit expressions for E, σ, and J.

Writing everything in coordinate notations amounts to taking

$E = \mathbb{R}^n \oplus \mathbb{R}^n$ (with usually $n = 3N$, N being the number of particles),

$$E \ni v = (x,p),$$

$$\sigma((x,p),(x',p')) = \tfrac{1}{2}(p\cdot x' - x\cdot p'),$$

$$J((x,p)) = (p,-x).$$

Hence, $\Omega_J(v) = \exp[-\frac{1}{4}(x^2+p^2)]$ and

$$dv = \frac{1}{(2\pi)^n}d^nx\,d^np.$$

A symplectic transformation can be represented by a matrix $\begin{pmatrix} A & B \\ C & D\end{pmatrix}$, where A, B, C, D, are real $n\times n$ matrices such that

$$S((x,p)) = (Ax+Bp, Cx+Dp).$$

The fact that S is symplectic is equivalent to

$$\begin{cases} A\,{}^tC - C\,{}^tA = 0, \\ B\,{}^tD - D\,{}^tB = 0, \\ C\,{}^tB - D\,{}^tA = 1. \end{cases}$$

Another explicit but less frequently used notation system is Bargmann's. Here one takes

$$E = \mathbb{C}^n,$$

$$E \ni v = z,$$

$$\sigma(z,z') = \mathrm{Im}(\bar z\cdot z') = \frac{1}{2i}(\bar z\cdot z' - z\cdot\bar z'),$$

$$J(z) = iz.$$

Hence, $\Omega_J(z) = \exp[-\frac{1}{2}(z)^2]$ and

$$dv = (1/\pi^n)d(\mathrm{Re}\,z)d(\mathrm{Im}\,z).$$

9. APPLICATIONS

We have computed the operators $W_J(S)$ of the metaplectic representation on one hand, and on the other hand the corresponding classical functions. Both these results can be used for applications.

A. Applications of the classical function formula

We give here some explicit calculations of the classical function corresponding to a given symplectic transformation. In the first three cases the symplectic transformations form a one-parameter subgroup of $\mathrm{Sp}(E,\sigma)$ which is defined as the classical evolution group for a quadratic Hamiltonian. Since for any quadratic Hamiltonian h the quantum mechanical evolution operator $\exp(iQ_ht)$ is exactly given by $W(S_t)$, where S_t is the one-parameter symplectic transformation group associated to h, one sees that the calculated functions are, at least formally, the twisted exponentials of h (see also Refs. 8 and 18). It is to be noted that one can show, using some recent results,[19] that these functions really are the twisted exponentials (not only formally), i.e., that the series of the twisted exponential makes sense in $\mathscr{S}'$, and does converge (again in $\mathscr{S}'$) to w_{S_t}. This means that the quite complicated proofs (see, for example, Ref. 18) for this convergence in particular cases are no longer necessary.

We give our different results in the x–p notation. Since we are here on the level of the classical functions, the results are independent of the particular representation of the Weyl commutation relations we used:

(1) The harmonic oscillator ($n=1$): $H=\frac{1}{2}(x^2+p^2)$ gives rise to the evolution

$$\begin{cases} x_t = x_0 \cos t + p_0 \sin t, \\ p_t = -x_0 \sin t + p_0 \cos t; \end{cases}$$

hence, $(x_t, p_t) = S_t(x_0, p_0)$, with

$$S_t = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}.$$

Calculating the classical function corresponding to this, we find [we can apply Eq. (7.2) since S_t is nonexceptional whenever $t \neq (2k+1)\pi$; for the special values $t=(2k+1)\pi$ we have $S_t = -\mathbb{1}$ and $w_{S_t} = \frac{1}{2}\delta(x)\delta(p)$.]

$$w_{S_t}(x,p) = (\cos[t/2])^{-1}\exp(-i(x^2+p^2)\tan[t/2]).$$

This is the result found in Refs. 9 and 18.

(2) The same for $H=\frac{1}{2}(p^2-x^2)$ gives

$$S_t = \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix}$$

and

$$w_{S_t}(x,p) = \left(\cosh\frac{t}{2}\right)^{-1} e^{-2i(p^2-x^2)\tanh(t/2)}.$$

(3) The same for $H=\frac{1}{2}p^2+x$ gives $(x_t,p_t)=S_t(x_0,p_0)+a_t$, with

$$S_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

and $a_t = (-\frac{1}{2}t^2, -t)$. We have $w_{S_t}(x,p) = e^{-itp^2}$; hence,

$$w_{S_t,a_t}(x,p) = e^{-2i(p^2t/2)+tx+t^3/8)}.$$

Again these are the same expressions as in Ref. 9.

In our last calculation we treat a "general" exception S. It is general in the sense that no J can be found such that $J\ker(1+S)=\ker(1+S)$, which compels us to use the nonsimplified formula (7.1).

(4) Take $\left(\begin{smallmatrix} -1 & \alpha \\ 0 & -1 \end{smallmatrix}\right)$, with $\alpha>0$. We have $\eta_{J,S} = (i/\sqrt{2})\sqrt{2+i\alpha}$ and

$$\int db\, \Omega_J(b+Sb-2v)e^{i\sigma(b,Sb)+2i\sigma(Sb,v)+2i\sigma(v,b)}$$

$$= (\text{after some calculation})\ \frac{\sqrt{\pi}}{2}\frac{\sqrt{2}}{\sqrt{\alpha}}\frac{1}{\sqrt{\alpha-2i}}$$

$$\times\delta(p)e^{-4ix^2/\alpha}.$$

Hence,

$$w_S(x,p) = \tfrac{1}{2}\sqrt{\pi/\alpha}\, i\sqrt{i}\,\delta(p)e^{-4ix^2\alpha^{-1}}.$$

B. Applications of the expression for $W_J(S)$

We have [see Eq. (6.3b)] $W_J(S) = \eta_{J,S}P_J\circ U_S\,|_{\mathscr{H}_J}$. Hence, for any φ, ψ in $\mathscr{H}_J$,

$$(\varphi, W_J(S)\psi) = \eta_{J,S}\int dv\, \overline{\varphi(v)}\,\psi(S^{-1}v). \tag{9.1}$$

Suppose we are interested in the time evolution operator e^{itH} associated with a quadratic Hamiltonian H. Dequantizing H we get a quadratic function h on phase space, for which the corresponding classical time evolution on phase space is given by a symplectic one-parameter group $(S_h)_t$. It is easy to check that $e^{itH} = W(S_h)_t$. Hence, the matrix elements of the time evolution operator e^{itH} for an at most quadratic Hamiltonian are given by

$$(\varphi, e^{itH}\psi) = \eta_{J,S_{h,t}}\int dv\, \overline{\varphi(v)}\,\psi(S_{h,-t}v). \tag{9.2}$$

This formula is of course only true if the chosen representation of the Weyl commutation relations is a $\mathscr{H}_J$ representation. However, we can use an extension for arbitrary representation spaces.

Indeed, let $\mathscr{H}$ be any Hilbert space carrying an irreducible representation of the Weyl commutation relations [usually one chooses $\mathscr{H}=L^2(\mathbb{R}^n)$ with the Schrödinger representation]. Choose a nice complex structure J on E, σ, and let $\tilde{\Omega}_J \in \mathscr{H}$ be the ground eigenstate of the harmonic oscillator Hamiltonian corresponding to $h_J(v) = s_J(v,v) = \sigma(v,Jv)$. [Usually one takes $J(x,p)=(p,-x)$; hence, $h(v)=\frac{1}{2}(x^2+p^2)$; $\tilde{\Omega}_J$ is then—in the Schrödinger representation—the well-known Hermite function $\pi^{-n/2}\times\exp(-\frac{1}{2}x^2)$.] We define the coherent states $\tilde{\Omega}_J^a$ to be the translated [by $W(a)$] of $\tilde{\Omega}_J$: $\tilde{\Omega}_J^a = W(a)\tilde{\Omega}_J$. For any vector ψ in $\mathscr{H}$ we define the function ϕ_ψ by

$$\phi_{J,\psi}(a) = (\Omega_J^a, \psi)_{\mathscr{H}}.$$

One can easily check that, as a function of a, these $\phi_{J,\psi}$ are elements of $\mathscr{H}_J$. The converse is also true: To any function in $\mathscr{H}_J$ corresponds a unique vector in $\mathscr{H}$ for which the relation above holds. The matrix elements of the evolution operator e^{itH} for any quadratic Hamiltonian $H=Q_h$ are then given by

$$(\varphi, e^{itH}\psi) = \eta_{J,S_{h,t}}\int_E da\, \overline{\phi_{J,\varphi}(a)}\phi_{J,\psi}(S_{h,-t}a). \tag{9.3}$$

So once the classical solutions of the Hamiltonian equations for the Hamiltonian h are known, we can compute any matrix element of the quantum evolution operator for the corresponding Hamiltonian $H=Q_h$. This Hamiltonian H, though at most quadratic, may be quite nontrivial, e.g., a system of N particles, in a homogeneous electromagnetic field (with arbitrary strength), with harmonic oscillator pair potentials, is described by a Hamiltonian falling into this class.

The procedure given above for applying our formula for $W_J(S)$ even if the representation chosen is not a $\mathscr{H}_J$ representation can of course also be applied if one is not interested in one-parameter subgroups but in the whole symplectic group: We can define a projective representation of $\mathrm{Sp}(E,\sigma)$ on any Hilbert space $\mathscr{H}$ carrying an irreducible representation of the Weyl commutation relations

$$(\varphi, W(S)\psi) = \eta_{J,S}\int da\, \overline{\phi_{J,\varphi}(a)}\phi_{J,\psi}(S^{-1}a). \tag{9.4}$$

In the case where $\mathscr{H}=L^2(\mathbb{R}^n)$, with the Schrödinger representation

$$(W(x_a,p_a)\psi)(x) = \exp\left(-\frac{i}{2}x_a p_a\right)e^{ip_a x}\psi(x-x_a),$$

one can check that this yields

$$(\varphi, W(S)\psi) = \iint dx\, dx'\, \overline{\varphi(x)}\, U_S(x,x')\psi(x'),$$

where $U_S(x,x')$ is given, up to a phase factor, by expression (3.27) in Ref. 4 for the cases considered there. The phase factor occurs because we really have a (projective) representation of the whole group $\mathrm{Sp}(E,\sigma)$ while in Ref. 4 only individual symplectic transformations were studied.

10. REMARKS

(1) In the preceding section we showed how one can reconstruct, using our expression in $\mathscr{H}_J$, the metaplectic representation on any Hilbert space carrying an irreducible representation $W(v)$ of the Weyl commutation relation. To do this, we introduced the coherent states (with respect to some J) in $\mathscr{H}$. We can avoid these coherent states in the reconstruction if we use the classical functions w_S: Let Π be the representation on $\mathscr{H}$ of phase space parity ($v \to -v$). Then define $W(S)$ on $\mathscr{H}$ as

$$W(S) = 2^n \int dv\, w_S(v) W(2v)\, \Pi.$$

(2) We have given explicit expression (6.3b) and (7.4) for the operator $W_J(S)$. [In fact, Eq. (9.4) shows us that expression (7.4) is also valid in other representation spaces than $\mathscr{H}_J$.] We can use these expressions to calculate the matrix elements of $W_J(S)$ between coherent states:

$$\begin{aligned} W_J(S)(a,b) &= (\Omega_J^a, W_J(S)\Omega_J^b) \\ &= e^{i\sigma(Sb,a)}(\Omega_J^{a-Sb}, W_J(S)\Omega_J) \\ &= \eta_{J,S} e^{i\sigma(Sb,a)} \int dv\, \overline{\Omega_J^{a-Sb}(v)}\, \Omega_{SJS^{-1}}(v). \end{aligned}$$

Using Eq. (A5) this gives

$$\begin{aligned} W_J(S)(a,b) = (\eta^*_{J,S})^{-1}\exp[i\sigma(Sb,a) - i\sigma(a-Sb,\\ J\hat{Z}(a-Sb)) - \sigma(a-Sb,\hat{Z}(a-Sb))], \end{aligned}$$

with

$$\hat{Z} = -(J + SJS^{-1})^{-1}.$$

It is easy to check that this is in fact the same expression as in Bargmann.[2]

(3) Formula (7.1) for w_S depends on the choice of J. So let us denote for the time being this function by $w_{S,J}$. For two J, J' there exists of course a relation between $w_{S,J}$ and the $w_{S,J'}$. Since one sees easily from Eq. (6.3) that $\eta_{J,S'^{-1}SS'} = \eta_{S'JS'^{-1},S}$, a simple substitution in the integration in Eq. (7.1) gives us the following relation between $w_{S,J'}$ and $w_{S,J}$ (we put $J' = S'JS'^{-1}$):

$$w_{S,J'}(v) = w_{S,S'JS'^{-1}}(v) = w_{S'^{-1}SS',J}(S'^{-1}v). \tag{10.1}$$

On the other hand, we know that for any function f on phase space

$$f(S'^{-1}v) = S'f(v) = (w_{S',J} \circ f \circ w^*_{S',J})(v),$$

where $\circ$ denotes the twisted product (see for instance Refs. 11 and 13). Substituting $w_{S^{-1}SS',J}$ for f, and introducing the multipliers ρ, we get

$$w_{S'^{-1}SS',J}(S'^{-1}v) = \rho_J(S', S'^{-1}SS')\rho_J(SS', S'^{-1}w_{S,J}(v).$$

Combining this with Eq. (10.1), we see that

$$w_{S,J'}(v) = \rho_J(S', S'^{-1}SS')\rho_J(SS', S'^{-1})w_{S,J}(v).$$

So, up to a sign depending on J, J', and S, $w_{S,J'}$ is equal to $w_{S,J}$. If we choose to consider our representation as a double valued representation of $\mathrm{Sp}(E,\sigma)$ instead of as a projective representation, this implies that the double valued representation $S \nrightarrow \pm w_{S,J}$ is independent of J.

(4) Formula (9.4) is only valid for linear canonical transformations. In fact, once the canonical transformation T is nonlinear, there does not exist any more a bounded operator V_T satisfying $\forall f$: $Q_f V_T = V_T Q_{f \circ T^{-1}}$. (This can easily be seen if one realizes that up to a constant this V_T would have to be unitary. One can then use an argument found in Ref. 10 to show that T cannot be linear.) One can of course try to find V_T satisfying the relation above for just n independent functions f_j (see, for example, Ref. 20). The operator constructed in this way is however dependent on the choice of the f_j.

ACKNOWLEDGMENTS

I would like to thank Professor P. Huguenin for several useful remarks. It is a pleasure to thank Professor A. Grossmann for many helpful discussions and for his constant interest and encouragement. Thanks are also due for the hospitality of the CPT2, CNRS Marseille, where part of this work was done.

APPENDIX A

All the calculations in this Appendix are based on the following general principle:

Let B be a real linear map $E \to E$ such that

$$\sigma(u,Bv) = \sigma(v,Bu), \quad \forall u,v \in E, \tag{A1}$$

$$\sigma(u,Bu) > 0, \quad \text{if } u \neq 0, \tag{A2}$$

then the function $\Omega_B(v) = \exp[-\frac{1}{2}\sigma(v,Bv)]$ is integrable, and

$$\int dv \Omega_B(v) = 2^n(\det B)^{-1/2}. \tag{A3}$$

Here we choose the positive square root of det B.

By a simple analyticity argument one can extend (A3) to all complex combinations $B + iC$ of real linear maps from E to E, where B is chosen as above [B satisfies both Eqs. (A1) and (A2)] and C is symmetric [i.e., it satisfies Eq. (A1)]. For any such complex combinations we have again

$$\int dv\, e^{-\sigma(v,Bv)/2} e^{-i\sigma(v,Cv)/2} = 2^n[\det(B+iC)]_c^{-1/2}. \tag{A3'}$$

Here we have introduced the notation $[\det(B+iC)]_c^{\pm 1/2}$ in the following meaning: let $f_\pm : [0,1] \to \mathbb{C}$ be a continuous function with $f_\pm(0) \in \mathbb{R}_+$ and $f_\pm(\xi) = [\det(B + i\xi C)]^{\pm 1/2}$. The continuity of f and its initial value in $\mathbb{R}_+$ select without ambiguity one of the two possible roots of $[\det(B+i\xi C)]^{\pm 1}$ as the value of $f_\pm(\xi)$ at any ξ. Then we define

$$[\det(B+iC)]_c^{\pm 1/2} = f_\pm(1).$$

As usual in Gaussian integrals, the integration variable in Eq. (A3) can be shifted by a complex vector:

$$\int dv\, \Omega_B(v + a + ib) = 2^n(\det B)^{-1/2}, \qquad \text{(A3")}$$

where we define $\sigma(u + iu', v + iv')$ to be the obvious complex linear extension:

$$\sigma(u + iu', v + iv') = \sigma(u,v) - \sigma(u',v') + i\sigma(u',v) + i\sigma(u,v').$$

For any real linear map B satisfying both Eqs. (A1) and (A2), we can construct $\hat{B} = -B^{-1}$ [B is regular because of Eq. (A2)]. It is easy to check that Eqs. (A1) and (A2) are again satisfied by $\hat{B}$. As a corollary of Eq. (A3") we have now

$$\int dv\, e^{i\sigma(a,v)}\Omega_B(v) = \int dv\, e^{i\sigma(\hat{B}a,Bv)}\, e^{-\sigma(v,Bv)/2}$$
$$= \int dv\, e^{-\sigma(v - i\hat{B}a, B(v - i\hat{B}a))/2}\, e^{-\sigma(a,\hat{B}a)/2}$$
$$= 2^n(\det B)^{-1/2}\Omega_{\hat{B}}(a).$$

Finally, note that the family of real linear maps satisfying Eqs. (A1) and (A2) is a convex cone containing the σ-allowed complex structures.

We can now start with our calculations. We begin with $\beta_{J',J}$. Equation (5.3) tells us that $\beta_{J',J}$ is given by

$$\beta_{J',J} = \left(\int da\, |(\Omega^a_{J'},\Omega_J)|^2 \right)^{-1/2}.$$

So we start by calculating $(\Omega^a_{J'},\Omega_J)$. Put $Z = J + J'$. Then

$$(\Omega^a_{J'},\Omega_J) = \int dv\, e^{-i\sigma(a,v)}\Omega_{J'}(a)e^{\sigma(a,J'v)}\Omega_{J'}(v)\Omega_J(v)$$
$$= \Omega_{J'}(a) \int dv\, e^{i\sigma(iJ'a - a,v)}\Omega_Z(v)$$
$$= 2^n(\det Z)^{-1/2}\Omega_{J'}(a)\Omega_{\hat{Z}}(a - iJ'a)$$
$$= 2^n(\det Z)^{-1/2}\, e^{-i\sigma(a,J'\hat{Z}a)}$$
$$\times e^{-\sigma(a,(J' + \hat{Z} + J'\hat{Z}J')a)/2}.$$

Since, however, $Z = J + J'$, and $J^2 = J'^2 = -1$, we have $J'ZJ = -Z$; hence, $J\hat{Z}J' = -\hat{Z}$, or

$$J'\hat{Z}J' = Z\hat{Z}J' - J\hat{Z}J' = -J' + \hat{Z}. \qquad \text{(A4)}$$

This implies

$$(\Omega^a_{J'},\Omega_J) = 2^n(\det Z)^{-1/2}\, e^{-i\sigma(a,J'\hat{Z}a)}\Omega^2_{\hat{Z}}(a). \qquad \text{(A5)}$$

Hence,

$$\int da\, |(\Omega^a_{J'},\Omega_J)|^2 = 2^{2n}(\det Z)^{-1} \int da\, \Omega^4_{\hat{Z}}(a)$$
$$= 2^n(\det Z)^{-1}(\det \hat{Z})^{-1/2}$$
$$= 2^n(\det Z)^{-1/2}.$$

So finally

$$\beta_{J',J} = \left(\int da\, |(\Omega^a_{J'},\Omega_J)|^2 \right)^{-1/2} = 2^{-n/2}(\det Z)^{1/4}$$
$$= 2^{-n/2}[\det(J + J')]^{1/4}. \qquad \text{(A6)}$$

We now compute $\alpha(J'',J',J)$. From Eq. (5.7') we see that

$$\alpha(J'',J',J) = \exp\{i \arg[(\Omega_{J'},P_J\Omega_{J''})]\}.$$

Put $Z_1 = J + J'$, $Z_2 = J + J''$. Using Eq. (A5) we can now calculate $\arg(\Omega_{J'},P_J\Omega_{J''})$:

$$\arg((\Omega_{J'},P_J\Omega_{J''})) = \arg\left[\int da(\Omega_{J'},\Omega^a_J)(\Omega^a_J,\Omega_{J''}) \right]$$
$$= \arg\Big(\int da \exp[-\sigma(a,(\hat{Z}_1 + \hat{Z}_2)a)]$$
$$- i\sigma(a,J(\hat{Z}_2 - \hat{Z}_1)a) \Big).$$

From Eq. (A4) we see that $J\hat{Z}_i = -1 - \hat{Z}_iJ$; hence, $J(\hat{Z}_2 - \hat{Z}_1) = -(\hat{Z}_2 - \hat{Z}_1)J$. Combining this with the fact that J, $\hat{Z}_i$ satisfy Eq. (A1), we see now that both $\hat{Z}_1 + \hat{Z}_2$ and $J(\hat{Z}_2 - \hat{Z}_1)$ fulfill condition (A1), while $\hat{Z}_1 + \hat{Z}_2$ obviously satisfies Eq. (A2). Hence,

$$\arg((\Omega_{J'},P_J\Omega_{J''})) = \arg([\det(\hat{Z}_1 + \hat{Z}_2$$
$$- iJ(\hat{Z}_1 - \hat{Z}_2)]_c^{-1/2}). \qquad \text{(A7)}$$

The determinant in Eq. (A7) can be simplified. Indeed

$$\hat{Z}_1 + \hat{Z}_2 = \hat{Z}_1(-Z_2 - Z_1)\hat{Z}_2,$$
$$J(\hat{Z}_1 - \hat{Z}_2)$$
$$= J\hat{Z}_1(-Z_2 + Z_1)\hat{Z}_2$$
$$= -\hat{Z}_1J(-Z_2 + Z_1)\hat{Z}_2 + \hat{Z}_1Z_1(-Z_2 + Z_1)\hat{Z}_2.$$

Hence,

$$\det(\hat{Z}_1 + \hat{Z}_2 - iJ(\hat{Z}_1 - \hat{Z}_2))$$
$$= \det\hat{Z}_1 \det(-Z_2 - Z_1 - iJZ_2 + iJZ_1 + iZ_1Z_2 - iZ_1Z_1)$$
$$\times \det\hat{Z}_2$$
$$= \det(\hat{Z}_1\hat{Z}_2)\det(2J + J' + J'' + i\mathbb{1} + iJ'J'').$$

Finally,

$$\alpha(J'',J',J) = \exp(i \arg([\det(2J + J' + J''$$
$$- i\mathbb{1} - iJ'J'')]_c^{1/2})). \qquad \text{(A8)}$$

Our last calculation concerns the coefficient in Eq. (7.2).

We have to calculate

$$I = \int db\Omega_J((\mathbb{1} + S)b)e^{i\sigma(b,Sb)}$$
$$= [\det(\mathbb{1} + S)]^{-1} \int db\Omega_J(b)e^{-i\sigma(b,(\mathbb{1}+S)^{-1}b)}$$
$$= [\det(\mathbb{1} + S)]^{-1} \int db \exp\left[-\tfrac{1}{2}\sigma\left(b,Jb + i\frac{\mathbb{1} - S}{\mathbb{1} + S}b\right)\right].$$

One can easily check that $(\mathbb{1} - S)(\mathbb{1} + S)^{-1}$ satisfies Eq. (A1) (see also Ref. 10). Hence,

$$I = \frac{2^n}{\det(\mathbb{1} + S)} \left[\det\left(J + i\frac{\mathbb{1} - S}{\mathbb{1} + S}\right)\right]_c^{-1/2}$$
$$= \frac{2^n}{\sqrt{\det(\mathbb{1} + S)}} (\det[J(\mathbb{1} + S) + i(\mathbb{1} - S)])_c^{-1/2}$$
$$= \frac{2^n}{\sqrt{\det(\mathbb{1} + S)}} (\det[\mathbb{1} + S - iJ(\mathbb{1} - S)])_c^{-1/2}.$$

Combining this with the other coefficient in $w_S(v)$, this yields the result stated in Eq. (7.3).

APPENDIX B

Our first calculation here will be the decomposition of

$\tilde{\alpha}^2$ (see Sec. 6). Before computing this, we derive some simple relations which will turn out to be very useful.

The first of these is

$$J(\mathbb{1} \pm iJ) = J \mp i\mathbb{1} = \mp(\mathbb{1} \pm iJ); \tag{B1}$$

hence,

$$(\mathbb{1} + iJ)(\mathbb{1} - iJ) = (\mathbb{1} - iJ)(\mathbb{1} + iJ) = 0, \tag{B2}$$

$$(\mathbb{1} \pm iJ)^2 = 2(\mathbb{1} \pm iJ). \tag{B3}$$

On the other hand, we already mentioned the existence for any complex structure J of J-symplectic bases, i.e., of bases e_j, f_j of E such that $f_j = Je_j$, $\sigma(e_j,e_k) = \sigma(f_j,f_k) = 0$, $\sigma(e_j,f_k) = \delta_{jk}$. With respect to such a basis J is represented by the matrix

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

Hence, there exists a complex unitary matrix U such that UM_JU^{-1} has the form

$$i\begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix}.$$

Let now L be any linear map from E to itself, with matrix representation M_L w.r.t. a J-symplectic basis. We can write UM_LU^{-1} as $\binom{X\ Y}{Z\ W}$, where X, Y, Z, W are $n \times n$ matrices.

Now,

$$U((\mathbb{1} - iM_J) + M_L(\mathbb{1} + iM_J))U^{-1} = 2\begin{pmatrix} \mathbb{1} & Y \\ 0 & W \end{pmatrix},$$

$$U((\mathbb{1} - iM_J) + (\mathbb{1} + iM_J)M_L)U^{-1} = 2\begin{pmatrix} \mathbb{1} & 0 \\ Z & W \end{pmatrix},$$

$$U((\mathbb{1} - iM_J) + (\mathbb{1} + iM_J)M_L(\mathbb{1} + iM_J))U^{-1} = 2\begin{pmatrix} \mathbb{1} & 0 \\ 0 & 2W \end{pmatrix}.$$

This implies

$$\begin{aligned}\det[(\mathbb{1} - iJ) + L(\mathbb{1} + iJ)] &= \det[U(\mathbb{1} - iM_J) + M_L(\mathbb{1} + iM_J))U^{-1}] \\ &= 2^{2n}\det W\end{aligned}$$

and analogously

$$\det[(\mathbb{1} - iJ) + (\mathbb{1} + iJ)L] = 2^{2n}\det W,$$

$$\det[(\mathbb{1} - iJ) + (\mathbb{1} + iJ)L(\mathbb{1} + iJ)] = 2^{3n}\det W.$$

Hence,

$$\begin{aligned}\det[(\mathbb{1} - iJ) + L(\mathbb{1} + iJ)] &= 2^{-n}\det[(\mathbb{1} - iJ) + (\mathbb{1} + iJ)L(\mathbb{1} + iJ)] \\ &= \det[(\mathbb{1} - iJ) + (\mathbb{1} + iJ)L].\end{aligned} \tag{B4}$$

We can now proceed to compute the decomposition of $\tilde{\alpha}^2(S_1,S_2)$.

Let S_1, S_2 be any symplectic transformations. Define

$$J_1 = S_1^{-1}JS_1, \quad J_2 = S_2JS_2^{-1},$$

$$Z_1 = J + J_1, \quad Z_2 = J + J_2,$$

$$\hat{Z}_1 = -Z_1^{-1}, \quad \hat{Z}_2 = -Z_2^{-1}.$$

Then (see Sec. 6 and Appendix A)

$$\tilde{\alpha}_J^2(S_1,S_2) = \exp\{-i\arg(\det[(\mathbb{1} - iJ)\hat{Z}_1 + (\mathbb{1} + iJ)\hat{Z}_2])\}.$$

Since $J\hat{Z}_1 = -\hat{Z}_1J - \mathbb{1}$ (see Appendix A), we have

$$\begin{aligned}\det[(\mathbb{1} - iJ)\hat{Z}_1 + (\mathbb{1} + iJ)\hat{Z}_2] &= \det[\hat{Z}_1(\mathbb{1} + iJ) + i\mathbb{1} + (\mathbb{1} + iJ)\hat{Z}_2] \\ &= (\det Z_1 \det Z_2)^{-1}\det[-iZ_1Z_2 \\ &\quad + (\mathbb{1} + iJ)Z_2 + Z_1(\mathbb{1} + iJ)] \\ &= (\det Z_1Z_2)^{-1}\det[(\mathbb{1} - iJ_1)Z_2 + (\mathbb{1} + iJ_1)Z_1]\end{aligned}$$

(we have used $-J + Z_1 = J_1$ and $Z_1J = J_1Z_1$). However,

$$\begin{aligned}\det[(\mathbb{1} - iJ_1)Z_2 + (\mathbb{1} + iJ_1)Z_1](\det Z_1)^{-1} &= \det[(\mathbb{1} + iJ_1) + (\mathbb{1} - iJ_1)Z_2(-\hat{Z}_1)] \\ &= \det[(\mathbb{1} + iJ_1) - Z_2\hat{Z}_1(\mathbb{1} - iJ_1)] \quad \text{[use Eq. (B4)]} \\ &= \det[(\mathbb{1} + iJ_1)Z_1 + Z_2(\mathbb{1} - iJ)](\det Z_1)^{-1} \\ &= \det[Z_1(\mathbb{1} + iJ) + Z_2(\mathbb{1} - iJ)](\det Z_1)^{-1}.\end{aligned}$$

Hence,

$$\tilde{\alpha}_J^2(S_1,S_2) = \exp\{-i\arg(\det[Z_1(\mathbb{1} + iJ) + Z_2(\mathbb{1} - iJ)])\}.$$

We have

$$\begin{aligned}[Z_1(\mathbb{1} + iJ) + Z_2(\mathbb{1} - iJ)]J &= -i(J + J_1)(\mathbb{1} + iJ) + i(J + J_2)(\mathbb{1} - iJ) \\ &= -i(-2i\mathbb{1} + J_1 - J_2 + iJ_1J + iJ_2J) \\ &= -(\mathbb{1} + iJ_1 + iJ - J_1J + \mathbb{1} - iJ_2 - iJ - J_2J) \\ &= -(\mathbb{1} + iJ_1)(\mathbb{1} + iJ) - (\mathbb{1} - iJ_2)(\mathbb{1} - iJ).\end{aligned}$$

Hence,

$$\det[Z_1(\mathbb{1} + iJ) + Z_2(\mathbb{1} - iJ)] = \det[(\mathbb{1} + iJ_1)(\mathbb{1} + iJ) + (\mathbb{1} - iJ_2)(\mathbb{1} - iJ)]$$

(remember that $\det J = \det S_1 = \det S_2 = 1!$, see Sec. 4) and

$$\begin{aligned}&= \det[S_1^{-1}(\mathbb{1} + iJ)S_1(\mathbb{1} + iJ) \\ &\quad + S_2(\mathbb{1} - iJ)S_2^{-1}(\mathbb{1} - iJ)] \\ &= 2^{-2n}\det\{[S_1^{-1}(\mathbb{1} + iJ) + S_2(\mathbb{1} - iJ)] \\ &\quad \times [(\mathbb{1} + iJ)S_1(\mathbb{1} + iJ) + (\mathbb{1} - iJ)S_1^{-1}(\mathbb{1} - iJ)]\} \\ &= 2^{-4n}\det[S_1^{-1}(\mathbb{1} + iJ) + S_2(\mathbb{1} - iJ)] \\ &\quad \times \det\{[(\mathbb{1} + iJ)S_1(\mathbb{1} + iJ) + (\mathbb{1} - iJ)] \\ &\quad \times [(\mathbb{1} + iJ) + (\mathbb{1} - iJ)S_2^{-1}(\mathbb{1} - iJ)]\} \\ &= 2^{-3n}\det[(\mathbb{1} + iJ) + S_1S_2(\mathbb{1} - iJ)] \\ &\quad \times \det[(\mathbb{1} - iJ) + S_1(\mathbb{1} + iJ)]\{\det[(\mathbb{1} - iJ) \\ &\quad + S_2^{-1}(\mathbb{1} + iJ)]\}^* \\ &= 2^{-2n}\det[(\mathbb{1} + iJ) + S_1S_2(\mathbb{1} - iJ)] \\ &\quad \times \{\det[(\mathbb{1} + iJ) + S_1(\mathbb{1} - iJ)]\}^* \\ &\quad \times \{\det[(\mathbb{1} + iJ) + S_2(\mathbb{1} - iJ)]\}^*.\end{aligned}$$

So finally

$$\begin{aligned}\tilde{\alpha}_J^2(S_1,S_2) = \exp\{i\arg(\det[(\mathbb{1} - iJ) + S_1S_2(\mathbb{1} + iJ)] \\ \times \det[(\mathbb{1} + iJ) + S_1(\mathbb{1} - iJ)] \\ \times \det[(\mathbb{1} + iJ) + S_2(\mathbb{1} - iJ)])\}.\end{aligned} \tag{B5}$$

This is exactly the decomposition of $\tilde{\alpha}^2$ as used in Sec. 6.

Our next calculation is the computation of $|\det[(\mathbb{1}-iJ)+S(\mathbb{1}+iJ)]|$:

$$
\begin{aligned}
&|\det[(\mathbb{1}-iJ)+S(\mathbb{1}+iJ)]|^2\\
&=\det[(\mathbb{1}-iJ)+S(\mathbb{1}+iJ)]\det[(\mathbb{1}+iJ)+(\mathbb{1}-iJ)S]\\
&=\det\{[(\mathbb{1}-iJ)+S(\mathbb{1}+iJ)]\\
&\quad\times J[(\mathbb{1}+iJ)+(\mathbb{1}-iJ)S]\}\\
&=\det[2i(\mathbb{1}-iJ)S-2iS(\mathbb{1}+iJ)]\\
&=2^{2n}\det(iS+JS-iS+SJ)\\
&=2^{2n}\det(JS+SJ).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&|\det[(\mathbb{1}-iJ)+S(\mathbb{1}+iJ)]|\\
&\quad=2^n[\det(SJ+JS)]^{1/2}.
\end{aligned}
\tag{B6}
$$

Finally, we give here the connection with Bargmann's constant $(\det\lambda)^{-1/2}$.[2] We introduce the x–p notation (see also Sec. 8): $S((x,p))=(Ax+Bp,Cx+Dp)$. In Bargmann's notations one has $\lambda=\frac{1}{2}(D+A+iB-iC)$, and $\nu_g=(\det\lambda)^{-1/2}=2^{n/2}[\det(A+D+iB-iC)]^{-1/2}$. This constant ν_g is in fact the matrix element $(\Omega_J,W_J(S)\Omega_J)$ (see Ref. 2). We have

$$
\begin{aligned}
(\Omega_J,W_J(S)\Omega_J)&=\eta_{J,S}(\Omega_J,\Omega_{SJS}\;,)\\
&=\eta_{J,S}\,\beta^{\;2}_{J,SJS}\;,=(\eta^*_{J,S})^{-1}\\
&=2^n\{\det[(\mathbb{1}+iJ)+S(\mathbb{1}-iJ)]\}^{-1/2}.
\end{aligned}
$$

However,

$$
S=\begin{pmatrix}A&B\\C&D\end{pmatrix},\quad J=\begin{pmatrix}0&\mathbb{1}\\-\mathbb{1}&0\end{pmatrix};
$$

hence,

$$
\begin{aligned}
&\det(\mathbb{1}+iJ)+S(\mathbb{1}-iJ)\\
&=\det\begin{pmatrix}\mathbb{1}+A+iB & i\mathbb{1}+B-iA\\ -i\mathbb{1}+C+iD & \mathbb{1}+D-iC\end{pmatrix}\\
&=\det\begin{pmatrix}A+D+i(B-C) & -i(A+D)+B-C\\ -i\mathbb{1}+C+iD & \mathbb{1}+D-iC\end{pmatrix}\\
&=\det\begin{pmatrix}A+D+i(B-C) & 0\\ -i\mathbb{1}+C+iD & 2\mathbb{1}\end{pmatrix}\\
&=2^n\det(A+D+iB-iC).
\end{aligned}
$$

So

$$
(\Omega_J,W_J(S)\Omega_J)=2^{n/2}(\det(A+D+iB-iC))^{-1/2}.
$$

Comparing this result with Bargmann's (B7) we see that they coincide, as was to be expected.

APPENDIX C

We give here the details of the calculation leading to formula (7.1):

$$
w_S(v)=2^n\eta_{J,S}\int db\,(\Omega^b_J,\Pi(v)\,\Omega^{Sb}_{SJS}\;,),
\tag{C1}
$$

with

$$
\begin{aligned}
&\int db\,(\Omega^b_J,\Pi(v)\,\Omega^{Sb}_{SJS}\;,)\\
&\quad=\int db\,(\Pi(v)\Omega^b_J,\Omega^{Sb}_{SJS}\;,)\\
&\quad=\int db\int dc\,e^{2i\sigma(v,b)}\overline{\Omega^{2v-b}_J(c)}\,\Omega^{Sb}_{SJS}\;,(c)\\
&\quad=\int db\int dc\,e^{2i\sigma(v,b)}e^{-2i\sigma(v,c)+i\sigma(b,c)}\\
&\qquad\times e^{i\sigma(Sb,c)}\Omega_J(2v-b-c)\Omega_J(b-S^{-1}c)\\
&\quad=\int db\int dc\,e^{2i\sigma(v,b)}\\
&\qquad\times e^{-2i\sigma(v,Sc)+i\sigma(b,Sc)+i\sigma(b,c)}\Omega_J(2v-b-Sc)\Omega_J(b-c)\\
&\quad=\int db\int dc\,e^{i\sigma(2v-c-Sc,b)}\\
&\qquad\times e^{-2i\sigma(v,Sc)}\Omega_J\left(\sqrt{2b}+\frac{Sc-c-2v}{\sqrt{2}}\right)\\
&\qquad\times\Omega_J\left(\frac{Sc+c-2v}{\sqrt{2}}\right)\\
&\quad=2^{-n}\int dc\left(e^{-2i\sigma(v,Sc)}e^{-i\sigma(2v-c-Sc,Sc-c-2v)/2}\right.\\
&\qquad\left.\times\Omega_J\left(\frac{Sc+c-2v}{\sqrt{2}}\right)\int db\,e^{i\sigma(2v-c-Sc,b)/\sqrt{2}}\Omega_J(b)\right)\\
&\quad=\int dc\,e^{-2i\sigma(v,Sc)}e^{i\sigma(2v-Sc,c)}\Omega^2_J\left(\frac{Sc+c-2v}{\sqrt{2}}\right)\\
&\quad=\int dc\,e^{i\varphi(c/2,v,Sc/2)}\Omega_J(Sc+c-2v).
\end{aligned}
$$

Combining this result with Eq. (C1), we get formula (7.1).

For the requantization of $w_S(v)$ we have to calculate

$$
\begin{aligned}
&2^n\int dv\,w_S(v)\Pi(v)\\
&\quad=2^{2n}\eta\int dv\int db\,\Omega_J(2v-c-Sc)e^{i\varphi(c/2,v,Sc/2)}\Pi(v).
\end{aligned}
\tag{C2}
$$

We give here the calculation of this integral:

$$
\begin{aligned}
I&=\int dv\int db\,\Omega_J(2v-c-Sc)\,e^{i\varphi(b/2,v,Sb/2)}\Pi_J(v)\\
&=\int dv\int db\,\Omega_J(-2v)\,e^{i\varphi(b,2v+b+Sb,Sb)/4}\\
&\quad\times W_J(2v+b+Sb)\Pi\\
&=\int dv\int db\,\Omega_J(2v)\,e^{-i\sigma(b,Sb)+2i\sigma(Sb,v)+2i\sigma(v,b)}\\
&\quad\times W_J(b+Sb)W_J(2v)\Pi\,e^{2i\sigma(v,b+Sb)}\\
&=\int db\,W_J(b+Sb)\,e^{-i\sigma(b,Sb)}\int dv\,\Omega_J(2v)\,e^{4i\sigma(v,b)}\Pi(v).
\end{aligned}
$$

Using the notations of Ref. 11, we have

$$
\Omega_J e^{4i\sigma(v,b)}=2^{-n}\{b,-b\mid v\};
$$

hence (see Ref. 11, Sec. 5.B.1),

$$
\begin{aligned}
\int dv\,\Omega_J(2v)\,e^{4i\sigma(v,b)}\Pi_J(v)&=2^{-n}\int dv\{b,-b\,|v\}\Pi_J(v)\\
&=2^{-2n}Q_J(\{b,-b\,|\cdot\})\\
&=2^{-2n}|\Omega^{-b}_J)(\Omega^b_J|.
\end{aligned}
$$

This implies

$$I = 2^{-2n} \int db\, W_J(Sb) W_J(b) |\Omega_J^{-b})(\Omega_J^b|$$
$$= 2^{-2n} \int db\, W_J(Sb) |\Omega_J)(\Omega_J^b|$$
$$= 2^{-2n} \int db\, |\Omega_J^{Sb})(\Omega_J^b| \,.$$

Combining this with Eq. (C2), we get Eq. (7.4).

[1]I.E. Segal, Math. Scand. **13**, 31 (1963).
[2]V. Bargmann, "Group representations in Hilbert spaces of analytic functions," in *Analytic Methods in Physics*, edited by R.P. Gilbert and R.G. Newton (Gordon and Breach, New York, 1968).
[3]C. Itzykson, Commun. Math. Phys. **4**, 92 (1967).
[4]M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971); P. Kramer, M. Moshinsky, and T.H. Seligman, "Complex extensions of canonical transformations and quantum mechanics," in *Group Theory and Its Applications*, (Academic Press, New York, San Francisco, and London, 1975) Vol. III and other papers by these authors.
[5]M.H. Boon and T.H. Seligman, J. Math. Phys. **14**, 1224 (1973).
[6]J. Manuceau, A. Verbeure, Commun. Math. Phys. **9**, 293 (1968).
[7]A. Grossman, "Geometry of real and complex canonical transformations in quantum mechanics," Talk given at the VIth International Colloquium on Group Theoretical Methods, Tübingen, 18–22, July 1977.
[8]G. Burdet and M. Perrin, J. Math. Phys. **16**, 2172 (1975); Lett. Math. Phys. **2**, 93 (1977).
[9]Ph. Combe, R. Rodriguez, M. Sirugue-Collin, and M. Sirugue, "Classical functions associated with some groups of automorphisms of the Weyl group," Talk given at the Vth International Colloquium on Group Theoretical Methods, Montreal.
[10]P. Huguenin, Lett. Math. Phys. **2**, 321 (1978); Mécanique de l'espace de phase, Cours 3ème cycles, Neuchâtel (1975).
[11]I. Daubechies and A. Grossmann, "An integral transform related to quantization," J. Math. Phys. **21** (1980).
[12]D. Kastler, Commun. Math. Phys. **1**, 14 (1965).
[13]A. Grossmann and P. Huguenin, Helv. Phys. Acta **51**, 252 (1978).
[14]V. Bargmann, Commun. Pure Appl. Math. **48**, 187 (1961).
[15]A. Grossmann, Commun. Math. Phys. **48**, 191 (1976).
[16]E. Artin, *Geometric Algebra* (Interscience, New York and London, 1957).
[17]I. Satake, Adv. Math. 7, 83 (1971).
[18]F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Ann. Phys. (Leipzig) **111**, 111 (1978).
[19]F. Hansen, "An algebraic approach to quantum mechanics," preprint (1979).
[20]P.A. Mello and M. Moshinsky, J. Math. Phys. **11**, 2017 (1975).

The hydrogen atom: Quantum mechanics on the quotient of a conformally flat manifold

G. A. Ringwood
Department of Physics, University of Antwerpen, and Department of Mathematics, University of Durham
J. T. Devreese
Institute for Applied Mathematics and Department of Physics, University of Antwerpen (R. U.C.A. and U.I.A.) Universiteitsplein 1, 2610 Wilrijk, Belgium

(Received 13 December 1979; accepted for publication 22 February 1980)

The regularization of the Kepler problem proposed by Kustaanheimo and Stiefel provides an example of quantum mechanics on the quotient of a conformally flat manifold.

1. INTRODUCTION

The quantization of a particle on a Riemannian manifold described by the classical Lagrangian

$$\tfrac{1}{2}g_{\alpha\beta}\dot{q}^\alpha\dot{q}^\beta - V(q) \tag{1.1}$$

is of interest, particularly in view of its repercussions in field theory. Despite the eminent reasonableness of the Schrödinger equation

$$i\partial\psi/\partial t = (-\tfrac{1}{2}\Delta_2 + V(q))\psi,$$

where Δ_2 is the scalar Laplace operator, as Dowker[1] says, "Logically there is no reason why the Schrödinger equation should take this particular covariant form nor, indeed, why it should be covariant at all." Dowker[1] goes on to say that the pragmatic approach to quantization is to guess the Schrödinger equation and compare it with physical reality.

Duru and Kleinert[2] describe a path integral solution for the hydrogen atom. This solution, although inspired, is rather cavalier, both in its treatment of the underlying transformation of Kustaanheimo and Stiefel[3] and the well known ambiguities of the path integral formulation.[4,5] This comment reinterprets the solution without the use of path integrals and then goes on to give the skeletal form of the path integral omitted by Duru and Kleinert.[2] The interpretation provides an example of a physical situation, all be it once removed, in which a particle moves on a Riemannian space which is conformally flat. The Schrödinger equation takes a conformally invariant form. The physical situation is recovered as a quotient of this conformally flat space so this solution also provides a rare example of quantum mechanics on a quotient space.[6]

2. THE HYDROGEN ATOM ONCE REMOVED

In order to regularize the Kepler problem Kustaanheimo and Stiefel[3] construct a map from R^4 into physical space R^3. The map is given by $x^i = A^i{}_\alpha q^\alpha$ (Roman indices run from 1 to 3 and Greek indices run from 1 to 4), where A is the matrix

$$\begin{pmatrix} q^3 & q^4 & q^1 & q^2 \\ -q^2 & q^1 & q^4 & q^3 \\ -q^1 & q^2 & q^3 & -q^4 \end{pmatrix}.$$

In terms of polar coordinates

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = r\begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix},$$

$$\begin{pmatrix} q^1 \\ q^2 \\ q^3 \\ q^4 \end{pmatrix} = r^{1/2}\begin{pmatrix} \sin\frac{1}{2}\theta\cos\frac{1}{2}(\alpha+\phi) \\ \sin\frac{1}{2}\theta\sin\frac{1}{2}(\alpha+\phi) \\ \cos\frac{1}{2}\theta\cos\frac{1}{2}(\alpha-\phi) \\ \cos\frac{1}{2}\theta\sin\frac{1}{2}(\alpha-\phi) \end{pmatrix}. \tag{2.1}$$

It is clear that points on the circle in R^4 parameterized by α are mapped onto a single point in R^3. Thus R^3 can be realized as a quotient of R^4 by a one-parameter subgroup, SO2, of the full rotation group SO4. (Note that the quotient for $r = 1$ is a quotient of spheres: $S^3/S^1 = S^2$.) Dowker[6] describes how to treat quantum mechanics on quotient spaces. Let M denote a manifold, H a Lie group of transformations of M, and M/H the quotient of M by H. If q and h are representatives of M and H, respectively, let hq denote the action of H on M. The Green function on M/H can be described in terms of the Green function on M by $G_{M/H}(q,q';E) = \int_H dh\, G_M(hq,q';E)$.

3. QUANTUM MECHANICS ON CONFORMALLY FLAT SPACE

The classical Coulomb Lagrangian on R^3

$$\tfrac{1}{8}\dot{x}^i\dot{x}^i + \mu r^{-1}, \quad \mu > 0,$$

where for a reason which will be apparent in the next equation the mass is taken as $\frac{1}{4}$, becomes

$$\tfrac{1}{2}q^2\dot{q}^\alpha\dot{q}^\alpha + \mu q^{-2}$$

on R^4, where $q^2 = q^\alpha q^\alpha$, and this is of the form (1.1), where the metric is conformally flat. The scalar curvature R is $-18q^{-4}$ so that the Lagrangian takes the scalar form

$$\tfrac{1}{2}g_{\alpha\beta}\dot{q}^\alpha\dot{q}^\beta + \mu(-R/18)^{1/2}.$$

The canonical momenta of the system are $p_\alpha = q^2\dot{q}^\alpha = g_{\alpha\beta}\dot{q}^\beta$ and the classical Hamiltonian is

$$\tfrac{1}{2}g^{\alpha\beta}p_\alpha p_\beta - \mu(-R/18)^{1/2}.$$

The conformal nature of the classical free particle leads one to guess that the quantum free particle should possess conformal invariance. With this in mind the wavefunctions

 0022-2488/80/061390-03$1.00

are taken to be tensor densities of weight $\frac{1}{4}$ so that the normalization condition

$$\int d^4q\, g^{1/4}|\psi|^2 = 1$$

is somewhat unconventional. The Laplacian for a tensor density of weight $\frac{1}{4}$ is $g^{1/8}\Delta_2 g^{-1/8}$, where $\Delta_2 = g^{-1/2}\partial_\alpha g^{\alpha\beta} g^{1/2}\partial_\beta$. The quantum free particle Hamiltonian is taken to be

$$-\tfrac{1}{2}(g^{1/8}\Delta_2 g^{-1/8} - R/6) = -\tfrac{1}{2}q^{-2}\partial_\alpha^2,$$

where the additional term in the curvature makes the Hamiltonian conformally invariant. That is, under a conformal transformation $g \rightarrow \Omega^2 g$ in four dimensions

$$(g^{1/8}\Delta_2 g^{-1/8} - R/6) \rightarrow \Omega^{-2}(g^{1/8}\Delta_2 g^{-1/8} - R/6).$$

Although the forms of the normalization and the Hamiltonian were argued on invariance principles, the choice is ultimately dictated by the known solution of the problem on the quotient space R^3. It will be shown that this Hamiltonian gives the known answer to the Coulomb problem. The free particle Hamiltonian is Hermitian only with the normalization employed above. (The quantum momenta must then take the Hermitian forms $\hat{p}_\alpha = -ig^{-1/8}\partial_\alpha g^{1/8}$.) The above is the only choice which makes the quantum system covariant.

With the free particle Hamiltonian as given the Green function for a particle in a "Coulomb" potential satisfies

$$(E + \tfrac{1}{2}q^{-2}\partial_\alpha^2 + \mu q^{-2})G(q, q'; E) = \delta(q - q')g'^{-1/4}, \tag{3.1}$$

where the form of the delta distribution on the right hand side of the equation is due to the normalization. For the classical system Kustaanheimo and Stiefel show that the solution of the "Coulomb" problem is essentially harmonic. The solution of (3.1) can be given in terms of the heat kernel for the four dimensional harmonic oscillator $K(q, q'; s)$, where s is inverse temperature. The identity

$$0 = \int_0^\infty ds \exp(\mu s)(-\partial/\partial s + \tfrac{1}{2}\partial_\alpha^2 - \tfrac{1}{2}\omega^2 q^2)K(q, q'; s),$$

valid for $\mu < 2\omega$, can be rearranged to give

$$[-\tfrac{1}{2}\omega^2 - (-\tfrac{1}{2}q^{-2}\partial_\alpha^2 - \mu q^{-2})]\int_0^\infty ds\,(-)\exp(\mu s)K(q,q';s) = g^{-1/4}\delta(q - q'),$$

and with the identification $E = -\frac{1}{2}\omega^2$ (the energy is taken to be negative) the integral expression is the Green function in (3.1):

$$G(q,q';E) = -\int_0^\infty ds \frac{\omega^2}{4\pi^2 sh^2\omega s}\exp\left\{\mu s - \frac{\omega}{2sh\omega s} \times [(q^2 + q'^2)ch\omega s - 2q^\alpha q'^\alpha]\right\}.$$

(The positive energy solution can be obtained by analytic continuation.)

Integrating over the range of α, $(0, 4\pi)$, the Green function on the physical space R^3 then becomes (the scale of the α integral is), from (2.1),

$$G(r,r';E) = -\frac{1}{16}\int_0^\infty ds \frac{\omega^2}{\pi sh^2\omega s} I_0\left(\frac{\omega}{sh\omega s}(rr')^{1/2}\cos\tfrac{1}{2}\theta\right) \times \exp\left[\mu s - \frac{\omega ch\omega s}{2sh\omega s}(r + r')\right],$$

where θ is the angle between the vectors **r** and **r**′.

Changing the variable of integration, $\zeta = ch\omega s/sh\omega s$, the Green function is

$$\frac{-p_0}{8\pi}\int_1^\infty d\zeta \left(\frac{\zeta + 1}{\zeta - 1}\right)^\gamma I_0(2p_0(\zeta^2 - 1)^{1/2}(rr')^{1/2}\cos\tfrac{1}{2}\theta) \times \exp[-\zeta p_0(r + r')], \tag{3.2}$$

where $\gamma = \mu/2\omega$ and $p_0^2 = -E/2$, valid for $\gamma < 1$. For values of γ outside this range the integral (3.2) can be replaced by a contour integral in the positive sense around the cut from 1 to ∞ on the real axis:

$$-\int_1^\infty d\zeta \rightarrow \frac{\exp(i\pi\gamma)}{2i\sin\pi\gamma}\oint d\zeta.$$

Then (3.2) with this modification is the Green function for the hydrogen atom given by Hostler.[7] This confirms that the choice of Hamiltonian and normalization are correct.

4. THE SKELETAL PATH INTEGRAL

The propagator on R^4 can be written formally as

$$g^{1/8}\exp[i\tau(\tfrac{1}{2}\Delta_2 - V^*)]g^{-1/8}\delta(q - q')g'^{-1/4},$$

where the potential V^* is

$$R/12 - \mu(-R/18)^{1/2}.$$

The path integral is given by a "folding" together of small time propagators. Replacing the delta distribution by its Fourier decomposition

$$(2\pi)^{-4}\int d^4p \exp(ip_\alpha \Delta q^\alpha),$$

where $\Delta q^\alpha = q^\alpha - q'^\alpha$, and performing the usual procedure of evaluating $\exp[i\tau(\frac{1}{2}\Delta_2 - V^*)]$ by its small time approximation $1 + i\tau(\frac{1}{2}\Delta_2 - V^*)$ and then exponentiating, the small time propagator is given by

$$\int d^4p\,(2\pi)^{-4}g'^{-1/4}\exp[ip_\alpha\Delta q^\alpha - i\tau(\tfrac{1}{2}g^{\alpha\beta}p_\alpha p_\beta - \mu(-R/18)^{1/2})]. \tag{4.1}$$

The phase space path integral is then

$$\int \mathscr{D}^4q\, g^{1/4}\int \mathscr{D}^4p\, g^{-1/4}\exp\left[i\int dt\, p\dot{q} - (\tfrac{1}{2}q^{-2}p^2 - \mu q^{-2})\right],$$

which has the classical Hamiltonian in the exponential. This may be unexpected in view of the discussion of Dowker.[5] It is, however, the form used by Duru and Kleinert[2] and (4.1) gives the skeletal decomposition. Performing the p integration in (4.1) gives the skeletal components of the coordinate path integral

$$(2\pi i\tau)^{-2}g^{1/2} \times \exp\left[\frac{i}{2\tau}g_{\alpha\beta}\Delta q^\alpha\Delta q^\beta + i\tau\mu(-R/18)^{1/2}\right]g'^{-1/4}.$$

ACKNOWLEDGMENT

One of the authors (GAR) wishes to express his gratitude to A. Lagendijk.

[1]J.S. Dowker, J. Phys. **A 7**, 1256 (1974).
[2]H. Duru and H. Kleinert, Phys. Lett **B 84**, 185 (1979).
[3]P. Kustaanheimo and E. Stiefel, J. Reine Angew. Math. **218**, 204 (1965).
[4]R. Fanelli, J. Math. Phys. **16**, 1729 (1975).
[5]J.S. Dowker, J. Math. Phys. **17**, 1873 (1976).
[6]J.S. Dowker, J. Phys. **A 5**, 936 (1972).
[7]L.H. Hostler, J. Math. Phys. **5**, 591 (1964).

Decaying systems with degenerate livsic matrix[a)]

E. Katznelson
Department of Physics and Astronomy, Tel Aviv University, Ramat-Aviv, Israel

(Received 13 December 1978; accepted for publication 1 May 1979)

The spectral analysis given by Wong for the resolvent of a non-self-adjoint operator with arbitrary multiplicity is utilized for the description of the time evolution of an unstable system. After studying the case for which the operator is independent of the resolvent variable z, the Wong analysis is extended to the physically interesting case for which the operator depends on z. The case of infinite multiplicity is treated, and it is found that the flow of probability through the generalized eigenstates is analogous to the approach to equilibrium in statistical mechanics.

I. INTRODUCTION

Reduced motion governs the time evolution of a finite dimensional subspace K. It is represented by $Pe^{-iHt}P$, where P is the projection into the subspace K. This kind of operator is used in the description of decaying systems.

Let H be the self-adjoint Hamiltonian of the physical system defined on the Hilbert space $\mathscr{H}$. With essential spectrum $\sigma_l(H)$ and point spectrum $\sigma_P(H)$ and let P be an orthogonal projection on a finite dimensional subspace K of $\mathscr{H}$. The probability for a state $\Psi \in K$ created at $t = 0$ to decay is (the φ_i from an orthonormal basis in K)

$$\begin{aligned} Q(t) &= \int |\langle E | U(t) | \psi \rangle |^2 \, dE \\ &= 1 - \sum_m |\langle \phi_m | U(t) | \psi \rangle |^2 \\ &= 1 - \sum_m |\langle \phi_m | PU(t)P | \psi \rangle |^2, \end{aligned} \tag{1.1}$$

where $E \in \sigma_l(H)$ and $U(t) = e^{-iHt}$.

The total time- evolution $U(t) = e^{-iHt}$ and the resolvent $R(z) = (z - H)^{-1}$ are related to each other by the inverse Laplace transform

$$U(t) = \frac{1}{2\pi i} \oint R(z) e^{-izt} \, dz , \tag{1.2}$$

where the integration path is around the spectrum of H. If we project this into K, we can express the reduced motion $U'(t) = PU(t)P$ by the reduced resolvent $R'(z) = PR(z)P$ as

$$U'(t) = \frac{1}{2\pi i} \oint R'(z) e^{-izt} \, dz . \tag{1.3}$$

The operator $R'(z)$ on K is meromorphic on the complement of $\sigma_l(H)$ and invertible when $\mathrm{Im} z \neq 0$ (Howland[1]). The formula

$$[z - h(z)]^{-1} \equiv R'(z) \tag{1.4}$$

defines an operator $h(z)$ on K which is meromorphic in z on the complement of $\sigma_l(H)$ and has only real singularities. Following Howland,[1] $h(z)$ will be called the Livsic matrix of H and K. Horwitz and Marchand[2] made the following assumptions: (a) When $\dim P = n$ is finite, the reduced resolvent $R'(z)$ can be continued from above through the spectrum $\sigma_l(H)$, and is regular analytic in the second sheet except for n distinct simple poles situated in the lower half plane near $\sigma_l(H)$. (b) The rank of analytic continuation of $R(z)$ is n in the regularity domain.

We shall relax the first assumption, and assume that in the second sheet $R'(z)$ is analytic except for l distinct poles, where the sum of the multiplicities is less than or equal to n [bifurcation is not considered because it can cause singularities in $h(z)$].

Stodolsky[3] discussed degeneracy of this type in the two dimensional case for application to molecular spectroscopy. We wish to treat the general case here. Wong[4] has given a complete mathematical discussion for the problem where the non-Hermitian matrix $h(z)$ does *not* depend on z. We shall, however, be able to adapt his approach to a study of the degenerate decay problem for which there is a nontrivial "z" dependence.

In accordance with the assumption made by Wong,[4] we shall assume that (c) $R'(z)$ has a Laurent expansion about any of its isolated poles:

$$R'(z) = \sum_{n=-r(b)}^{\infty} (z - E_b)^n B_n , \tag{1.5}$$

where $r(b)$ (finite) is the order of the pole at E_b, and

$$B_n = \frac{1}{2\pi i} \oint_{\Gamma_b} (z - E_B)^{-n-1} R'(z) \, dz , \tag{1.6}$$

where the contour Γ_b encloses only the singularity E_b. When the integration path in Eq. (1.3) is deformed into the second sheet from above through $\sigma_l(H)$, we pick up the contribution from the poles in the second sheet plus some contribution from the branch points. For intermediate times, the contributions of the poles dominate the integral.

In Sec. II, as a review of the technique of Wong and its application to the decay problem, we investigate the case where $h(z) = h$ (constant), its spectral representation, and its contribution to the "time evolution". To understand the mechanism of Wong's method for this case, we discuss and example of a matrix that can be diagonalized and study its limit to a nondiagonalizable form. In addition to analytic continuation of the resolvent through the fixed cut on the positive axis, we introduce the technique of Balslev and Combes[5] which effectively rotates the continuous spectrum.

[a)]Work supported in part by the U.S.-Israel Binational Science Foundation.

This method has many calculational and conceptual advantages in applications to the problem of unstable systems (Horwitz and Sigal[6]).

In the next section an approximate form of $h(z)$ (z dependent) is introduced. We show that locally $h(z)$ has the Wong form and therefore his results can be used to study the general case.

In the last section we investigate the limit of infinite degeneracy, and show that where can be a fixed point in K to which the motion may evolve in a way that appears *analogous* to an approach to equilibrium in statistical mechanics.

II. LIVSIC MATRIX INDEPENDENT OF z

Consider the simple case when $h(z) = h$ (constant) in the second sheet: It has the following form:

$$h(z) = \sum_i (E_i P_i + N_i), \tag{2.1}$$

where P_i are projections satisfying the conditions

$$P_i P_j = \delta_{ij} P_i \quad (E_i \neq E_j \text{ for } i \neq j) \tag{2.2}$$

and N_i are nilpotents satisfying the conditions

$$N_i P_j = \delta_{ij} N_i, \quad (N_i)^{r(i)} = 0 \quad [r(i) = \dim P_i]. \tag{2.2a}$$

$R'(z)$ has in this case a Laurent expansion about any of its poles E_b

$$R'(z) = \sum_{n=-r(b)}^{\infty} (z - E_b)^n B_n, \tag{2.3}$$

where

$$B_{-1} = B^2_{-1} = P_b, \quad B_{-2} = N_b,$$
$$B_{-n} = N_b^{n-1} \quad (n \geqslant 2),$$
$$B_{-m} B_{-n} = B_{-n-m-1}, \quad P_b B_0 = B_0 P_b = 0,$$
$$B_n = (-1)^n B_0^{n+1} \tag{2.4}$$

and for two distinct isolated poles E_a and E_b the product of two operator coefficients A_m and B_n vanishes for m and n less than zero. The proof of this form is due to Wong.[4] In K, h is a matrix and we shall use its Jordan canonical form (Kato[7]).

For the description of the degenerate Livsic matrix we shall need the following definition.

Definition (2.1): λ_1 is an eigenvalue of h associated with the eigenvector ϕ_0^R if there is a vector ϕ_1^R such that

$$(h - \lambda_1 I)\phi_1^R = \phi_0^R.$$

ϕ_1^R is called a generalized eigenvector of the first type. If there is a vector ϕ_2^R such that

$$(h - \lambda_1 I)\phi_2^R = \phi_1^R,$$

ϕ_2^R is called generalized eigenvector of the second type. ϕ_{i+1}^R is called a generalized eigenvector of $(i+1)$ type if

$$(h - \lambda_1 I)\phi_{i+1}^R = \phi_i^R,$$

where ϕ_i^R is a generalized eigenvector of the (i) type.

As evident from the definitions, $(h - \lambda I)$ is a stepping operator on the set of generalized eigenvectors. We shall prove at the end of this section that the set of eigenvectors and generalized eigenvectors are a complete set. It will be shown that the time evolution associated with an nth order pole is as follows:

$$V(t) = \frac{1}{2\pi i}\oint (\lambda_i I - h)^{-1} e^{-i\lambda t}\, d\lambda$$
$$= e^{-i\lambda_i t}\left\{P_i + (-it)N_i + \frac{(-it)^2}{2!}N_i^2 + \cdots + \frac{N_i^{(r(i)-1)}(-it)^{V(i)-1}}{[r(i)-1]!}\right\}$$

and the time evolution of ϕ_j^R is

$$V(t)\phi_j = e^{-i\lambda_i t}\left\{\frac{(-it)^{j-1}}{(j-1)!}\phi_1^R + \frac{(-it)^{j2}}{(j-2)!}\phi_2^R + \cdots + \phi_j^R\right\}.$$

Russkanen[8] has discussed the occurrence of nonexponential decays due to the coalescence of poles in a Lee model using generalized eigenvectors, but only considered the case of $h(z) = \text{const}$. Before giving the general proof of the statements, it will be useful to consider a three dimensional example in some detail.

A. Three dimensional example

A simple example of a matrix h that illustrates the main ideas is the three dimensional case

$$h = \begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{pmatrix}. \tag{2.5}$$

We shall use Schmidt's form of h which is given by

$$(h - I\lambda) = E(\lambda)D(\lambda)F(\lambda), \tag{2.6}$$

where $E(\lambda)$ and $F(\lambda)$ are matrix valued functions (which are products of elementary transformations) whose determinants are constant (do not depend on λ) and do not vanish. $D(\lambda)$ is a diagonal matrix.

In this example, it is given by

$$D(\lambda) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & (\lambda - \lambda_1)^3 \end{pmatrix}, \tag{2.7}$$

$$E(\lambda) = \begin{pmatrix} 1 & 0 & 0 \\ (\lambda - \lambda_1) & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda - \lambda_1 & 1 \end{pmatrix}, \tag{2.8a}$$

$$F(\lambda) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -(\lambda - \lambda_1)^2 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & (\lambda_1 - \lambda) & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.8b}$$

Matrices $E(\lambda)$ and $F(\lambda)$ are written in factored form for the purpose of calculating inverses. If λ is the eigenvalue of h with eigenvector

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

then

$$D(\lambda)F(\lambda)\begin{pmatrix}x\\y\\z\end{pmatrix}=0 \qquad (2.9)$$

because $E(\lambda)$ has a constant nonzero determinant.

The subspace in which $D(\lambda)$ can vanish is characterized by the vector

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

The eigenvector corresponding to the subspace for which $D(\lambda)F(\lambda)$ vanishes is therefore given by

$$[F(\lambda)]^{-1}\begin{pmatrix}0\\0\\1\end{pmatrix} \qquad (2.10)$$

In terms of the definition

$$\begin{pmatrix}1\\0\\0\end{pmatrix}=\phi_0^R,\quad \begin{pmatrix}0\\1\\0\end{pmatrix}=\phi_0^R,\quad\text{and}\quad \begin{pmatrix}0\\0\\1\end{pmatrix}=\phi_2^R, \qquad (2.11)$$

one obtains

$$\begin{pmatrix}x\\y\\z\end{pmatrix}=\phi_0^R-(\lambda_1-\lambda)\phi_1^R+(\lambda_1-\lambda)^2\phi_2^R, \qquad (2,12)$$

where ϕ_0^R is an eigenvector, and ϕ_1^R and ϕ_2^R are the generalized eigenvectors of the first and second type, respectively.

In the same way [using Eq. (2.8b)] the left "eigenvector" is

$$\begin{aligned}(x,y,z)&=(0,0,1)-(\lambda_1-\lambda)(010)+(\lambda_1-\lambda)^2(100)\\&=\phi_0^L-(\lambda_1-\lambda)\phi_1^L-(\lambda_1-\lambda)^2\phi_2^L. \qquad (2.13)\end{aligned}$$

It is easy to check that

$$(h-\lambda_1 I)\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}1\\0\\0\end{pmatrix},$$

$$(h-\lambda_1 I)\begin{pmatrix}0\\0\\1\end{pmatrix}=\begin{pmatrix}0\\1\\0\end{pmatrix}.$$

As indicated in the introduction, the inverse Laplace transform of the resolvent will be interpreted as the time evolution operator for the reduced motion. We shall consider in what follows only the contribution of the poles to the time evolution, and redefine

$$U(t)=\frac{1}{2\pi i}\oint_{\text{poles}}(\lambda I-h)^{-1}e^{-i\lambda t}\,d\lambda, \qquad (2.14)$$

$$(\lambda I-h)^{-1}=\begin{bmatrix}\dfrac{1}{\lambda_1-\lambda} & \dfrac{1}{(\lambda_1-\lambda)^2} & \dfrac{1}{(\lambda-\lambda_1)^3}\\ 0 & \dfrac{1}{\lambda_1-\lambda} & \dfrac{1}{(\lambda_1-\lambda)^2}\\ 0 & 0 & \dfrac{1}{\lambda_1-\lambda}\end{bmatrix}, \qquad (2.15)$$

where h is the matrix given in Eq. (2.5). The inverse operator (2.15) can be written in the form of Eq. (2.3), where $P=I$, and

$$N=\begin{pmatrix}0&1&0\\0&0&1\\0&0&0\end{pmatrix}. \qquad (2.16)$$

Then $U(t)$ is obtained from Eqs. (2.14) and (2.15) to be

$$U(t)=e^{-i\lambda_1 t}\begin{pmatrix}1 & -it & (-it)^2/2\\ 0 & 1 & -it\\ 0 & 0 & 1\end{pmatrix}. \qquad (2.17)$$

The time evolution of a vector "u" is $U(t)u$. Therefore, the time evolution of the eigenvector and the generalized eigenvectors are

$$\begin{aligned}U(t)\phi_0^R&=e^{-i\lambda_1 t}\phi_0^R,\\ U(t)\phi_1^R&=e^{-i\lambda_1 t}[(-it)\phi_0^R+\phi_1^R],\\ U(t)\phi_2^R&=e^{-i\lambda_1 t}\left[\frac{(-it)^2}{2!}\phi_0^R+(-it)\phi_1^R+\phi_2^R\right].\end{aligned} \qquad (2.18)$$

B. Generalization of the result to arbitrary finite dimensions

We now generalize the result of the three dimensional case to an arbitrary number of dimensions. Using Eqs. (2.1)–(2.2a), we see that it is enough to look at one term of the sum

$$h=\sum_i(\lambda_i P_i+N_i)$$

because the terms act in independent subspaces. Let P_i have r dimensions; then, $\lambda_i P_i+N_i=h_i$, where

$$h_i=\begin{bmatrix}\lambda_i & \eta_1 & 0 & \cdot & \cdot & 0\\ 0 & \lambda_i & \eta_2 & & & \\ \cdot & & \cdot & \cdot & & \\ \cdot & & & \cdot & \cdot & \\ 0 & & & & \lambda_i & \eta_{r-1}\\ 0 & & & & 0 & \lambda_i\end{bmatrix}, \qquad (2.19)$$

Let us assume that none of the n_i are zeros, since if one of them were zero the matrix would be decomposed into two blocks.

We consider first the case $n_1=n_2=n_{r-1}=1$. The only eigenvector is ψ_1, a column vector with a one in the first entry, and zeros elsewhere; the eigenvalue is λ_i. The generalized eigenvectors are $(\psi_1,\psi_2,\psi_3,\ldots,\psi_{r-1})$, where ψ_j has one in the jth entry and zeros elsewhere. These vectors fulfill the definition (2.1):

$$(h_i-\lambda_i)\psi_{j-1}=\psi_j\,. \qquad (2.20)$$

There are only $(r-1)$ generalized eigenvectors satisfying Eq. (2.20) because

$$(h_i-\lambda_1)\psi_r=(h_i-\lambda_i)^r\psi_1$$

and

$$(h_1-\lambda_1)^r=N_i^r=0.$$

Theorem (2.1) (*Ref.* 9): The generalized eigenvectors are a complete set. (We call a generalized eigenvector of type zero an eigenvector.)

Proof: The eigenvector is determined by $(A-\lambda_0 I)g_1 = 0$. If A has only one dimension, then the proof is obvious. If not, we define g_2 by the following:

$$g_i = (A-\lambda_0 I)g_2$$

and so on until we cannot find a solution to

$$g_1 = (A-\lambda_0 I)g_{i+1}\,.$$

$(g_1, g_2, \ldots, g_r)$ are independent because if not we can find a minimal set $(g_1, \ldots, g_{r_0})$ such that

$$\sum_{i=1}^{r_0} \alpha_i g_i = 0, \quad \alpha_{r_0} \neq 0\,,$$

then

$$0 = (A-\lambda_0 I)\sum_{i=1}^{r_0}\alpha_i g_i = \sum_{i=2}^{r_0}\alpha_i g_{i-1} = 0,$$

in contradiction to the minimal assumption on r_0.

The computation of the time evolution (see the introduction) involves the computation of $(\lambda - h_i)^{-1}$. Let $(\lambda-\lambda_i) = a$; then $(\lambda I - h_i) = aI(I-B)$, where

$$B = \frac{1}{a}N \tag{2.21}$$

and using the equality for nilpotent B $(I = P_h)$, i.e.,

$$I = (I-B)(I+B+B^2+B^3+\cdots+B^n\cdots), \tag{2.22}$$

we find that

$$(I-B)^{-1} = \frac{1}{a}(I+B+B^2+B^3+\cdots) \tag{2.23}$$

and

$$(\lambda I - h_i)^{-1} = \left(I\frac{1}{\lambda-\lambda_i}\right) + N_i\frac{1}{(\lambda-\lambda_i)^2} + \cdots + (N_i)^n\frac{1}{(\lambda-\lambda_i)^n} + \cdots\,. \tag{2.24}$$

The sum (2.24) has a finite number of terms because $(N)^r = 0$, where r is the dimension of the matrix.

Then, $V(t)$ is

$$V(t) = \frac{1}{2\pi i}\oint(\lambda I - h_i)^{-1}e^{-i\lambda t}\,d\lambda = e^{-it\lambda_i}\left\{P_i + (-it)N_i + \frac{(-it)^2}{2!}N_i^2 + \cdots + \frac{(N_i)^{(r-1)}(-it)^{(r-1)}}{(r-1)!}\right\} \tag{2.25}$$

and the time evolution of ψ_j is

$$V(t)\psi_j = e^{-i\lambda_i t}\left\{\frac{(-it)^{j-1}}{(j-1)!}\psi_1 + \frac{(-it)^{j-2}}{(j-2)!}\psi_2 + \cdots + \psi_j\right\}. \tag{2.26}$$

Remarks (2.1): (1) For the case where the equality $\eta_1 = \eta_2 = \cdots = \eta_{r-1} = 1$ does not hold but $\eta_i \neq 0$, $\forall\, i, \phi_i$ the $(i-1)$th generalized eigenvector is equal to $(1/C)\psi_i$ (where ψ_i is as defined above), and

$$C = \prod_{j=1}^{i}\eta_i\,.$$

The equations (2.21) and (2.22) hold in this case and Eq. (2.25) becomes [with the new $V(t)$]

$$V(t)\psi_j = e^{-i\lambda_i t}\left\{C_i^j\frac{(-it)^{j-1}}{(j-1)!}\psi_1 + C_i^j\frac{(-it)^{j-2}}{(j-2)!}\psi_2 + \cdots + C_i^j\psi_j\right\}, \tag{2.27}$$

where (with $\eta_0 = 1$)

$$C_i^j = \eta_j\,\eta_{j-1}\cdots\eta_i\,.$$

Hence, we obtain

$$V(t)\phi_j = e^{-i\lambda_i t}\left(d_i^j\frac{(-it)^{j-1}}{(j-1)!}\phi_1 + \cdots + d_i^j\phi_j\right), \tag{2.28}$$

where

$$d_i^j = C_i^j/C\,.$$

(2) Using Eqs. (2.20) and (2.24) we find that

$$\frac{1}{2\pi i}\oint_{\text{Poles}}(\lambda I - h_i)^{-1}\,d\lambda = P_i = \sum|\psi_i\rangle\langle\psi_i|\,.$$

(3) If we use

$$h_i(\eta) = \lambda_i P_i + \eta N_i \tag{2.29}$$

instead of $h_i = h_i(1)$, the rhs of Eq. (2.24) becomes

$$\frac{1}{\lambda-\lambda_i}I + \eta N_i\frac{1}{(\lambda-\lambda_i)^2} + \ldots + (\eta N_i)^{r-1}\frac{1}{(\lambda-\lambda_i)^r}; \tag{2.30}$$

then, $V(t)$ is

$$e^{-i\lambda_i t}\left\{P_i + (-i\eta t)N_i + \frac{(-i\eta t)^2}{2!}N_i^2 + \cdots + \frac{(-i\eta t)^{n!}}{n}(N_i)^n + \cdots + \frac{(-i\eta t)^{r-1}}{(r-1)!}N_i^{r-1}\right\} \tag{2.31}$$

Note that in the limit $\eta \to 0$ Eq. (2.30) becomes $1/(\lambda-\lambda_i)$ and Eq. (2.31) becomes $e^{-i\lambda_i t}P_i$ as in usual Hamiltonian. In this limit $(\eta\to 0)$, $h_i(\eta)$ cannot make transitions between the ψ_i and there is no nonexponential term in $V(t)$.

C. Discussion

The non-Hermitian operator h was different from a diagonalizable Hamiltonian in the following ways:

(1) We have to find the eigenvectors and the generalized eigenvectors to get a full decomposition of h. For a diagonalizable Hamiltonian the eigenvectors alone provide this description, because in this case all the P_i in Eq. (2) are one dimensional.

(2) In the nondiagonalizable degenerate case, in the block associated with λ_i, the eigenvector and the generalized eigenvectors are vectors with "energy" equal to the eigenvalue in the sense of the expectation value (in the nonorthogonal basis), i.e.,

$$(\psi_i^t, h^n\psi_i) = (\lambda_i)^n = (\lambda_i)^n\|\psi_i\|\,,$$

where ψ_i is the $(i-1)$ generalized eigenvector and ψ_i^t is the dual to ψ_i in the nonorthogonal basis, i.e., the corresponding left eigenvector, and generalized left eigenvector, as in Eqs. (2.12) and (2.13). Linear combinations of the ψ_i's are not, however, vectors with energy λ_i. For example,

$$[(\psi_1+\psi_2)^t, h(\psi_1+\psi_2)]$$
$$=[(\psi_1+\psi_2)^t, \lambda_i((\psi_1+\psi_2)+\psi_1)]$$
$$=\|\psi_1+\psi_2\|\lambda_i+1 .$$

(3) The time evolution operator on the eigenvector ψ_1 and on the generalized eigenvector ψ_2 differs in two essential features; on ψ_2 it is not a multiplication operator and it has nonexponential behavior.

To clarify the meaning of these results let us consider a three dimensional example which is diagonalized, and study its limit to nondiagonalizable form. The matrix h has only one eigenvector. If it is changed by adding small quantities to the diagonal terms

$$A=\begin{pmatrix}\lambda_1 & 1 & 0\\ 0 & \lambda_1+\epsilon & 1\\ 0 & 0 & \lambda_1+v\end{pmatrix},$$

then the new matrix A has three eigenvalues and eigenvectors. The eigenvalues are

$$\lambda=\lambda_1,\quad \lambda=\lambda_1+\epsilon,\quad \text{and } \lambda=\lambda_1+v \tag{2.32}$$

and the corresponding eigenvectors are

$$\phi'_{R_0}=\begin{pmatrix}1\\0\\0\end{pmatrix},\quad \phi'_{R_1}=\begin{pmatrix}1\\ \epsilon\\0\end{pmatrix},\quad \text{and}\quad \phi'_{R_2}=\begin{pmatrix}1\\ r\\ v(v-\epsilon)\end{pmatrix}. \tag{2.33}$$

The three left eigenvectors [corresponding to the ordering (2.32)] are

$$\phi'_{L_0}=(v\epsilon,-v,1),\ \phi'_{L_1}=(0,-(v-\epsilon),1),\ \phi'_{L_2}=(1,0,0). \tag{2.34}$$

Note that where $v_1\epsilon\to 0$, then ϕ'_{R_1} and ϕ'_{R_2} are equal to ϕ'_{R_0}, which is equal to the eigenvector of Eq. (2.5), and ϕ'_{L_0} and ϕ'_{L_2} are equal to ϕ'_{L_2}, which is equal to the left eigenvector of Eq. (2.5).

Stodolsky[3] pointed out that when the eigenvectors depend on parameters, it can happen that the number of eigenvectors becomes less than the dimension of the space for certain values of the parameters. He also states that the eigenvectors are "orthogonal" in the following sense: If we define the adjoints by

$$(\phi'_{R_0})^+=\phi'_{L_0},\quad (\phi'_{R_1})^+=\phi'_{L_1},\quad (\phi'_{R_2})^+=\phi'_{L_2},$$

then one obtains an orthogonality of the type

$$(\phi'^{+}_{R_0}\,\phi'_{R_1})=(\phi'^{+}_{R_{01}}\,\phi'_{R_2})=(\phi_1^+\,\phi'_{R_2})=0, \tag{2.35}$$

as can be checked using Eqs. (2.33) and (2.34).

The eigenvectors can be taken as

$$\phi'_{R_0}=\begin{pmatrix}1\\0\\0\end{pmatrix},\quad \phi'^{+}_{R_0}=\left(1,-\frac{1}{\epsilon},\frac{1}{v\epsilon}\right),$$

$$\phi'_{R_1}=\begin{pmatrix}1/\epsilon\\1\\0\end{pmatrix},\quad \phi'^{+}_{R_1}=\left(0,1,\frac{-1}{(v-\epsilon)}\right), \tag{2.36}$$

$$\phi'_{R_2}=\begin{pmatrix}\dfrac{1}{v(v-\epsilon)}\\ \dfrac{1}{v-\epsilon}\\ 1\end{pmatrix},\quad \phi'^{+}_{R_2}=(0,0,1).$$

After normalization, the eigenvectors fulfill a completeness relation. Let

$$Q_1=\phi_{R_0}\phi^+_{R_0}=\begin{pmatrix}1 & -1/\epsilon & 1/v\epsilon\\ 0&0&0\\0&0&0\end{pmatrix}, \tag{2.37}$$

$$Q_2=\phi'_{R_1}\phi^+_{R_1}=\begin{pmatrix}0 & 1/\epsilon & 1/\epsilon(\epsilon-v)\\ 0&1&1/(\epsilon-v)\\0&0&0\end{pmatrix}, \tag{2.38}$$

$$Q_3=\phi'_{R_2}\phi^+_{R_2}=\begin{pmatrix}0 & 0 & 1/v(v-\epsilon)\\ 0&0&1/(v-\epsilon)\\0&0&1\end{pmatrix}, \tag{2.39}$$

then the completeness relation is

$$\sum_{i=1}^{3}Q_i=1 .$$

Furthermore, Q_1, Q_2, and Q_3 are idempotents, i.e.,

$$Q_1^2=Q_1,\quad Q_2^2=Q_2,\quad \text{and } Q_3^2=Q_3,$$

as can be seen from Eq. (2.37). The nondegenerate matrix A can therefore be written in the form of the Q's:

$$A=\sum_{i=1}^{3}C_iQ_1,$$
$$C_1=\lambda_1,\quad C_2=\lambda_1+\epsilon,\quad C_3=\lambda_1+v. \tag{2.40}$$

We are now in a position to diagonalize A. The matrix S which diagonalizes A is

$$S=\begin{pmatrix}1 & 1/\epsilon & -1/(\epsilon-v)v\\ 0&1&1/(v-\epsilon)\\0&0&1\end{pmatrix} \tag{2.41}$$

and S^{-1} is equal to

$$S^{-1}=\begin{pmatrix}1 & -1/\epsilon & 1/v\epsilon\\ 0&1&1/(\epsilon-v)\\0&0&1\end{pmatrix}. \tag{2.42}$$

The diagonal matrix A' can be found using Eqs. (2.38), (2.40) and (2.41):

$$A'=S^{-1}AS=S^{-1}\Big(\sum_i C_i\phi'_{R_{i+1}}\phi'^{+}_{R_{i+1}}\Big)S$$
$$=\sum_i(C_iS^{-1}\phi'_{R_1}\phi^+_{R_1}S)$$
$$=\begin{pmatrix}\lambda_1&0&0\\0&\lambda_1+\epsilon&0\\0&0&\lambda_1+v\end{pmatrix}. \tag{2.43}$$

where $\epsilon_1 v$ goes to zero, then S and S^{-1} are singular matrices, so A cannot be diagonalized in this limit. The time evolution operator associated with A', the nondegenerate case in diagonal form, is

$$V(t)=\frac{1}{2\pi i}\oint(\lambda I-A')^{-1}e^{-i\lambda t}\,d\lambda\,.$$

Using the form (2.43) for A', one obtains

$$V(t)=\begin{pmatrix} e^{-i\lambda_1 t} & 0 & 0\\ 0 & e^{-i(\lambda_1+\epsilon)t} & 0\\ 0 & 0 & e^{-i(\lambda_1+v)t}\end{pmatrix}. \qquad (2.44)$$

To understand the source of the nonexponential term in the time evolution of the generalized eigenvectors in the degenerate case, consider, for example, the generalized eigenvector ϕ_R^1 defined in Eq. (2.11) in the basis appropriate for A'. ϕ_R^1 in this basis is equal to

$$S^{-1}\phi_R^1=S^{-1}\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}0\\1\\0\end{pmatrix}-\frac{1}{\epsilon}\begin{pmatrix}1\\0\\0\end{pmatrix}\equiv\bar{\phi}_R^1 \qquad (2.45)$$

and its time evolution is

$$V(t)\bar{\phi}_R^1=e^{-i(\lambda_1+\epsilon)t}\begin{pmatrix}0\\1\\0\end{pmatrix}-\frac{1}{\epsilon}e^{-i\lambda_1 t}\begin{pmatrix}1\\0\\0\end{pmatrix}. \qquad (2.46)$$

In the original basis the time evolution has the form

$$S\{V(t)\bar{\phi}_R^1\}=e^{-i(\lambda_1+\epsilon)t}\begin{pmatrix}1/\epsilon\\1\\0\end{pmatrix}-\frac{1}{\epsilon}e^{-i(\lambda_1 t)}\begin{pmatrix}1\\0\\0\end{pmatrix} \qquad (2.47)$$

and when $\epsilon\to 0$, Eq. (2.47) becomes

$$\lim_{\epsilon\to 0}S\{V(t)\bar{\phi}_R^1\}$$

$$=\lim_{\epsilon\to 0}e^{-i\lambda_1 t}\left[\begin{pmatrix}1/\epsilon\\1\\0\end{pmatrix}+(-it)\epsilon\begin{pmatrix}1/\epsilon\\1\\0\end{pmatrix}-\begin{pmatrix}1/\epsilon\\0\\0\end{pmatrix}\right]$$

$$=e^{i\lambda_1 t}\begin{pmatrix}0\\1\\0\end{pmatrix}+(-it)e^{-i\lambda_1 t}\begin{pmatrix}1\\0\\0\end{pmatrix}, \qquad (2.48)$$

which is equal to Eq. (2.18).

In the last paragraph we saw that the nonexponential behavior comes from two eigenvalues that become degenrate, but this condition is not sufficient, as we can see from the following example:

$$\bar{h}=\begin{pmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_1 & 1\\ 0 & 0 & \lambda_1\end{pmatrix}. \qquad (2.49)$$

$\bar{h}$ has two eigenvectors with eigenvalue λ_1:

$$\psi_1=\begin{pmatrix}1\\0\\0\end{pmatrix}\quad\text{and}\quad\psi_2=\begin{pmatrix}0\\1\\0\end{pmatrix} \qquad (2.50)$$

and one generalized eigenvector

$$\psi_3=\begin{pmatrix}0\\0\\1\end{pmatrix}, \qquad (2.51)$$

while

$$\bar{A}=\begin{pmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_1+\epsilon & 1\\ 0 & 0 & \lambda_1+\epsilon\end{pmatrix}, \qquad (2.52)$$

corresponding to Eq. (2.49), is diagonalizable and has the eigenvectors ψ_1 with eigenvalue λ_1, and ψ_2, ψ_3 with eigenvalue $\lambda_1+\epsilon$.

Taking $\epsilon\to 0$ does not bring ψ_2 into ψ_1, and ψ_2 is not a generalized eigenvector as in Eq. (2.33). As we shall see, no nonexponential behavior is associated with ψ_2. This result can be verified by looking at the time evolution operator associated with $\bar{h}$:

$$\bar{V}(t)=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & -it\\ 0 & 0 & 1\end{pmatrix} \qquad (2.53)$$

and

$$\bar{V}(t)\psi_2=e^{-i\lambda_1 t}\psi_2\,. \qquad (2.54)$$

The conclusion is that two conditions must be satisfied for nonexponential behavior, namely that two eigenvalues become degenerate *and* the associated eigenvectors become equal.

The second difference in time evolution of ϕ_R^1, namely the term involving ϕ_R^0, is the result of the Hamiltonian's ability to connect generalized eigenvectors, as can be seen by direct calculation from Eq. (2.14). In fact,

$$h\phi_R^1=\phi_R^1+\lambda_1\phi_R^0$$

implies

$$U(t)\phi_R^1=e^{iht}\phi_R^1$$

$$=\left[1-iht+\frac{(-iht)^2}{2!}+\cdots+\frac{(-iht)^n}{n!}+\cdots\right]\phi_R^1$$

$$=\phi_R^1-it(\phi_R^1+\lambda_1\phi_R^0)+\frac{(-it)^2}{2!}(2\lambda_1\phi_R^0+\lambda_1^2\phi_R^1)+\cdots$$

$$=\phi_R^1\left[1-it\lambda_1+\frac{(-it)^2}{2!}\lambda_1^2+\cdots+\frac{(-it)^n}{n!}\lambda_1^n+\cdots\right]$$

$$+(-it)\phi_R^0\left[1+(-it)\lambda_1+\frac{(-it)^2}{2!}\lambda_1^2+\cdots+\frac{(-it)^n}{n!}\lambda_1^n+\cdots\right]$$

$$=e^{-i\lambda_1 t}\phi_R^1+(-it)e^{-i\lambda_1 t}\phi_R^0\,,$$

which is, of course, equal to Eqs. (2.17) and (2.48). This procedure, however, illustrates the transfer properties of the Hamiltonian.

Intuitively, one has a picture of the time evolution of the generalized eigenvector as the generalized eigenvector partly "decaying" the term $e^{-i\lambda_1 t}\phi_R^1$ and partly making a transition to ϕ_R^0 [the term $(-it)\phi_R^0$], at a rate proportional to t. ϕ_R^0 itself decays at a rate $e^{-i\lambda_1 t}$ so the time evolution of the part in this subspace is $ite^{-i\lambda_1 t}$.

The last example, Eq. (2.49), where $\bar{h}$ cannot transfer ψ_1 to ψ_2 and consequently there is no term like $(it)e^{-i\lambda_1 t\psi_1}$in the time evolution of ψ_1 [Eq. (2.54)], is consistent with this picture. To understand this effect, we shall consider another limiting case.

In the general three dimensional case, suppose

$$h=\begin{pmatrix}\lambda_1 & \eta_1 & 0\\ 0 & \lambda_1 & \eta_2\\ 0 & 0 & \lambda_1\end{pmatrix}. \tag{2.56}$$

Then, we obtain for the resolvent

$$(\lambda-h)^{-1} = \begin{bmatrix}\dfrac{1}{\lambda-\lambda_1} & \dfrac{\eta_1}{(\lambda-\lambda_1)^2} & \dfrac{\eta_1\eta_2}{(\lambda-\lambda_1)^3}\\ 0 & \dfrac{1}{\lambda-\lambda_1} & \dfrac{\eta_2}{(\lambda-\lambda_1)^2}\\ 0 & 0 & \dfrac{1}{\lambda-\lambda_1}\end{bmatrix} \tag{2.57}$$

and for the time evolution

$$V(t)=e^{-i\lambda_1 t}\times\begin{bmatrix}1 & -i\eta_1 t & \dfrac{\eta_1\eta_2(-it)^2}{2!}\\ 0 & 1 & -i\eta_2 t\\ 0 & 0 & 1\end{bmatrix}.$$

The time evolution of the eigenvectors is given by

$$V(t)\begin{pmatrix}0\\1\\0\end{pmatrix}=e^{-i\lambda_1 t}\left[\begin{pmatrix}0\\1\\0\end{pmatrix}-i\eta t\begin{pmatrix}1\\0\\0\end{pmatrix}\right]$$

and

$$V(t)\begin{pmatrix}0\\0\\1\end{pmatrix}=e^{-i\lambda_1 t}\left[\begin{pmatrix}0\\0\\1\end{pmatrix}+(-i\eta_2 t)\begin{pmatrix}0\\1\\0\end{pmatrix}+\eta_1\eta_2\frac{(-it)^2}{2!}\begin{pmatrix}1\\0\\0\end{pmatrix}\right].$$

Taking $\eta_1\to 0$ and $\eta_2\to 1$, Eq. (2.56) implies Eq. (2.49); then

$$V(t)\begin{pmatrix}0\\1\\0\end{pmatrix}=e^{-i\lambda_1 t}\begin{pmatrix}0\\1\\0\end{pmatrix}$$

and

$$V(t)\begin{pmatrix}0\\0\\1\end{pmatrix}=e^{-i\lambda_1 t}\left[\begin{pmatrix}0\\0\\1\end{pmatrix}+(-it)\begin{pmatrix}0\\1\\0\end{pmatrix}\right],$$

as we have seen before. On the other hand, if $\eta_1\to 1$ and $\eta_2\to 0$, then

$$V(t)\begin{pmatrix}0\\1\\0\end{pmatrix}=e^{-i\lambda_1 t}\left[\begin{pmatrix}0\\1\\0\end{pmatrix}+(-it)\begin{pmatrix}1\\0\\0\end{pmatrix}\right]$$

and

$$V(t)\begin{pmatrix}0\\0\\1\end{pmatrix}=e^{-i\lambda_1 t}\begin{pmatrix}0\\1\\0\end{pmatrix}.$$

In the first case ($\eta_1\to 0$, $\eta_2\to 1$), h cannot make a transition between

$$\begin{pmatrix}0\\1\\0\end{pmatrix} \text{ and } \begin{pmatrix}1\\0\\0\end{pmatrix}$$

and there is no nonexponential term in

$$V(t)\begin{pmatrix}0\\1\\0\end{pmatrix}.$$

In the second case ($\eta_1\to 1$, $\eta_2\to 0$), h cannot make a transition between

$$\begin{pmatrix}0\\0\\1\end{pmatrix} \text{ and } \begin{pmatrix}0\\1\\0\end{pmatrix}$$

and

$$V(t)\begin{pmatrix}0\\0\\1\end{pmatrix}$$

has no nonexponential term. In both cases, h cannot make transitions between

$$\begin{pmatrix}0\\0\\1\end{pmatrix} \text{ and } \begin{pmatrix}1\\0\\0\end{pmatrix}$$

and there is no term which includes t^2.

III. APPLICATION OF THE WONG DECOMPOSITION IN THE METHOD OF BALSLEV AND COMBES

In this section we will review the Aguilar and Combes definition of dilation analytic potentials and the first and second Balslev–Combes theorems. The analysis of the decay using this technique does not require that the Livsic matrix be independent of z. Aguilar and Combes pointed out[8] that the Wong decomposition can be applied in the framework of their method. This application is discussed in this section and the decay problem is studied.

We first briefly review the paper of Aguilar and Combes,[10] Balslev and Comes[5] and Simon[11] establish the notation and remind the reader of the principal results.

We will need the following definitions:

(1) Sobolev spaces: Let T_0 be the free Hamiltonian $-\nabla^2$ on $H=L^2(R^3)$; then the Sobolev space H_{+m} is $H_m=\{\psi\in D(T_0^m)\}$ with the norm $\|\psi\|_m=\|(T_0+1)_2^m\psi\|$, e.g., for $m=2$, the scalar product is $(\phi,\psi)=(T_0\phi, T_0\psi)+(\phi_1\psi)+2(T_0\phi_1\psi)$

(2) The dilatation group in $L^2(R^n)$ is defined as

$$[U(\theta)\phi](x)=e^{(n/2)\theta}[\phi(e^\theta x)],\quad \theta\in R,\quad \phi\in L^2(R^n) \tag{3.1}$$

and in momentum space

$$[U(\theta)\phi](P)=e^{(n/2)\theta}[\phi(e^{-\theta}x)], \tag{3.2}$$

and its generator is $A=\frac{1}{2}(\mathbf{r}\cdot\nabla+\nabla\cdot\mathbf{r})$ or $\frac{1}{2}(x\cdot p+p\cdot x)$.

(3) An unbounded operator V on $L^2(R^3)$ is said to be a dilatation analytic potential (said to be in C_α) if and only if (a) the domain of V, $D(V)=D(T_0)$ (the domain of T_0), and V are symmetric; (b) the induced operator V: $H_{+2}=DT_0\to H\equiv L^2(R^3)$ is complete; (c) the operator $V(\theta)$: $H_{+2}\to H$

given by

$$V(\theta) = U(\theta)V[U(\theta)]^{-1}, \quad \theta \in R \tag{3.3}$$

has analytic continuation ot the strip $\{\theta \mid |\text{Im}\theta| < \alpha\}$; (d) a dilatation analytic vector is a vector g for which $U(\theta)g$ has analytic continuation to a strip $|\text{Im}\theta| < \alpha$.

The vector space N_α of all the dilatation analytic vectors is precisely the vector space of all vectors ϕ for which

$$\sum_{n=0}^{\infty} \frac{\|A\phi\|^n}{n!}$$

has a radius of convergence α or more.[12]

Balslev and Combes use the operator $T_0 + V$ where T_0 is the free Hamiltonian of the n-body system with center of mass removed

$$T_0 = -\sum_{i=1}^{n} \nabla^2, \quad \text{on } L^2(R^{3n-3})$$

and

$$V = \sum_{0<i<j<n-1} V_{ij}(r_{ij}), \quad r_{ij} = \mathbf{r}_i - \mathbf{r}_j, \quad r_0 = 0.$$

We assume that V_{ij} as an operator on $L^2(R^3 d^3 r_{ij})$ is in some fixed C_α for every i and j. Then $H = T_0 + V$ is a self-adjoint operator on $D(T_0)$. Let $V(\theta)$ be in the group of dilatations on R^{3n-3}. Then

$$V(\theta) = U(\theta)V[U(\theta)]^{-1} = \sum_{i<j} V_{ij}(\theta), \quad \theta \in R$$

has a continuation into the strip $\{\theta \mid |\text{Im}\theta| < \alpha\}$ and thus using Eq. (3.2),

$$H(\theta) = U(\theta)H[U(\theta)]^{-1} = e^{-2\theta}T + V(\theta), \quad \theta \in R \tag{3.4}$$

has continuation to the same strip, and $H(\theta)$ is non-Hermitian for θ complex.

In order to study the spectrum of H, we have to study the thresholds of the cuts. Let $D = [D_i \ldots D_k]$ be a decomposition of $(0,1,\ldots,n-1)$ into $k \geqslant 2$ clusters, i.e., $D_i \cap D_j = \phi$ if $i \neq j$, $\cup_{i=1}^{\beta} D = [0,1,\ldots,n-1]$.

Let H_{D_i} be the Hamiltonian for the cluster D_i i.,e., $H_{D_i} = T_{0D_i} + V_{D_i}$, where T_{0D_i} is the kinetic energy of the particles in D_i. A bound state energy of ΣH_{D_i}, i.e., a sum of energies $E_{D_1}, E_{D_2}, \ldots E_{D_n}$ with E_{D_i}, an eigenenergy of H_{D_i} called the k-body threshold. The family of all these thresholds is denoted Σ. Similarly, we define thresholds of $H(\theta)$ and denote them as $\Sigma(\theta)$.

A. The first Balslev–Combes theorem (Ref. 5)

Under the assumption that all the two-body potentials

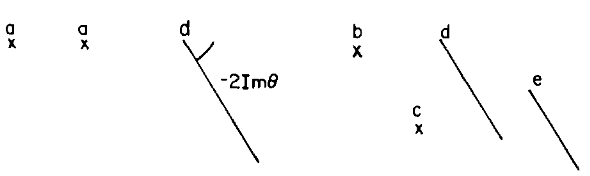

FIG. 1. The Spectrum of $H(\theta)$: (a) discrete eigenvalues; (b) continuous embedded eigenvalues; (c) complex eigenvalues; (d) thresholds of continuous spectrum in H; (e) complex threshold.

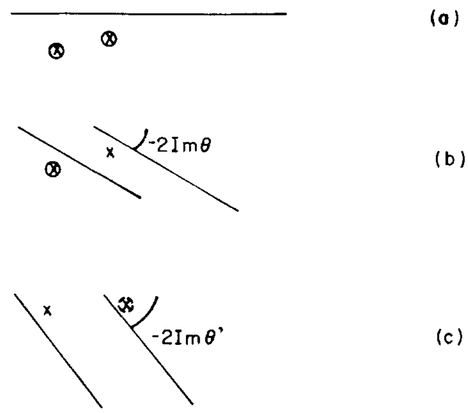

FIG. 2. Complex eigenvalue uncovered and recovered in Imθ varies: ⊗ —covered eigenvalue; × —uncovered eigenvalue; (×)—recovered eigenvalue.

$V_{ij} \in C_\alpha$, the spectrum of $H(\theta)[0 < \text{Im}\theta < \alpha]$ is explicitly given as (see Fig. 1) (a) $[z + e^{-2\theta}r \mid \text{all } z \in \Sigma(\theta), r \in R_\#]$; (b) a set $\sigma_i(\theta)$ of isolated points of the spectrum which are eigenvalues of finite (geometric or algebraic) multiplicity. Moreover,

(1) the real eigenvalues and thresholds of $H(\theta)$ are precisely those of H.

(2) All nonreal eigenvalues and thresholds of $H(\theta)$ lie in the sector

$$[z \mid 0 > \arg(z - \Sigma_{\min}) > 2\,\text{Im}\theta, \quad \Sigma_{\min} = \inf[x \mid x \in \Sigma \cap R].$$

Their presence depends only on Imα. All of the complex poles are in the second sheet of the resolvent (i.e., they are independent of Reθ).

(3) Complex thresholds and eigenvalues of $H(\theta)$ which are isolated from other parts of the essential spectrum of $H(\theta)$ are in $\sigma[H(\theta)]$ if Imθ' is sufficiently near Imθ.

One has the following picture of what happens to the spectrum of H as Imθ increases. For Im$\theta = 0$ there is an essential spectrum beginning at the lowest threshold of H, which is a union of half lines $[\lambda\ \infty)$ for each $\lambda \in \Sigma$, and a set of bound states, some below the continuum and some that may be embedded in the continuum as Imθ "turned up" from below; the bound states stay fixed, but the continuous spectrum swings out into the lower half plane. As it swings out, it can "uncover" some complex eigenvalues and thresholds which stay fixed, unless they happen to get covered again (see Fig. 2).

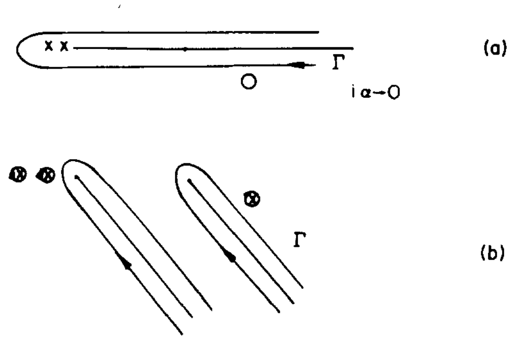

FIG. 3. Behavior of $\langle\psi, e^{-iHt}\psi\rangle$ [in (a), Γ does not enclose the complex pole].

B. The second Balsley–Combes theorem

Let $\psi \in N_\alpha$. Let H be an n-body system with the two-body potential in C_α. Then, $[\psi,(H-z)^{-1}\psi] = f(z)$ originally defined for $z \in C/\text{spec}(H)$ has a (many sheeted) continuation onto the union of the complements of the spectrum of all $H(\theta)$ with $|\text{Im}\theta| < \alpha$ {which we can easily see by inserting $U(\theta)...[U(\theta)]^{-1}$ in the scalar product}.

Augilar and Combes[10] show that in their study of the two-particle case

$$R(\theta,z) = [H(\theta)-z]^{-1} = [e^{-2\theta}T_0 + V(\theta) - z]^{-1}$$
$$= \sum_{n=-\nu(b)} (z-E_b)^n B_n(E,\theta)$$

near the complex pole E_b, as in Wong's Hamiltonian. Van Winter[13] achieved a similar result (in finite dimensional subspaces) by using complex canonical variables.

We shall use $H(\theta)$ (a non-Hermitian Hamiltonian having Wong's form) to calculate the decay of a state ψ, where ψ and $H(\theta)$ have the following properties: (1) $\psi \in N_{\alpha+\epsilon}$; (2) $H(i\alpha)$ has eigenvalue at $E = E_0 - (i\Gamma/2)$ of multiplicity two associated with the eigenvector ϕ_0 and generalized eigenvector ϕ_1; (3) $\psi(i\alpha)$, the analytic continuation of $U(\theta)\psi$ to $\theta = i\alpha$, is close to the subspace of ϕ_0 and ϕ_1 in the L_2 norm.

The probability that ψ (our state at time $t=0$) remains ψ at time t is

$$P(t) = |\langle\psi,e^{-iht}\psi\rangle|^2 = |a(t)|^2.$$

Using

$$\langle\psi,e^{iht}\psi\rangle = \frac{1}{2\pi i}\oint_C \langle\psi|(H-\lambda)^{-1}\psi\rangle\, e^{-i\lambda t}\, d\lambda,$$

where C is a contour going around $\sigma(H)$ (Fig. 3). We can now rotate C and replace $\langle\psi|(H-\lambda)^{-1}\psi\rangle$ with $\langle U(\theta),\psi[H(\theta)-\lambda]^{-1}\,|\,U(\theta)\psi\rangle$ to find (for $\theta = i\alpha$)

$$a(t) = \frac{1}{2\pi i}\oint_{C_1} \langle\psi(i\alpha)\,|[H(i\alpha)-\lambda]^{-1}\psi(i\alpha)\rangle e^{-i\lambda t}\, d\lambda = \frac{1}{2\pi i}\oint_{C_1}\{\langle\psi(i\alpha)|\phi_0\rangle\langle_0\phi\,|[H(i\alpha)-\lambda]^{-1}|\phi_0\rangle\langle_0\phi\,|\psi(i\alpha)\rangle$$
$$+ \langle\psi(i\alpha)|\phi_1\rangle\langle_1\phi\,|[H(i\alpha)-\lambda]^{-1}|\phi_1\rangle\langle_1\phi\,|\psi(i\alpha)\rangle + \langle\psi(i\alpha)|\phi_0\rangle\langle_0\phi\,|[H(i\alpha)-\lambda]^{-1}|\phi_1\rangle$$
$$\times\langle_1\phi\,|\psi(i\alpha)\rangle\}e^{-i\lambda t}\, d\lambda + R(t),$$

where C_1 is around E and $R(t)$ represents contributions from the other poles and cuts. (We use $\eta_1 = 1$ as in Sec. II.C. We use the Laurent expansion about E of $[H(\theta)-\lambda]^{-1}$:

$$[H(\theta)-\lambda]^{-1} = \frac{1}{E-\lambda}(|\phi_0\rangle\langle_0\phi\,| + |\phi_1\rangle\langle_0\phi\,|) + \frac{1}{(E-\lambda)^2}|\phi_0\rangle\langle_1\phi\,| + \sum_{n=0}^{\infty} B_n(\theta,E)(E-\lambda)^n$$

and the fact that $B_n|\phi_0\rangle = B_n|\phi_1\rangle = 0$; then,

$$a(t) = e^{-iEt}[\langle\psi(i\alpha)|\phi_0\rangle\langle_0\phi\,|\psi(i\alpha)\rangle + \langle\psi(i\alpha)|\phi_1\rangle\langle_1\phi\,|\psi(i\alpha)\rangle + (it)\langle\psi(i\alpha)|\phi_0\rangle\langle_1\phi\,|\psi(i\alpha)\rangle] + R(t). \tag{3.5}$$

If the multiplicity of the pole is r and $\psi(i\alpha)$ has the property (3), where (3) is changed to $\psi(i\alpha)$ is close in norm to the subspace of $\phi_0,\phi_1,...,\phi_r$, then

$$a(t) = e^{-iEt}\Big\{\sum_{i=0}^{r}\langle\psi(i\alpha)|\phi_i\rangle\langle i\phi\,|\psi(i\alpha)\rangle + (-it)\sum_{i=0}^{r-1}\langle\psi(i\alpha)|\phi_i\rangle\langle_{i+1}\phi\,|\psi(i\alpha)\rangle + \cdots$$
$$+ \frac{(-it)^n}{n!}\sum_{i=0}^{r-n}\langle\psi(i\alpha)|\phi_i\rangle\langle_{i+n}\phi\,|\psi(i\alpha)\rangle + \cdots\Big\} + R(t). \tag{3.6}$$

If the multiplicity of the pole is r with eigenvectors with J_l generalized eigenvectors associated to the Jth eigenvector, then

$$a(t) = e^{-iEt}\Big\{\sum_{j=1}^{l}\sum_{i=0}^{j}\sum_{k=0}^{j_l-1}\langle\psi(i\alpha)|\phi^j_k\rangle\langle^j_{k+i}\phi\,|\psi(i\alpha)\frac{(-it)^i}{i!}\Big\}.$$

We can rewrite Eq. (3.6) in the form

$$a(t) = e^{-iEt}\Big\{\sum_{i=0}^{r}\langle\psi(i\alpha)|\phi_0\rangle\langle_i\phi\,|\psi(i\alpha)\rangle\frac{(-it)^i}{i!} + \sum_{i=0}^{r-1}\langle\psi(i\alpha)|\phi_1\rangle\langle_{i+1}\phi\,|\psi(i\alpha)\rangle\frac{(-it)^i}{i!}$$
$$+ \sum_{i=0}^{r-2}\langle\psi(i\alpha)|\phi_2\rangle\langle_{i+2}\phi\,|\psi(i\alpha)\rangle\frac{(-it)^i}{i!} + \cdots$$
$$+ \sum_{i=0}^{1}\langle\psi(i\alpha)|\phi_{r-1}\rangle\langle_{i+r-1}\phi\,|\psi_{i\alpha}\rangle\frac{(-it)^i}{i!} + \langle\psi(i\alpha)|\phi_r\rangle\langle\phi_r\,|\psi(i\alpha)\rangle\Big\} = R(t).$$

If we use the more general form of Eq. (2.30), i.e., with $\eta_1 \neq 1$, we have

$$a(t) = e^{-iEt}\Big\{\sum_{i=0}^{r}\sum_{j=0}^{r-i}\langle\psi(i\alpha)|\phi_i\rangle\langle_{i+j}\phi\,|\psi(i\alpha)\rangle\frac{(-i\eta t)^j}{j!}\Big\} + R(t). \tag{3.7}$$

This result makes explicit the contribution to the time evolution of the reduced motion amplitude from the pole at E. The function $R(t)$ contains contributions from other poles and cuts, and a similar analysis could be applied to each of these other poles.

IV. Z DEPENDENT CASE

Although the Balslev–Combes method permits direct application of a procedure utilizing the Wong decomposition, we wish to show that analytic continuation through the real semiaxis cut of the Livsic matrix can be treated in a similar way. Since the Livsic matrix depends in general on z, this treatment involves an extension of the Wong decomposition to its use as an expansion in the neighborhood of each complex pole. This procedure may be useful in cases when the requirement of the Balslev–Combes theorems are not fulfilled.

Our goal in this section is to find an approximate form of a degenerate Livsic $h(z)$ (z dependent) with only one zero. We shall use the following approximation: the contribution to $U(t)$ near a pole E_b comes primarily from this pole. If we neglect the contribution from the cut and other poles. Then $P_b U(t) P_b$ [where P_b is a projection ot the subspace on which $R'(z)$], the resolvent of the Livsic matrix $h(z)$, is singular when $E \to E_b$ is calculated by

$$P_b U(t) P_b \cong \frac{1}{2\pi i} \oint_{\Gamma_b} R'(z) e^{-izt}\, dz, \tag{4.1}$$

where Γ_b is a contour enclosing only the singularity E_b. Using Eq. (4.1) and the relation

$$\lim_{t \to 0} U(t) = I$$

in our approximation, we have

$$\lim_{t\to 0} P_b U(t) P_b = P_b \cong \lim_{t \to 0} \frac{1}{2\pi i} \oint_{\Gamma_b} R'(z) e^{-izt}\, dz. \tag{4.2}$$

Using Eqs. (2.3) and (4.2)

$$\begin{aligned} P_b \cong \lim_{t\to 0} \{ & e^{-iE_b t} (B_{-(r(b))} (-it)^{r(b)-1} + \cdots \\ & + B_{-2}(-it) + B_{-1})\} \\ & = B_{-1}, \end{aligned} \tag{4.3}$$

we conclude that $B^2_{-1} = B_{-1}$ in the sense that $P_b \simeq B_{-1}$.

Using Eqs. (1.5) and (4.3), we get extra information on the structure of the resolvent, by considering

$$B_{-1} \cong B_{-1} B_{-1} = \frac{1}{2\pi i} \int_{\Gamma'_b} dz' \int_{\Gamma_b} dz R'(z') R'(z), \tag{4.4}$$

where Γ'_b is completely inside Γ_b; to compute Eq. (4.4) we give a generalized resolvent identity. For clarity of notation in the following, let us denote $h(z)$ by $H(z)$. Using Eq. (1.4), we obtain

$$z' - H(z) = [z - H(z)] - (z - z')$$

so that

$$R'(z)[z' - H(z)] = I - R'(z)(z - z'),$$

$$\begin{aligned} & R'(z)[z' - H(z')] + R'(z)[H(z') - H(z)] \\ & \quad = I - R'(z)(z - z'), \end{aligned}$$

or

$$\begin{aligned} & R'(z) + R'(z)[H(z') - H(z)] R'(z') \\ & \quad = R'(z') - R'(z) R'(z')(z - z'). \end{aligned}$$

Then, we have

$$\begin{aligned} & (z - z') R'(z) R'(z') \\ & \quad = R'(z') - R'(z) + R'(z)[H(z') - H(z)] R'(z'). \end{aligned} \tag{4.5}$$

Using Eq. (4.5) in (4.3), one finds

$$\begin{aligned} B_{-1} = & \frac{1}{(2\pi i)^2} \int_{\Gamma_b} dz \int_{\Gamma'_b} dz' \\ & \times [R'(z) - R'(z')](z' - z)^{-1} + \frac{1}{(2\pi i)^2} \\ & \times \int_{\Gamma_b} dz \int_{\Gamma'_b} dz' \frac{R(z)[H(z') - H(z)] R(z')}{(z' - z)}. \end{aligned} \tag{4.6}$$

The first term of Eq. (4.6) is equal to B_{-1} (Wong[4]); therefore, the second term should be approximately zero. Our next step is to suggest a form for $H(z)$, consistent with the validity of the approximations which we wish to use. Let us assume that H has a Taylor expansion about E_b:

$$H(z) = H(E_b) + \sum_{i=1}^{\infty} (z - E_b)^i C_i. \tag{4.7}$$

Then, the second term in Eq. (4.6) (using Sec. II.C) is

$$\begin{aligned} & \frac{1}{(2\pi i)^2} \int_{\Gamma_b} dz \int_{\Gamma'_b} dz' \Big\{ \sum_{m=-r(b)}^{\infty} \sum_{n=1}^{\infty} \sum_{k=-r(b)}^{\infty} [B_m \\ & \times C_n B_R (z - E_b)^m (z - z')^{-1} (z' - E_b)^{k+n} \\ & \times - B_m C_n B_R (z - E_b)^{m+n} (z - z')^{-1} (z' - E_b)^k] \Big\}. \end{aligned} \tag{4.8}$$

Integration on Γ'_b of the most singular term in Eq. (4.8) gives

$$\begin{aligned} & \frac{1}{2\pi i} \int_{\Gamma_b} dz B_{-r(b)} \{ (z - E_b)^{-r(b) + [n - r(b)]} \\ & \times \theta[n - r(b)] - (z' - E_b)^{n - 2r(b)} \} C_n B_{-r(b)}, \end{aligned} \tag{4.9}$$

where

$$\theta(m) = 0, \quad m \geqslant 0, \quad \theta(m) = 1, \quad m < 0.$$

For this contribution to vanish, either $n - 2r(b)$ must be nonnegative or $B_{-r(b)} C_n B_{-r(b)} = 0$.

We shall, in fact, demand that for $2r(b) > n$ (we will need this stronger condition in what follows)

$$B_{-r(b)} C_n = C_n B_{-r(b)} = 0. \tag{4.10}$$

Consider now other less leading terms in Eq. (4.8), obtained by replacing one of the $B_{-r(b)}$ by $B_{-r(b)+l}$; then, if $n < 2r(b)$, this contribution will vanish by Eq. (4.10), and for $n \geqslant 2r(b)$, Eq. (4.9) becomes ($l > 0$)

$$\begin{aligned} & \frac{1}{2\pi i} \int_{\Gamma_b} dz B_{-r(b)} \{ (z - E_b)^{-r(b) + [n + l - r(b)]} \\ & \times \theta[n - r(b) + l] - (z - E_b)^{n - 2r(b) + l} \\ & \times \theta[r(b) + l] \} C_{-n} B_{-r(b)+l}, \end{aligned}$$

which vanishes because there is no pole.

In the same way we see that if we demand the relation

$$B_{-n} C_k = C_k B_{-n} = 0, \quad \text{for } k > 2n > 0, \tag{4.11}$$

all of the terms of (4.8) vanish. Let us consider [to compare with Eq. (2.4)]

$$B_{-n}B_{-m} = \frac{1}{(2\pi i)^2}\int_{\Gamma_b} dz \int_{\Gamma'_b} dz'(z-E_b)^{m-1}(z'-E_b)^{n-1}R'(z)R'(z')$$

$$= \frac{1}{(2\pi i)^2}\int_{\Gamma_b} dz \int_{\Gamma'_b} dz'\{(z-E_b)^{m-1}(z'-E_b)^{n-1}(z'-z)^{-1}$$

$$\times [R'(z) - R'(z')] + R'(z)[H(z') - H(z)]R'(z')\} \tag{4.12}$$

for $0<n, m \leqslant r(b)$.

The second term [similar in structure to Eq. (4.8)] vanishes by Eq. (4.11), and the remaining first term of Eq. (4.12) is equal to B_{-m-n} (by Wong[4]).

Equation (2.4) holds and $R'(z)$ is given by

$$R'(z) = P_b(z-E_b)^{-1} + \sum_{i=1}^{r(b)} N^i(z-E_b)^{-(i+1)} + \sum_{n=0}^{\infty} B_n(z-E_b)^n. \tag{4.13}$$

The difference between $R'(z)$ and Wong's resolvent is that $P_bB_0 = 0$ and $B_n = (-1)^n B_0^n$ cannot be proved in this case (here, these residues are associated with the Laurent expansion around points where then Livsic matrix is essentially different). After the form $R'(z)$ is found, we can calculate $H(z)$ as

$$P_0[I_pE_b - H(E_b)] = \frac{1}{2\pi i}\int_{\Gamma_b} R'(z)[E_bI_p - H(E_b)]\,dz$$

$$= \frac{1}{2\pi i}\int_{\Gamma_b}\{R'(z)[H(z)-z] - R'(z)(z-E_b) + R'(z)[H(z)-H(E_b)]\}$$

$$= 0 - N_b + \frac{1}{2\pi i}\int_{\Gamma_b} R'(z)[H(z)-H(E_b)]\,dz. \tag{4.14}$$

The last term is equal to

$$\frac{1}{2\pi i}\int_{\Gamma_b}\sum_{n=1}^{\infty}\sum_{m=-r(b)}^{\infty} B_mC_n(z-E_b)^m(z-E_b)^n\,dz = \sum_{n=1}^{\infty}\sum_{m=-r(b)}^{\infty} B_mC_n\theta(n+m). \tag{4.15}$$

This term vanishes for $n+m \geqslant 0$ by θ and for $m+n<0$, B_mC_n is zero [by Eq. (4.11)].

Then the form of P_bH_b is

$$P_b[H(E_b)] = P_bE_b + N_b \tag{4.16}$$

and

$$H(E_b)P_b = E_bP_b + N_b = P_bH(E_b). \tag{4.17}$$

Using Eqs. (4.11) and (4.13), $H(E_b)P_b$ can be calculated also, and we obtain the same results.

According to Eq. (4.11), $C_1P_b = 0$, but C_nP_b for $n \geqslant 1$ is so far unrestricted when $N_b = 0$. In this case, the general expansion (4.7) yields

$$H(z)P_b = P_bE_b + \sum_{i=2}^{\infty} C_iP_b(z-E_b)^i. \tag{4.18}$$

If $N_b = \Sigma_{i=1}^{r(b)-1} |i\rangle\langle i+1|$, where $|i\rangle$ is the $(i+1)$th generalized eigenvector, then using Eq. (4.11) and $N_b = B_{-2}, B_{-i} = N_b^{i-1}$ for 1,2,3,...,$r(b)$, we find the following relations: for $n<r(b)$,

$$C_{2n}P_b = \sum_{i=0}^{n-1}\sum_{j=1}^{n} f_{ij}\,|j\rangle\langle r(b)-i| \tag{4.19}$$

and for $2n+1<2r(b)$,

$$C_{2n+1}P_b = \sum_{i=0}^{n-1}\sum_{j=0}^{n} g_{ij}\,|j\rangle\langle r(b)-i| \tag{4.20}$$

and for $n \geqslant r(b)$, C_nP_b is so far unrestricted. However, the formal requirements provide restrictions $P_b[z-H(z)]R(z)P_b = P_b = R(z)[z-H(z)]$, on the B_n and C_n operators. Explicitly,

$$P_b[z-H(z)]R(z)P_b$$

$$= [(z-E_b)P_b - N_b]\Big[P_b\frac{1}{z-E_b} + \sum_{i=1}^{r(b)} N^i(z-E_b)^{-(i+1)}\Big] + [(z-E_b)P_b - N_b]\Big[\sum_{n=0}^{T} B_nP_b(z-E_b)^n\Big]$$

$$-\sum_{n=2}^{\infty} P_bC_n(z-E_b)^n\Big[P_b\frac{1}{z-E_b} + \sum_{i=1}^{r(b)-1} N^i(z-E_b)^{-(i+1)}\Big] - \sum_{\substack{n=0\\m=0}}(P_bC_nB_mP_b)(z-E_b)^{m+n} = P_b\,;$$

hence,

$$\sum_{n=0}^{\infty} P_b\,[B_n(z-E_b)^{n+1} - N_b B_n P_b (z-E_b)^n] - \sum_{n=2}^{\infty} P_b\Big[C_n(z-E_b)^{n-1} + \sum_{i=1}^{r(b)-1} C_n N^i (z-E_b)^{-i-1+n}\Big]P_b$$

$$\times \sum_{\substack{n=2\\ M=0}} P_b C_n B_n P_b (z-E_b)^{n+m} = 0. \tag{4.21}$$

$P_b R(z)[z-H(z)]P_b = P_b$ gives the same equation, but the ordering of the B_n's and C_n's is changed. Collecting equal powers of $(z-E)$, and using the conditions written previously, one easily finds solutions.

A. Relation between Laurent coefficients from the expansion around two different poles

In the following we will study two zeros or $H(z) - z$. In this case, we must consider two poles of the resolvent. Our goal is to prove that the residues are approximately orthogonal, assuming that the pole approximation is valid. We assume the form

$$H(z) = \bar{H} + \sum_{n=1}^{\infty} [(z-E_b)^n C_n + (z-E_a)_n F_n] + \sum_{i,j=1}^{\infty} (z-E_a)^i (z-E_b)^j D_{ij}\,, \tag{4.22}$$

where $\bar{H}$ is independent of (z) (it is a constant operator). Then the second term of Eq. (4.6), i.e.,

$$\frac{1}{(2\pi i)^2}\int_{\Gamma_b'} dz' \int_{\Gamma_b} dz (z'-z)^{-1} R'(z)[H(z') - H(z)]R'(z'), \tag{4.23}$$

must be assumed to be approximately zero for the validity of the pole approximation. From this it will be shown that the residues are orthogonal.

Putting $H(z)$ of Eq. (4.22) into the last integral, we find

$$\frac{1}{(2\pi i)^2}\int_{\Gamma_b'} dz' \int_{\Gamma_b} dz \sum_{m=-r(b)}^{\infty}\sum_{n=1}^{\infty}\sum_{k=-r(b)}^{\infty} [B_m C_n B_k (z'-E_b)^{k+n}(z-E_b)^m - B_m C_n B_k (z-E_b)^{m+n}(z'-E_b)^k]$$

$$\times (z'-z)^{-1} + \frac{1}{(2\pi i)^2}\int_{\Gamma_b'} dz' \int_{\Gamma_b} dz \sum_m \sum_n \sum_k B_m F_n B_k (z-E_b)^m (z'-E_a)^k (z-z')[(z-E_a)^n - (z'-E_a)^n]$$

$$+ \frac{1}{(2\pi i)^2}\int_{\Gamma_b} dz \int_{\Gamma_b'} dz' B_m D_{ij} B_k (z-E_b)^m (z-z')^{-1} (z'-E_b)^k [(z'-E_b)^j (z'-E_a)^i - (z-E_b)^j (z-E_a)^i]\,. \tag{4.24}$$

The treatment of the first term is the same as in the last section; then, $B_{-n}C_k = C_k B_{-n} = 0$ for $k < 2n$; in the second term we look at the most singular part, and require that it will be zero. The result of this requirement is

$$F_n B_{-r(b)} = B_{-r(b)} F_n = 0, \quad n > 0. \tag{4.25}$$

The same requirement (that every term will be zero) for any B_{-n} $(m > 0)$ gives $F_n B_{-m} = B_{-m} F_n = 0$, $m,n > 0$. The third term in Eq. (4.25) gives the same kind of term as the first term of Eq. (4.24) multiplied by $(z - E_a)$ in some power. Using the result that we found in the first term, we get

$$D_{ij} B_{-n} = B_{-n} D_{ij} = 0, \quad j < 2n. \tag{4.26}$$

The proof that the above conditions imply

$$B_{-m} B_{-n} = B_{1-m-n}\,,$$

for $0 < n$, $m \leqslant r(b)$, is the same as in Eq. (4.12). Then, Eqs. (4.13) and (4.14) hold near $z = E_b$.

Assuming that $R'(z)$ near $z = E_a$ has the expansion

$$R'(z) = \sum_{n=-r(a)}^{\infty} A_n (z-E_a)^n,$$

then, by the same argument, we demand, for the validity of the pole approximation near $z = E_a$,

$$A_{-n} C_k = C_k A_{-n} = 0 \quad (k < 2n), \quad A_{-n} F_k = F_k A_{-n} = 0 \quad (k < 2n), \quad D_{ij} A_{-n} = A_{-n} D_{ij} = 0 \quad (i < 2n). \tag{4.27}$$

The relations corresponding to Eq. (4.14) are

$$P_b[E_b - H(E_b)] = -N_b = -B_{-2}\,, \qquad P_a[E_a - H(E_a)] = -N_a = -A_{-2}\,. \tag{4.28}$$

The product of B_{-n} and A_{-m} for $n,m \geqslant 0$ is

$$A_{-n} B_{-m} = \frac{1}{(2\pi i)^2}\int_{\Gamma_a} dz \int_{\Gamma_b} dz' (z-E_a)^{n-1} R'(z)[H(z') - H(z)]R'(z')(z-z')^{-1}(z-E_b)^{m-1}, \tag{4.29}$$

where Γ_a encloses E_a, and Γ_b encloses E_b, but they do not cross. By integration of the first term of Eq. (4.29) we get

$$\frac{1}{(2\pi i)}\int_{\Gamma_b} dz' \frac{R(z)}{(z-z')}[H(z)-\bar{H}]R(z')(z-E_a)^{n-1}(z'-E_b)^{n-1}$$

$$= \int_{\Gamma_b} dz' \frac{R(z)}{(z'-z)}\Big[\sum_{k=1}^{\infty}\sum_{l=-r(b)}^{\infty}[C_kB_l(z'-E_b)^{k+l+m-1}(z-E_a)^{n-1}]$$

$$+\sum_{b=1}\sum_{l=-r(b)}F_kB_l(z-E_b)^{l+m-1}(z'-E_a)^{n+k+1}+\sum_{ij}\sum_{l=-r(b)}D_{ij}[B_l(z'-E_b)^{l+m-1+j}(z-E_a)^{i+n-1}].$$

[$R(z)$ is analytic in Γ_b.] The first term vanishes (for $k\geqslant l$ since there is no pole in this case, and for $k<l$ since $C_kB_l=0$, and the second term vanishes since $l<0$ implies $F_nB_l=0$. The third term vanishes since $D_{ij}B_l=0$ for $j<-2l$. By the same method the second term of Eq. (4.29) vanishes in the integration on Γ_b. In particular, the following identities hold:

$$A_{-1}B_{-1}=P_bP_a=P_aP_b=0,$$
$$A_{-1}B_{-2}=P_aN_b=N_bP_a=0, \qquad (4.30)$$
$$A_{-2}B_{-1}=P_bN_a=N_aP_b=0,$$

and we have again a resolvent formally of Wong type. Using Eqs. (4.28) and (4.30), we see that $\bar{H}$ has the Wong form. Using

$$P_aH(E_a)=E_aP_a+N_a,$$
$$P_bH(E_b)=E_bP_b+N_b,$$

we find that

$$\bar{H}=\sum_i(P_iE_i+N_i).$$

It should be emphasized that the Wong type structure follows from assuming the validity of the pole approximation. To the extent that the pole approximation is not exact, the structure of the resolvent can differ from the precise Wong form. However, the Wong decomposition can be used to analyze the leading time dependence of the contribution to the amplitude from the neighborhood of each pole.

To illustrate the effect of degeneracy in the S matrix, we discuss the example of the Lee–Friedrichs model[2], where $H=H_0+V$, and V connects the continuum only to a few discrete states $|k\rangle$. Then the off shell scattering amplitude is

$$\langle E\,|T(z)|E\rangle=\langle E\,|V|K\rangle\langle K\,|\frac{1}{z-H}|K'\rangle\langle k'|V|E\rangle. \qquad (4.31)$$

One may then approximate $\langle K\,|1/(z-H)|K'\rangle$ by a Wong type resolvent in the neighborhood of each pole, which contains the effect of the pole, as an effective propagator. Then, using Eq. (4.13), the contribution from all of the poles to the S matrix in the neighborhood $E_p\approx E_b$ is

$$\langle P\,|S\,|P'\rangle=\delta^3(P-P')-2\pi i\int(E_p-E_{p'})$$
$$\times\Big\{\sum_{l=0}^{r(b)-1}\frac{\langle P'|V|K_b\rangle\langle(K+l)_b\,|V|P'\rangle}{(E_p-E_b)^{l+1}}$$
$$+\sum_{n=0}^{\infty}\langle P\,|V|R_b\rangle\langle R_b\,|B_n\,|R'_b\rangle$$
$$\times\langle R'_b\,|V|P'\rangle(E_p-E_b)^n\Big\}, \qquad (4.32)$$

where E_b is the pole of the resolvent with multiplicity $r(b)$, $|\,k_b\rangle$ is the (K_b-1)th generalized eigenvector of the pole at E_b, i.e., $\phi^R_{R_b}=|\,k_b\rangle$ and $|\,k_b\rangle=\phi^l_{R_b}$ as defined previously.

Here we have used the following representation of $(N^b_n)^n$:

$$N_b=\sum_{i=1}^{r(b)}|\,i\rangle\langle i+1\,|,\quad (N_b)^2=\sum_{i=1}^{r(b)}|\,i\rangle\langle i+2\,|,$$
$$(N_b)^n=\sum_{i=1}^{r(b)-n}|\,i\rangle\langle i+n\,|. \qquad (4.33)$$

The first term in Eq. (4.32) is a generalization of the Breit–Wigner form that would be obtained from a pole approximation to the usual Lee model. The sum $\Sigma_{n=0}^{\infty}\langle P'\,|VB_nV\,|P\rangle(E_p-E_b)^n$ should be considered as backgroud, even though it contains contributions from other poles, because it is analytic in E_p and vanishes at $E_p=E_b$, while the first term becomes large.

V. INFINITE DIMENSIONAL CASE

One difference between eigenvectors and generalized eigenvectors is that only for eigenvectors does $U(t)$ become a multiplication operator. This raises the question of whether it is possible to find a linear combination of generalized eigenvectors that will have the property that $U(t)$ can be a multiplication operator.

Theorem: Let h be in the form of Eq. (2.31). Let $h=\lambda_iP_i+\eta N_i$. Denote by $|\,\psi_i\rangle$ the ith generalized eigenvector $(0\leqslant i\leqslant \dim P_i=n)$; then, for n finite, $U(t)$ is a multiplication operator only on the ray $a\psi_0$ (ψ_0 is an eigenvector), but for n infinite, $U(t)$ is a multiplication operator also on

$$\phi=a\sum_{i=1}^{\infty}\psi_i$$

Proof: $U(t)$ is a multiplication operator on $\phi=\Sigma_{i=0}^{\infty}a_i\psi_i$ (for $i>\dim P_i$; suppose that $a_i=0$) if

$$U(t)\phi=C(t)\phi. \qquad (5.1)$$

In this case,

$$\lim_{t\to 0}\frac{\partial^nU(t)}{\partial t^n}\phi=\lim_{t\to 0}\frac{\partial_tC(t)}{\partial t^n}\phi=b_n\phi,$$

where b_n is a constant.

Using the form of $U(t)$ [Eq. (2.31)], we find

$$U(t)\phi=e^{-i\lambda_it}\Big[\sum_{j=0}^{\infty}\psi_j\sum_{k=j}^{\infty}a_k\frac{(-i\eta t)^{k-j}}{(k-j)!}\Big] \qquad (5.2)$$

and

$$\lim_{t\to 0}\frac{\partial}{\partial t}[U(t)\phi]$$
$$=-\Big(i\lambda_i\sum_{j=0}^{\infty}a_j\psi_j+\sum_{j=0}^{\infty}\psi_ja^{j+1}\Big)$$
$$=-i\lambda\phi+\eta\sum_{j=0}^{\infty}(a_{j+1}\psi_j).$$

In order that condition (5.1) holds, it is required that

$$\sum_{j=0}^{\infty} a_{j+1}\psi_j = \frac{(b_1 + i\lambda)}{i\eta}\phi,$$

which implies that $a_j = a$ for every j or that $a_j = 0$ for $j > 0$. For the finite case, only the second condition holds.

To compute $C(t)$ for the case $\phi = a\Sigma_{i=0}^{\infty}\psi_i$, one uses Eq. (5.2):

$$U(t)\sum_{i=0}^{\infty} a\psi_i = e^{-i\lambda_i t}\left[\sum_{j=0}^{\infty} a\psi_j \sum_{k=j}^{\infty} \frac{(-i\eta t)^{k-j}}{(k-j)!}\right] = e^{-i(\lambda_i+\eta)t} \quad (5.3)$$

If $\eta = -i\,\mathrm{Im}\,\lambda_i$ (see Appendix of Ref. 14), a value where the part of $h(\eta)$ that transfers one subspace to others is equal to the width, then ϕ is a stable state $[\lambda_i = E_0 - (i\Gamma/2)]$, i.e., a fixed point

$$U(t) = e^{-iE_0 t}\phi. \quad (5.4)$$

If we have at the beginning any other state $\phi' = \Sigma_{i=1}^{\infty} a_i\psi_i$, then if a finite number of terms in ϕ are not zero,

$$\lim_{t\to\infty} U(t)\phi' = 0,$$

but if an infinite number of terms in ϕ are not zero, let us define a by setting

$$(\phi',\phi') = a^2(\phi,\phi).$$

Now, it is an identity that

$$\phi' = a\phi + (\phi' - a\phi)$$

and hence

$$U(t)\phi = ae^{-iE_0 t} + U(t)[\phi' - a\phi].$$

if Γ is positive, then the second term decays, and one obtains

$$\lim_{t\to\infty} U(t)\phi' = e^{-iE_0\lambda t}a\phi.$$

For example, let $\phi' = \Sigma_{n=1}^{\infty}\psi_{2n-1}$:

$$(\phi',\phi) = \tfrac{1}{2}(\phi,\phi);$$

then,

$$U(t)\phi' = e^{-i\lambda_i t}\left\{\sum_{n=1}^{\infty}\psi_{2n-1}\left[1 - \sum_j \frac{(-i\eta t)^{2j}}{(2j)!}\right] + \sum_{n=0}\psi_{2n}\frac{(-i\eta t)^{2j-1}}{(2j-1)!}\right\}$$

$$= e^{-i\lambda_i t}\sum_{n=0}^{\infty}[\cosh(-i\eta t)\psi_{2n+1} + \sinh(-i\eta t)\psi_{2n}]$$

$$= \tfrac{1}{2}\sum_{i=1} e^{-i\lambda_i t - i\eta t}\phi_i + \tfrac{1}{2}\sum e^{-i\lambda_i t + i\eta t}(\psi_{2i+1} - \psi_{2i})$$

$$= \tfrac{1}{2}e^{-iE_0 t}\phi + \tfrac{1}{2}e^{-iE_0 t - \Gamma t}\sum(\psi_{2n+1} + \psi_{2n})$$

$$= U(t)\frac{\phi}{2} - U(t)\left(\phi' - \frac{\phi}{2}\right).$$

The second term decays as $e^{-\Gamma t}$. Taking $\phi' = \Sigma_{i=0}^{\infty}\psi_{2i}$ $[(\phi'\phi') = 1]$, then

$$U(t)\phi' = \tfrac{1}{2}e^{-iE_0 t}\phi + \tfrac{1}{2}e^{-iE_0 t}\sum(\psi_{2n} - \psi_{2n-1}).$$

For any $\phi' = \Sigma_{i=1}^{\infty} a_i\psi_i$, let us consider the entropy of the probability distribution given by $|a_i|^2$. This is

$$S = -\sum_{i=1}^{r}|a_i|^2\ln|a_i|^2. \quad (5.5)$$

Note that S is maximum when $a_1 = a_2 = \ldots = a_r = 1/r^{1/2}$ in which case $\phi = \phi'$ and

$$\max S = -\ln\frac{1}{r} = \ln r. \quad (5.6)$$

As for the finite degeneracy case, we shall compare this result to $H(i\alpha)$ of Balslev and Combes. Let $H(i\alpha)$ have only one pole of degeneracy r; then (we denote ϕ_i the ith generalized eigenvector), we have

$$a(t) = e^{-iE_0 t}\left[\sum_{i=0}^{r}\sum_{j=0}^{r-i}\langle\psi(i\alpha)\mid\phi_i\rangle \times\langle|\phi_{i+j}\mid\psi(i\alpha)|\rangle\frac{(-i\eta t)^j}{j!}\right] + R(t). \quad (5.7)$$

If $\langle\psi(i\alpha)|\phi_i\rangle$ is constant for every i (recall the earlier discussion where ϕ was a state with an equal amount and phase of each of the ψ_i $\phi = \Sigma_i\psi_i$; the interpretation here is the same), then

$$a(t) = e^{-iEt}\sum_{i}^{r}\sum_{j=1}^{r-i}\langle\psi(i\alpha)|\phi_i\rangle \times\langle{}_i\phi\,|\psi(i\alpha)\rangle\frac{(-i\eta t)^j}{j!} + R(t), \quad (5.8)$$

and when $r\to\infty$,

$$a(t) = e^{-iEt}\sum_{i=0}^{\infty}|\langle\psi(i\alpha)|\phi_i\rangle|^2 e^{-i\eta t} + R(t). \quad (5.9)$$

Using

$$\sum|\phi_i\rangle\langle\phi_i| = P_b,$$

we obtain

$$a(t) = e^{-i(E+\eta)t}\langle P_b\psi(i\alpha)|P_b\psi(i\alpha)\rangle + R(t). \quad (5.10)$$

if $\eta = i\Gamma/2$, then we may write

$$a(t) = e^{-iE_0 t}\langle P_b\psi(i\alpha)|P_b\psi(i\alpha)\rangle + R(t). \quad (5.11)$$

When E_b is a real number, $\langle P_b\psi(i\alpha)|P_b\psi(i\alpha)\rangle$ in close to the norm of $\psi(i\alpha)$ by property (3) of Sec. III and $R(t)$ is "small" for the same reason (i.e., the scalar product is small). We call $R(t)$ the fluctuation.

The generalization of the result to large but finite multiplicity "r" includes two associated aspects. The definition of the fluctuation (i.e., a small correction or large correction lasting for a small time) and the times "t_0" and "t_r" are so that for every time "t" such that $t_r > t > t_0$, $a(t)$ can be written asymptotically as a term of fixed magnitude (which may be zero) plus a term that can be included in our definition of the fluctuation. Using Eq. (5.8),

$$a(t) = e^{-iEt}\sum_{i=0}^{r_1}\sum_{j=0}^{\infty}\langle\psi(i\alpha)|\phi_i\rangle\langle\phi_i|\psi(i\alpha)\rangle\frac{(-i\eta t)^j}{j!} - \sum_{i=0}^{r_1}\sum_{j=r_1-i}^{\infty}\langle\psi(i\alpha)|\phi_i\rangle\langle\phi_i|\psi(i\alpha)\rangle\frac{(-i\eta t)^j}{j!}$$

$$+\sum_{i=r_1}^{r}\sum_{j=0}^{r-1}\langle\psi(i\alpha)|\phi_i\rangle\langle\phi_i|\psi(i\alpha)\rangle\frac{(-i\eta t)^j}{j!}$$
$$+R(t), \tag{5.12}$$

where a good choice of r_1 is made, so that for every time t such that $t_r > t > t_0$, the last three terms of Eq. (5.12) are included in our definition of the fluctuation and $a(t)$ is well approximated by the first term. Equation (5.12) can be written in the form

$$a(t) = e^{-iE_0t}\langle P_{r_1}\psi(i\alpha)|P_{r_1}\psi(i\alpha)\rangle + R_1(t), \tag{5.13}$$

where

$$P_{r_1} = \sum_{i=1}^{r_1}|\phi_i\rangle\langle\phi_i|$$

and $R_1(t)$ is the sum of the last three terms in Eq. (5.12).

In this case, there is a metastable state, but, as in the statistical mechanics of systems with a finite number of degrees of freedom, no true equilibrium.

ACKNOWLEDGMENT

The author wishes to express his deep gratitude to Professor L.P. Horwitz for helpful and useful discussions and for careful reading of the manuscript.

[1] J.S. Howland, J. Math. Anal. Appl. **50**, 415 (1975).
[2] L.P. Horwitz and J.P. Marchand, Rocky Mountain J. Math. **1**, 225 (1971).
[3] L. Stodolsky, Phys. Rev. D **1**, 2683 (1970).
[4] J. Wong, J. Math. Phys. **10**, 1438 (1969).
[5] E. Balslev and J.M. Combes, Commun. Math. Phys. **22**, 280 (1971).
[6] L.P. Horwitz and I.M. Sigal, Helv. Phys. Acta **51**, 685 (1978).
[7] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1976).
[8] P.V. Ruuskanen, Nucl. Phys. B **22**, 253 (1970).
[9] Introduction to the Theory of Linear Nonselfadjoint Operators by I.C. Gohberg and M.G. Krein-Translation of Mathematical Monographs, Volume 18.
[10] Aguilar and J.M. Combes, Commun. Math. Phys. **22**, 269 (1971).
[11] B. Simon, Ann. Math. **97**, 247 (1973).
[12] E. Nelson, Ann. Math. **70**, 572 (1959).
[13] C. Van Winter, J. Math. Anal. Appl. **47**, 633 (1974).
[14] Y. Dothan and D. Horn, Phys. Rev. D **1**, 916 (1970).

Phase-integral calculation of physically important quantities for nonrelativistic bound *s* states of the linear central potential

Bo Thidé
Institute of Theoretical Physics, University of Uppsala, Thunbergsvägen 3, S-752 38 Uppsala, Sweden

(Received 26 November 1979; accepted for publication 17 January 1980)

Energy levels, normalization factors, quantal expectation values, and probability densities at the origin for nonrelativistic bound *s* states of the linear central potential are calculated by means of phase-integral formulas, given by N. Fröman in two recent papers. The accuracy of the phase-integral method is exhibited by a numerical comparison with exact results. During the last few years the potential in question has been widely used as a model potential describing quark confinement in heavy mesons.

1. INTRODUCTION

N. Fröman[1,2] has recently derived accurate phase-integral formulas for normalization factors, quantal expectation values, and probability densities at the origin, not involving wavefunctions, for the case of a single well potential $V(r)$ being either regular or Coulomb-like (case A), Coulomb-like and attractive (case B), or regular (case C) close to the origin.

The linear central potential $V(r) = ar$, where $a > 0$, has been widely used as quark-antiquark interaction potential in nonrelativistic charmonium spectroscopy (see, e.g., Refs. 3–7). The purpose of the present paper is to demonstrate that for *s* states of this potential, for which exact solutions of the Schrödinger equation are known in terms of the Airy function Ai, various physically important quantities can be accurately calculated by means of the phase-integral formulae given in Refs. 1 and 2. If higher orders of the phase-integral approximations are used, the agreement with exact results is excellent, particularly for excited states.

For $l = 0$, the time-independent, radial Schrödinger equation is, with obvious notations,

$$\frac{d^2u}{dr^2} + \frac{2\mu}{\hbar^2}(\mathscr{E} - ar)u = 0. \tag{1}$$

Introducing the dimensionless quantities (cf. Eqs. (2) and (3) in Ref. 8)

$$z = \left(\frac{2\mu a}{\hbar^2}\right)^{1/3} r, \tag{2}$$

$$A = \left(\frac{2\mu}{\hbar^2 a^2}\right)^{1/3} \mathscr{E}, \tag{3}$$

we transform (1) into

$$d^2u/dz^2 + Q^2(z)u = 0, \tag{4}$$

where

$$Q^2(z) = A - z. \tag{5}$$

Bound-state solutions are obtained when A assumes certain discrete values A_n, $n = 0,1,2,\cdots$, where n is the number of radial nodes of the corresponding eigenfunction, which we denote by $u_n(z)$ and assume to be normalized in such a way that

$$\int_0^\infty u_n^2(z)\,dz = 1, \quad n = 0,1,2,\cdots. \tag{6}$$

Since the particular physical potential under consideration is regular at the origin, and since the orbital angular momentum is equal to zero, one can use the phase-integral formulas pertaining to either case A or case C in Refs. 1 and 2, depending on whether one wants to work with modified or unmodified phase-integral approximations. In the present paper we shall use both case A and case C formulas to calculate energy levels, normalization factors, expectation values of positive powers of the radial coordinate, and probability densities at the origin.

2. ENERGY LEVELS

The energy spectrum can be obtained from the $(2N + 1)$ th-order phase-integral quantization condition, which, according to (19), (20a), and (20c) in Ref. 1, can be written

$$\tfrac{1}{2}\int_\Gamma q(z)\,dz = (n + \alpha)\pi, \quad n = 0,1,2,\cdots, \begin{cases}\alpha = \tfrac{1}{2} \text{ (case A)},\\ \alpha = \tfrac{3}{4} \text{ (case C)},\end{cases} \tag{7}$$

where, according to (9) in Ref. 1,

$$q(z) = Q_{\text{mod}}(z) \sum_{m=0}^{N} Y_{2m}, \tag{8}$$

with the first few functions Y_{2m} given by (11a–c) in Ref. 1. In (7) Γ stands for either of the contours Γ_A (case A) or Γ_C (case C) shown in Figs. 1 and 2, respectively. In these figures the heavy lines indicate how the complex z plane is cut in order to make $q(z)$ single-valued; the phase of $Q_{\text{mod}}(z)$ is chosen so as to make $Q_{\text{mod}}(z)$ real and positive on the upper lips of the cuts. Defining

$$L^{(2m+1)} = \tfrac{1}{2}\int_\Gamma Y_{2m} Q_{\text{mod}}\,dz, \tag{9}$$

we can also write the quantization condition (7) in the form

$$\sum_{m=0}^{N} L^{(2m+1)} = (n + \alpha)\pi, \quad n = 0,1,2,\cdots \begin{cases}\alpha = \tfrac{1}{2} \text{ (case A)},\\ \alpha = \tfrac{3}{4} \text{ (case C)}.\end{cases} \tag{10}$$

Writing

$$Q^2_{\text{mod}}(z) = Q^2(z) - \beta/4z^2 = A - z - \beta/4z^2, \tag{11}$$

where the constant β is chosen in case A equal to 1 (which for the first-order phase-integral approximation gives the Kramers–Kemble–Langer modification) and in case C

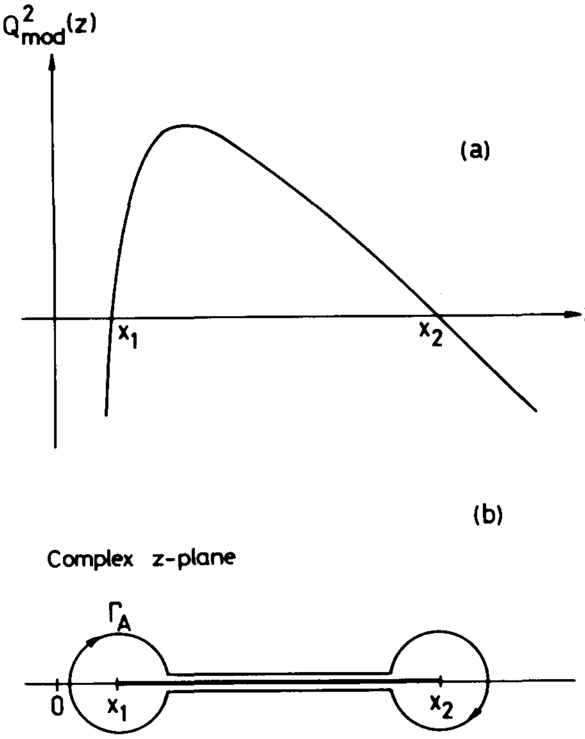

FIG. 1. This figure refers to case A. (a) shows $Q^2_{\rm mod}$ given by (11) with $\beta = 1$ as a function of z when z is real. (b) shows the branch cut (heavy line), and the closed contour of integration Γ_A encircling, in the negative sense, the generalized classical turning points, i.e., the two zeros x_1 and x_2 of $Q^2_{\rm mod}$ lying on the positive real axis; the third zero x_0 is located on the negative real axis.

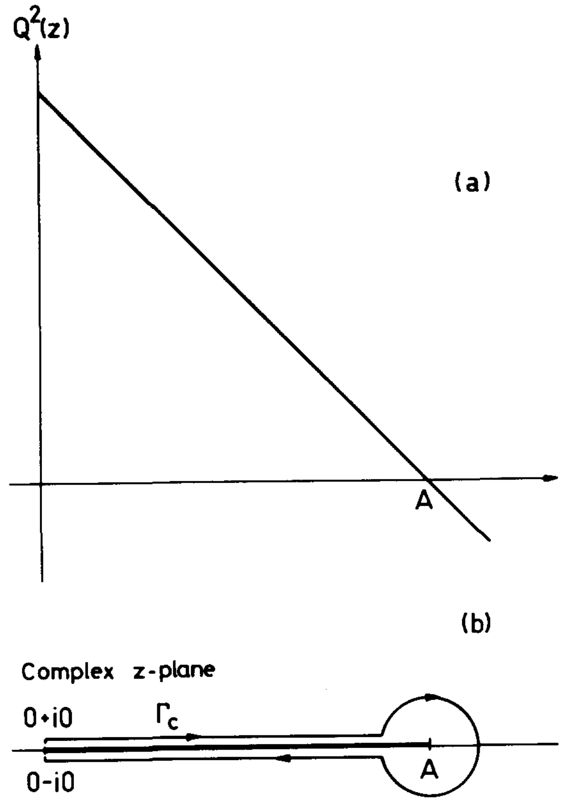

FIG. 2. This figure refers to case C. (a) shows Q^2 given by (5) as a function of z when z is real. (b) shows the branch cut (heavy line), and the nonclosed contour of integration Γ_C running from the point $z = 0 + i0$ on the upper lip of the cut around the classical turning point $x = A$ to the point $z = 0 - i0$ on the lower lip of the cut.

equal to 0 (which gives unmodified phase-integral approximations), and using Eqs. (12)–(14) in Ref. 1, we obtain for the functions $\epsilon_0(z)$ and $\epsilon_2(z)$, which appear in the definitions of Y_2 and Y_4, the following expressions:

$$\epsilon_0 = \frac{1}{[4(A-z)z^2-\beta]^3}[4(4\beta+5)z^6 - 32\beta Az^5 + 16\beta A^2z^4 + 4\beta(2\beta-11)z^3 - 8\beta(\beta-3)Az^2 + \beta^2(\beta-1)], \tag{12a}$$

$$\epsilon_2 = \frac{z^2}{[4(A-z)z^2-\beta]^6}[3456(4\beta+5)z^{10} - 45056\beta Az^9 + 54784\beta A^2z^8 + 256\beta(18\beta - 116A^3 - 585)z^7 - 768\beta(13\beta - 8A^3 - 286)Az^6 + 256\beta(29\beta - 528)\times A^2z^5 - 32\beta[9\beta(\beta-154) + 64(\beta-15)A^3]z^4 + 192\beta^2(3\beta-196)Az^3 - 128\beta^2(\beta-69)A^2z^2 - 144\beta^3(\beta+8)z + 64\beta^3(\beta+3)A]. \tag{12b}$$

Case A: When $\beta = 1$ the integral in (9) can be evaluated in terms of the complete elliptic integrals $K(k)$, $E(k)$ and $\Pi(\alpha^2,k)$ of the first, second, and third kind, respectively, by employing standard methods described in Ref. 9. The labor involved in this evaluation is greatly reduced if one separates off total derivatives in the integrand, yielding zero upon integration around the closed contour Γ_A.

Letting $x_0 < 0 < x_1 < x_2$, where x_0, x_1, x_2 are the real zeros of $Q^2_{\rm mod}(z)$ (cf. Fig. 1), and defining the modulus k by

$$k^2 = \frac{x_2 - x_1}{x_2 - x_0}, \tag{13}$$

we can write the first-, third-, and fifth-order contributions to the sum in the quantization condition (10) as follows

$$L^{(1)} = \frac{2(x_2-x_0)^{1/2}}{3}[(x_1-x_0)K(k) + AE(k)] - \frac{2x_2(x_1-x_0)}{(x_2-x_0)^{1/2}}\Pi(x_0k^2/x_1,k), \tag{14'}$$

$$L^{(3)} = \frac{A(x_2-x_0)^{1/2}}{27-16A^3}[4(A-x_0)x_0(x_1-x_0)K(k) + 3E(k)], \tag{14''}$$

$$L^{(5)} = \frac{A(x_2-x_0)^{1/2}}{15(27-16A^3)^3}\{2[A^2(2727-272A^3) - 3A(3568A^3+13257)x_0 + 90(128A^3+351)x_0^2](x_1-x_0)K(k) + (1088A^6 - 28188A^3 - 47385)E(k)\}. \tag{14'''}$$

In order to obtain the eigenvalues A_n we have inserted (14′)–(14‴) into the quantization condition (10) with $\alpha = \frac{1}{2}$ and solved the resulting equations numerically for $n = 0,1,2,3,4$. The results are given in Table I.

Case C: When $\beta = 0$ the integral (9) defining $L^{(2m+1)}$ can be evaluated in an elementary way. The result is (cf. Eq. (18) in Ref. 8)

TABLE I. For the quantum numbers $n = 0,1,2,3,4$, this table gives the exact eigenvalues A_n (= the absolute value of the $(n+1)$th zero of the Airy function Ai) taken from Table 1 in Ref. 16, and the corresponding approximate eigenvalues in (from top to bottom) the first-, third-, and fifth-order phase-integral approximations. The phase-integral values are obtained in case A from the quantization condition (10) with (14′)–(14‴), and in case C from (16′)–(16‴).

	$A_n = \left(\frac{2\mu}{\hbar^2 a^2}\right)^{1/3} \mathscr{E}_n$		
	Exact	Phase-integral approximations of orders 1,3,5	
n		Case A	Case C
		2.35	2.32
0	2.3381	2.336	2.339
		2.340	2.3376
		4.094	4.082
1	4.0879 49	4.0873	4.0881
		4.0884	4.0879 41
		5.525	5.517
2	5.5205 598	5.5202	5.5205 8
		5.5208	5.5205 589
		6.790	6.785
3	6.7867 081	6.7865	6.7867 16
		6.7869	6.7867 079
		7.947	7.943
4	7.9441 3359	7.9440	7.9441 37
		7.9442	7.9441 3354

$$L^{(2m+1)} = \tfrac{2}{3}A^{3/2} \frac{(-1)^{m+1} b_{2m}}{2^{6m-1}(2m-1)A^{3m}}, \tag{15}$$

where the first few b_{2m} are given in Eqs. (16a–f) in Ref. 8.

Inserting (15) into the quantization condition (10) with $\alpha = \frac{3}{4}$ and solving for A_n, we obtain

$$A_n = \xi_n^{2/3}, \quad \text{first-order approximation}, \tag{16′}$$

$$A_n = \left\{ \frac{\xi_n}{2} \left(1 + \sqrt{1 + \frac{5}{8\xi_n^2}} \right) \right\}^{2/3},$$

third-order approximation, (16″)

$$A_n = \left\{ \frac{\xi_n}{4} \left[1 + \sqrt{1 + \frac{5+\eta_n}{8\xi_n^2}} + \left(2 + 2\sqrt{1 + \frac{5+\eta_n}{8\xi_n^2}} + \frac{5-\eta_n}{8\xi_n^2} - \frac{\eta_n}{4\xi_n^2} \sqrt{1 - \frac{2210}{3\eta_n^2}} \right)^{1/2} \right] \right\}^{2/3},$$

fifth-order approximation, (16‴)

where

$$\xi_n = (n + \tfrac{3}{4}) \frac{3\pi}{2}, \tag{17}$$

and

$$\eta_n = \tfrac{1}{3}\{(5\{6605 + 15912\xi_n^2 + [442(572832\xi_n^4 + 475560\xi_n^2 - 911645)]^{1/2}\})^{1/3} + (5\{6605 + 15912\xi_n^2 - [442(572832\xi_n^4 + 475560\xi_n^2 - 911645)]^{1/2}\})^{1/3} - 5\}. \tag{18}$$

Numerical values of A_0, A_1, A_2, A_3, A_4 are given in Table I.

Since the first-order phase-integral approximation is identical to the first-order JWKB-approximation, we note by comparing with (16′) and (17) that formula (3.5) in Ref. 10 contains a misprint. Furthermore, the first-order JWKB

TABLE II. For the quantum numbers $n = 0,1,2,3,4$, this table gives the exact values of the square of the normalization factor, $|C_n|^2$, and the corresponding approximate values in (from top to bottom) the first-, third-, and fifth-order phase-integral approximations. The exact values are obtained from the formula $|C_n|^2 = [2\sqrt{\pi}\,\mathrm{Ai}'(-A_n)]^{-2}$, the value of the derivative of the Airy function at its $(n+1)$th zero, $\mathrm{Ai}'(-A_n)$, being taken from Table 1 in Ref. 16. The phase-integral values are obtained in case A from (20) in the first-order approximation and by numerical evaluation of (19) in higher orders, and in case C by using (23), by a direct evaluation of (24), and by using the simplified formulas (25′)–(25‴). In the first-order approximation the three different ways of calculating $|C_n|^2$ in case C yield exactly the same result.

	$\|C_n\|^2$				
	Exact	Phase-integral approximations of orders 1,3,5			
n		Case A	Case C, Eq. (23)	Case C, Eq. (24)	Case C, Eqs. (25)
		0.161	0.164	0.164	0.164
0	0.1618	0.1620	0.1615	0.1616	0.1616
		0.1615	0.1621	0.1620	0.1620
		0.1231	0.1237	0.1237	0.1237
1	0.1233 78	0.1234 1	0.1233 6	0.1233 7	0.1233 7
		0.1233 4	0.1233 80	0.1233 79	0.1233 80
		0.1062	0.1064	0.1064	0.1064
2	0.1063 0485	0.1063 2	0.1063 03	0.1063 03	0.1063 03
		0.1062 9	0.1063 050	0.1063 0492	0.1063 050
		0.0958 5	0.0959 8	0.0959 8	0.0959 8
3	0.0959 1703	0.0959 23	0.0959 165	0.0959 166	0.0959 166
		0.0959 10	0.0959 1705	0.0959 1704	0.0959 1704
		0.0886 3	0.0887 1	0.0887 1	0.0887 1
4	0.0886 7115	0.0886 75	0.0886 709	0.0886 710	0.0886 710
		0.0886 67	0.0886 7115	0.0886 7115	0.0886 7115

values given in the column headed $(m/a^2)^{1/3}E_n$(WKB) in Table I in Ref. 5 are in disagreement with the first-order values in our Table I which we claim not to contain any numerical errors.

3. NORMALIZATION FACTORS

Case A: In case A we insert (25a) in Ref. 1 into (23) in Ref. 1 to obtain a phase-integral formula for the normalization factor. Expressed in terms of the dimensionless quantities A and z, introduced in (2) and (3), this formula reads

$$\frac{1}{|C_n|^2} = \int_{\Gamma_A} \frac{dz}{q(z;A_n)}. \tag{19}$$

As was the case with the integral in the quantization condition (7) in case A, the integral in (19) can be expressed in terms of complete elliptic integrals. Using standard methods (see in particular formula (236.20) in Ref. 9), we easily obtain from (19) the first-order expression

$$\frac{1}{|C_n|^2} = \frac{4}{(x_2 - x_0)^{1/2}}\,[x_0K(k) + (x_2 - x_0)E(k)]. \tag{20}$$

However, the analytic evaluation of $|C_n|^2$ in higher orders with the aid of complete elliptic integrals is rather complicated, and therefore we have in the calculation of the third- and fifth-order values resorted to a numerical evaluation of the integral in (19). Numerical results are given in Table II.

Case C: Let us now consider case C. Expressing Eq. (23) in Ref. 1 in terms of the dimensionless quantities A and z, which amounts to formally replacing $2\mu/\hbar^2$ by unity, we obtain the following formula

$$\frac{1}{|C_n|^2} = 2\left(\frac{\partial}{\partial A}\int_{\Gamma_C} q(z;A)\,dz\right)_{A=A_n}. \tag{21}$$

Owing to the simplicity of $Q^2_{\rm mod}(z) = Q^2(z) = A - z$ in the particular problem under consideration we find that

$$(\partial/\partial A)q(z;A) = -(\partial/\partial z)q(z;A). \tag{22}$$

Using this in (21), we immediately obtain the formula

$$\frac{1}{|C_n|^2} = 4q(0 + i0;A_n). \tag{23}$$

If we insert the consistently truncated expression (25b) in Ref. 1 into (23) in Ref. 1, we obtain, as an alternative to (21), the following formula

$$\frac{1}{|C_n|^2} = \int_{\Gamma_C}\frac{dz}{q(z;A_n)} + (1-\delta_{N,0}) \times\left[\sum_{j=0}^{N-1}(-1)^j\left(\frac{1}{2q(z;A_n)}\frac{d}{dz}\right)^{2j+1}\frac{1}{q^2(z;A_n)}\right]_{z=0+i0}. \tag{24}$$

Note that in higher orders this formula will not yield exactly the same results as (21).

By recasting, in a consistent way, the sum in the right-hand member of Eq. (25b) in Ref. 1 into a simplified form not containing derivatives of $q(z)$ with respect to z, one obtains formulas (25b')–(25b''') in Ref. 1. Using these formulas in Eq. (23) in Ref. 1 and evaluating the integrals, we obtain

$$\frac{1}{|C_n|^2} = 4\sqrt{A_n}, \quad \text{first-order approximation}, \tag{25'}$$

$$\frac{1}{|C_n|^2} = 4\sqrt{A_n} + \frac{2\sqrt{\rho}}{3}\left\{2\arctan\sqrt{\rho/A_n} - \frac{\sqrt{3}}{2}\times\ln\frac{A_n+\rho+\sqrt{3\rho A_n}}{A_n+\rho-\sqrt{3\rho A_n}} + \arctan\frac{\sqrt{\rho A_n}}{A_n-\rho}\right\} + \frac{1}{2q^5(0+i0;A_n)}, \quad \text{third-order approximation}, \tag{25''}$$

$$\begin{aligned}\frac{1}{|C_n|^2} = {}& 4\sqrt{A_n} + \frac{64}{3\sqrt{2235}}\left\{\rho_1^{7/2}\left(2\arctan\sqrt{\rho_1/A_n} - \frac{\sqrt{3}}{2}\ln\frac{A_n+\rho_1+\sqrt{3\rho_1A_n}}{A_n+\rho_1-\sqrt{3\rho_1A_n}} + \arctan\frac{\sqrt{\rho_1A_n}}{A_n-\rho_1}\right)\right.\\ &-\rho_2^{7/2}\left(\tfrac{1}{2}\ln\frac{A_n+\rho_2+2\sqrt{\rho_2A_n}}{A_n+\rho_2-2\sqrt{\rho_2A_n}} + \tfrac{1}{2}\ln\frac{A_n+\rho_2+\sqrt{\rho_2A_n}}{A_n+\rho_2-\sqrt{\rho_2A_n}}\right.\\ &\left.\left.-\sqrt{3}\arctan\frac{\sqrt{3\rho_2A_n}}{A_n-\rho_2}\right)\right\} + \frac{1}{2q^5(0+i0;A_n)} - \frac{25}{16q^{11}(0+i0;A_n)}, \quad \text{fifth-order approximation},\end{aligned} \tag{25'''}$$

where

$$\rho = (5/32)^{1/3}, \tag{26a}$$

$$\rho_1 = \left(\frac{\sqrt{2235}+5}{64}\right)^{1/3}, \tag{26b}$$

$$\rho_2 = \left(\frac{\sqrt{2235}-5}{64}\right)^{1/3}, \tag{26c}$$

and where the principal value of arctan is to be used. Note that (25'')–(25'''), which are simpler than the explicit third- and fifth-order expressions obtained from (24) directly, do not give exactly the same results as (24) in the third and fifth orders, respectively.

In Table II we include numerical values of $|C_n|^2$ obtained in case C in the three alternative ways described above, i.e., by using (23), by using (24), and by using the simplified formulas (25')–(25'''). As can be seen from Table II the differences between these numerical values are essentially insignificant.

4. EXPECTATION VALUES

According to the phase-integral formulas (26a) and (26b) in Ref. 1 for the quantal expectation value of a multiplicative operator $f(z)$ with respect to a bound state with quantum number n, we have the $(2N+1)$ th-order formulas

TABLES IIIa–c. These tables give the expectation values $\langle z\rangle_n$, $\langle z^2\rangle_n$, and $\langle z^3\rangle_n$ for the quantum numbers $n = 0,1,2,3,4$. The exact values are obtained from (36a–c) with the use of the exact eigenvalues A_n. The phase-integral values of $\langle z^\nu\rangle_n$ in (from top to bottom) the first-, third-, and fifth-order approximations are obtained in case A from (29a–c) in the first-order approximation and by numerical evaluation of (27) in higher orders. In case C the results are obtained from (30a–c) in the first-order approximation and in higher orders in the two ways of either direct evaluation of (28) with (24) or of using the simplified form obtained by using (26b″–b‴) in Ref. 1 for the expression within curly brackets in (28) and formulas (25″)–(25‴) for $|C_n|^2$.

TABLE IIIa.

	$\langle z\rangle_n = \left(\frac{2\mu a}{\hbar^2}\right)^{1/3}\langle r\rangle_n$			
	Exact	Phase-integral approximations of orders 1,3,5		
n		Case A	Case C, Eqs. (24) and (28)	Case C, simplified form
		1.57	1.55	1.55
0	1.5587	1.557	1.560	1.560
		1.5581	1.5582	1.5584
		2.730	2.721	2.721
1	2.7253 00	2.7246	2.7254	2.7254
		2.7250	2.7252 9	2.7252 93
		3.683	3.678	3.678
2	3.6803 732	3.6800	3.6804 0	3.6803 9
		3.6802	3.6803 72	3.6803 726
		4.527	4.523	4.523
3	4.5244 721	4.5242	4.5244 82	4.5244 78
		4.5244	4.5244 718	4.5244 719
		5.298	5.295	5.295
4	5.2960 890	5.2959	5.2960 93	5.2960 91
		5.2960 1	5.2960 890	5.2960 890

$$\langle f(z)\rangle_n = |C_n|^2 \int_{\Gamma_A} f(z)\frac{dz}{q(z;A_n)} \quad \text{(case A)} \tag{27}$$

and

$$\langle f(z)\rangle_n = |C_n|^2\left\{ \int_{\Gamma_C} f(z)\frac{dz}{q(z;A_n)} + (1-\delta_{N,0}) \times\left[\sum_{j=0}^{N-1}(-1)^j\left(\frac{1}{2q(z;A_n)}\frac{d}{dz}\right)^{2j+1}\frac{f(z)}{q^2(z;A_n)}\right]_{z=0+i0}\right\} \quad \text{(case C)}, \tag{28}$$

where $|C_n|^2$ are the normalization factors obtained from Eqs. (19) and (24), respectively.

Case A: When $f(z) = z^\nu$, where ν is an integer, the integral in (27) can be expressed in terms of complete elliptic

TABLE IIIb.

	$\langle z^2\rangle_n = \left(\frac{2\mu a}{\hbar^2}\right)^{2/3}\langle r^2\rangle_n$			
	Exact	Phase-integral approximations of orders 1,3,5		
n		Case A	Case C, Eqs. (24) and (28)	Case C, simplified form
		2.86	2.87	2.87
0	2.916	2.914	2.922	2.920
		2.915	2.914	2.914
		8.88	8.89	8.89
1	8.9127 1	8.912	8.9135	8.9133
		8.9124	8.9126 2	8.9126 6
		16.23	16.23	16.23
2	16.2541 76	16.253	16.2544	16.2544
		16.2539	16.2541 6	16.2541 70
		24.55	24.55	24.55
3	24.5650 17	24.5644	24.5651	24.5651
		24.5648	24.5650 14	24.5650 15
		33.64	33.64	33.64
4	33.6582 71	33.6578	33.6583 3	33.6583 1
		33.6580	33.6582 70	33.6582 71

TABLE IIIc.

n	Exact	Case A	Case C, Eqs. (24) and (28)	Case C, simplified form
	$\langle z^3\rangle_n = \frac{2\mu a}{\hbar^2}\langle r^3\rangle_n$	Phase-integral approximations of orders 1,3,5		
		5.6	5.7	5.7
0	6.271	6.265	6.28	6.279
		6.274	6.267	6.268
		31.0	31.1	31.1
1	31.6583	31.650	31.661	31.660
		31.660	31.6579	31.6581
		76.7	76.8	76.8
2	77.3418 5	77.334	77.343	77.343
		77.344	77.3417 8	77.3418 1
		142.7	142.8	142.8
3	143.3276 6	143.320	143.3282	143.3281
		143.329	143.3276 4	143.3276 4
		229.0	229.0	229.0
4	229.6164 0	229.608	229.6168	229.6167
		229.618	229.6163 9	229.6164 0

integrals. Using a similar technique as in Sec. 3 (case A) for expressing the normalization factor in terms of complete elliptic integrals, we obtain in the first-order approximation

$$\langle z\rangle_n = \tfrac{2}{3}A_n, \tag{29a}$$

$$\langle z^2\rangle_n = \tfrac{8}{15}A_n^2 - \frac{1}{10}\frac{K(k)}{x_0K(k)+(x_2-x_0)E(k)}, \tag{29b}$$

$$\langle z^3\rangle_n = \tfrac{16}{35}A_n^3 - \tfrac{1}{7} - \tfrac{3}{35}A_n\frac{K(k)}{x_0K(k)+(x_2-x_0)E(k)}, \tag{29c}$$

$$\langle z^\nu\rangle_n = \frac{1}{2(2\nu+1)}(4\nu A_n\langle z^{\nu-1}\rangle_n - (\nu-1)\langle z^{\nu-3}\rangle_n), \quad \nu\geqslant 1. \tag{29d}$$

In higher orders the analytical evaluation of the expressions for the expectation values in terms of real complete elliptic integrals is rather cumbersome. Therefore, we have in the third and fifth orders computed the integrals in (27) numerically. Numerical results for $\langle z\rangle_n$, $\langle z^2\rangle_n$, and $\langle z^3\rangle_n$ are presented in Tables IIIa–c.

Case C: Letting $f(z)=z^\nu$ in (28) we can calculate the expectation values $\langle z^\nu\rangle_n$. We emphasize that in case C it is essential that $\nu\geqslant 0$ for the formulas to be valid; cf. the discussion after Eq. (26b‴) in Ref. 1.

In the first-order approximation formula (28) together with formula (24) immediately yields the following simple expressions:

$$\langle z\rangle_n = \tfrac{2}{3}A_n, \tag{30a}$$

$$\langle z^2\rangle_n = \tfrac{8}{15}A_n^2, \tag{30b}$$

$$\langle z^3\rangle_n = \tfrac{16}{35}A_n^3, \tag{30c}$$

$$\langle z^\nu\rangle_n = \frac{2\nu A_n}{2\nu+1}\langle z^{\nu-1}\rangle_n, \quad \nu\geqslant 1. \tag{30d}$$

If we calculate $\langle z^\nu\rangle_n$ in the third- and fifth-order phase-integral approximations by using (24) and (28) directly, we end up with rather unwieldy expressions, which we refrain from giving here. However, numerical values obtained from these latter expressions are included in Tables IIIa–c.

Let us now calculate $\langle z^\nu\rangle_n$ by using the simplified phase-integral formulas obtained by replacing the expression within the curly brackets in (28) (with $f(z)=z^\nu$), by the simpler expressions (26b′), (26b″), and (26b‴) in Ref. 1 (expressed in terms of A and z), appropriate to the order of approximation under consideration, and by using for $|C_n|^2$ in (28) the simplified formulas (25′)–(25‴). In the first-order approximation we get, of course, the expressions (30a–d). In the third- and fifth-order approximations the integrals appearing in both (28) and in the simplified formulas can be written

$$\int_{\Gamma_C}\frac{z^\nu dz}{q(z;A_n)} = (\delta_{\nu,0}-1)\sum_{\mu=0}^{\nu-1}\binom{\nu}{\mu}(-A_n)^{\nu-\mu}\int_{\Gamma_C}\frac{z^\mu dz}{q(z;A_n)} + (-1)^\nu 4\sum_{\lambda=0}^{[\nu/3]}\frac{(-1)^\lambda A_n^{[2(\nu-3\lambda)+1]/2}}{2(\nu-3\lambda)+1}\rho^{3\lambda} + \frac{2\rho^{(2\nu+1)/2}}{3}\Bigg\{2\arctan\sqrt{\rho/A_n} - \sin\left(\frac{(2\nu+1)\pi}{3}\right)\ln\frac{A_n+\rho+\sqrt{3\rho A_n}}{A_n+\rho-\sqrt{3\rho A_n}} + 2\cos\left(\frac{(2\nu+1)\pi}{3}\right)\arctan\frac{\sqrt{\rho A_n}}{A_n-\rho}\Bigg\},$$

$$\nu\geqslant 0, \text{ third-order approximation}, \tag{31′}$$

$$\int_{\Gamma_C}\frac{z^\nu dz}{q(z;A_n)} = (\delta_{\nu,0}-1)\sum_{\mu=0}^{\nu-1}\binom{\nu}{\mu}(-A_n)^{\nu-\mu}\int_{\Gamma_C}\frac{z^\mu dz}{q(z;A_n)}$$

$$+(-1)^\nu\frac{128}{\sqrt{2235}}\sum_{\lambda=0}^{[\nu/3]}\frac{(-1)^\lambda A_n^{[2(\nu-3\lambda)+1]/2}}{2(\nu-3\lambda)+1}\{\rho_1^{3(\lambda+1)}$$

TABLE IV. For the quantum numbers $n = 0,1,2,3,4$ this table gives the exact values of the quantity $(2\pi\hbar^2/\mu a)|\psi_n(0)|^2$ obtained from the exact formula (37) as well as the corresponding approximate values in (from top to bottom) the first-, third-, and fifth-order phase-integral approximations. The phase-integral values are obtained in case A from (32) and in case C from (33) with $|C_n|^2$ given by (23), (24), and (25′)–(25‴), respectively.

	$\frac{2\pi\hbar^2}{\mu a}\lvert\psi_n(0)\rvert^2$				
	Exact	Phase-integral approximations of orders 1,3,5			
n		Case A	Case C, Eqs. (33) and (23)	Case C, Eqs. (33) and (24)	Case C, Eqs. (33) and (25)
		1	1	1	1
0	1	1	1	1.001	1.0006
		1	1	0.9995	0.9996
		1	1	1	1
1	1	1	1	1.0000 4	1.0000 2
		1	1	0.9999 2	0.9999 97
		1	1	1	1
2	1	1	1	1.0000 06	1.0000 03
		1	1	0.9999 994	0.9999 997
		1	1	1	1
3	1	1	1	1.0000 02	1.0000 01
		1	1	0.9999 9991	0.9999 9998
		1	1	1	1
4	1	1	1	1.0000 007	1.0000 004
		1	1	0.9999 9998	1.0000 0000

$$+(-1)^{\lambda}\rho_2^{3(\lambda+1)}\} + \frac{64}{3\sqrt{2235}}\Bigg\{\rho_1^{(2\nu+7)/2} \times\Bigg[2\arctan\sqrt{\rho_1/A_n} - \sin\Big(\frac{(2\nu+1)\pi}{3}\Big)\ln\frac{A_n+\rho_1+\sqrt{3\rho_1A_n}}{A_n+\rho_1-\sqrt{3\rho_1A_n}} + 2\cos\Big(\frac{(2\nu+1)\pi}{3}\Big)\arctan\frac{\sqrt{\rho_1A_n}}{A_n-\rho_1}\Bigg] + \rho_2^{(2\nu+7)/2}\Bigg[\frac{(-1)^{\nu+1}}{2}\ln\frac{A_n+\rho_2+2\sqrt{\rho_2A_n}}{A_n+\rho_2-2\sqrt{\rho_2A_n}} - \sin\Big(\frac{(2\nu+1)\pi}{6}\Big)\ln\frac{A_n+\rho_2+\sqrt{\rho_2A_n}}{A_n+\rho_2-\sqrt{\rho_2A_n}} + 2\cos\Big(\frac{(2\nu+1)\pi}{6}\Big)\arctan\frac{\sqrt{3\rho_2A_n}}{A_n-\rho_2}\Bigg]\Bigg\},\quad \nu\geqslant 0,$$

fifth-order approximation, (31″)

where $\delta_{\nu,0}$ is the Kronecker symbol, $[\nu/3]$ means the integer part of $\nu/3$, and ρ, ρ_1, and ρ_2 are given by (26a–c). Inserting these explicit expressions (31′) and (31″) into the simplified formulas for $\langle z^{\nu}\rangle_n$, obtained as described above, we get numerical results in the third- and fifth-order approximations, which do not differ significantly from the corresponding result obtained by using (24) and (28) directly (see Tables IIIa–c).

Instead of evaluating the integrals occurring in the phase-integral formulas for $|C_n|^2$ and $\langle z^{\nu}\rangle_n$ in case C analytically, one can evaluate these integrals by numerical means. This is a very simple and rapid process since the integrands are smooth and slowly varying on the finite contour of integration Γ_C, which can be taken as the nonclosed circle from $z = 0 + i0$ to $z = 0 - i0$ centered on $z = A_n$ and with radius A_n.

5. PROBABILITY DENSITIES AT THE ORIGIN

Case A: In case A the probability density at the origin can be calculated by applying phase-integral formula (10) in Ref. 2, which, since $dV/dr = a$, immediately yields the result

$$|\psi_n(0)|^2 = \frac{\mu a}{2\pi\hbar^2}, \tag{32}$$

valid for any n in any order of the phase-integral approximations used. Comparing the exact formula (9) in Ref. 2 with the phase-integral formula (10) in Ref. 2 we realize that the latter formula yields the exact answer when dV/dr is a constant. Hence, (32) is exactly valid.

Case C: In case C the phase-integral formula for the probability density at the origin to be used is formula (19) in Ref. 2. The quantities C_n^2 and $q(0)$ appearing in this formula are those that are associated with the radial Schrödinger equation (1), in which the independent variable is r. Using instead the quantities $|C_n|^2$ and q that are associated with the differential equation (4) with (5), and thus are expressed in terms of the dimensionless quantities A and z, we get the following formula

$$|\psi_n(0)|^2 = \frac{2\mu a}{\pi\hbar^2}\,|C_n|^2 q(0+i0;A_n). \tag{33}$$

Depending on how we calculate $|C_n|^2$ in (33), we get slightly different results for $|\psi_n(0)|^2$; see Sec. 3 (case C). In particular we note that by inserting (23) into (33) we recover formula (32).

Numerical values of $(2\pi\hbar^2/\mu a)|\psi_n(0)|^2$ obtained from (32) and from (33), with $|C_n|^2$ evaluated in the three different ways described in Sec. 3 (case C), are given in Table IV.

Recently Müller-Kirsten, Hite, and Bose [Ref. 11, Sec. 5] have used a semiclassical formula for $|\psi_n(0)|^2$ derived by Quigg and Rosner[12] by the use of the first-order JWKB-approximation and the well-known simple qualitative argument for averaging the square of a rapidly oscillation wave-function under an integral sign. Because of such rough approximations one should at first sight expect the semiclassical formula for $|\psi_n(0)|^2$ in question to be valid only when the quantum number n is very large. However, the formula obtained by Quigg and Rosner is a special case of the phase-integral formula earlier derived by N. Fröman (Eq. (20) in Ref. 2), namely the special case that the first-order approximation is used and that $V(0) = 0$. From N. Fröman's derivation, based on a phase-integral method in which arbitrary-order phase-integral approximations are used, it is seen that the resulting formula (20) in Ref. 2 for $|\psi_n(0)|^2$, as well as the closely related formula (19) in Ref. 2 [and thus also (33) in the present paper], should be accurate also for small values of the quantum number n. The numerical results in Table IV are seen to support this statement in a direct way as regards formula (33) and thereby in an indirect way as regards formula (20) in Ref. 2.

6. EXACT EXPRESSIONS

The exact expressions for the bound-state solutions of (4) with (5) are (cf. Eq. (10) in Ref. 8)

$$u_n(z) = D_n \operatorname{Ai}(z - A_n), \quad n = 0,1,2,\cdots, \tag{34}$$

where D_n is a normalization factor and A_n, which is the eigenvalue, is equal to the absolute value of the $(n+1)$th zero of the Airy function Ai.

In (4), with (5) inserted, we set u equal to u_n, A equal to A_n, and multiply the resulting equation by $z^\nu u_n'$, $\nu \geqslant 0$, letting the prime denote differentiation with respect to z. Then we integrate from $z = 0$ to $z = +\infty$. In this way we obtain, for $\nu = 0$, the formula (cf. Eq. (6) in Ref. 13 and Eq. (6.9) in Ref. 14)

$$1/|D_n|^2 = [\operatorname{Ai}'(-A_n)]^2 \tag{35}$$

and, for $\nu = 1,2,3,\cdots$, the formulas

$$\langle z \rangle_n = \tfrac{2}{3} A_n, \tag{36a}$$

$$\langle z^2 \rangle_n = \tfrac{8}{15} A_n^2, \tag{36b}$$

$$\langle z^3 \rangle_n = \tfrac{16}{35} A_n^3 + \tfrac{3}{7}, \tag{36c}$$

$$\langle z^\nu \rangle = \frac{\nu}{2(2\nu+1)} [4A_n \langle z^{\nu-1} \rangle_n - (\nu - 1) \times (\nu - 2) \langle z^{\nu-3} \rangle_n], \quad \nu \geqslant 1. \tag{36d}$$

Inserting $V(r) = ar$ into Eq. (9) in Ref. 2, we obtain the formula (cf. Eq. (3) in Ref. 3, Eq. (69) in Ref. 5, Eq. (A3) in Ref. 10, and Eq. (12) in Ref. 15)

$$|\psi_n(0)|^2 = \frac{\mu a}{2\pi\hbar^2}. \tag{37}$$

In the numerical evaluation of the exact quantities $|D_n|^2$ with the use of formula (35), and $\langle z^\nu \rangle_n$ given by formulas (36a–d), use has been made of the values of $\operatorname{Ai}'(-A_n)$ and A_n given in Table 1 in Ref. 16. Note that due to the normalization used in this table the exact quantity that corresponds to the square of the phase-integral normalization factor, $|C_n|^2$, is the quantity $|D_n|^2/4\pi$, the numerical values of which are displayed in the column headed "Exact" in Table II for $n = 0,1,2,3,4$.

We remark that by using the differential equation (4) with (5) it is easy to obtain, in a well-known way, exact expressions for expectation values of powers of the differential operator d/dz (i.e., of the momentum operator) in terms of the expectation values (36).

After completing the work presented in this section, the present author became aware of a paper[17] containing exact calculations of (in our notations) $|D_n|^2$, $\langle z^\nu \rangle_n$ $(\nu \geqslant 1)$, and $|\psi_n(0)|^2$, along similar lines as just described. However, our formula (36d), which directly relates expectation values $\langle z^\nu \rangle_n$ for different ν, seems to be more straightforward and simpler to use than formulas (17) and (20) in Ref. 17. We also remark that Eqs. (11) and (12) in Ref. 17 contain obvious misprints.

ACKNOWLEDGMENTS

I should like to thank Dr. Nanny Fröman and Dr. Per Olof Fröman for introducing me to their phase-integral method and for continuous suggestions during the time the present work was carried out. I am also grateful to Dr. Finn Karlsson and Dr. Bengt Lundborg for fruitful discussions.

[1]N. Fröman, Phys. Rev. A **17**, 493 (1978).
[2]N. Fröman, J. Math. Phys. **19**, 1141 (1978).
[3]B.J. Harrington, S.Y. Park, and A. Yildiz, Phys. Rev. Lett. **34**, 168 (1975).
[4]J.F. Gunion and R.S. Willey, Phys. Rev. D **12**, 174 (1975).
[5]J.S. Kang and H.J. Schnitzer, Phys. Rev. D **12**, 841 (1975).
[6]H.J.W. Müller-Kirsten, Phys. Rev. D **12**, 1103 (1975).
[7]J. Pumplin, W. Repko, and A. Sato, Phys. Rev. Lett. **35**, 1538 (1975).
[8]B. Thidé, J. Math. Phys. **17**, 1222 (1976).
[9]P.F. Byrd and M.D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists*, revised second edition (Springer, Berlin–Heidelberg–New York, 1971).
[10]H.J. Schnitzer, Phys. Rev. D **13**, 74 (1976).
[11]H.J.W. Müller-Kirsten, G.E. Hite, and S.K. Bose, J. Math. Phys. **20**, 1878 (1979).
[12]C. Quigg and J.L. Rosner, Phys. Rev. D **17**, 2364 (1978).
[13]T.W. Nee, J.F. Koch, and R.E. Prange, Phys. Rev. **174**, 758 (1968).
[14]A.P. Balachandran, R. Ramachandran, J. Schechter, K.C. Wali, and H. Rupertsberger, Phys. Rev. D **13**, 361 (1976).
[15]J.S. Kang and H.J. Schnitzer, Phys. Rev. D **12**, 2791 (1975).
[16]M.E. Sherry, AFCRC-TR-59-135, "The zeros and maxima of the Airy function and its first derivative to 25 significant figures" (Air Force Cambridge Research Center, Bedford, Massachusetts, 1959).
[17]P. Leal Ferreira and J.A. Castilho Alcarás, Lett. Nuovo Cimento **14**, 500 (1975).

Prolongation structure for a nonlinear equation with explicit space dependence

A. R. Chowdhury and T. Roy
Physics Department, Jadavpur University, Calcutta-700032, India

(Received 27 November 1978; accepted for publication 22 June 1979)

A nonlinear Schroedinger equation with a term depending explicitly on a space variable, e.g., $i\psi_t + \psi_{xx} + (-2\alpha x + 2|E|^2)\psi = 0$ with $\psi = Ee^{i\phi}$ has been treated in the language of differential forms and prolongation. The inverse scattering equations previously invented by Liu and Chen are obtained. A unique feature of the analysis is explicit space–time dependence of the pfaffian forms.

From the very inception of the differential form approach[1] to the nonlinear partial differential equations, it has been felt that the only logical way to deduce the inverse scattering transform (IST) is through the technique of Lie and Cartan. But until now, the equations which have been considered do not contain explicit (x,t) dependence.[2] Thus, the structure of the Pfaffian forms also came out to be independent of (x,t). Here we report an application of the differential form approach to a nonlinear Schrödinger equation with a term $2\alpha x\psi$, considered by Chen and Liu[3] in connection with their investigation about soliton formation in nonuniform medium. It is also worth mentioning that the IST formalism for the equation under consideration was also obtained by them with which they constructed the N-soliton solution. The only difference was that the eigenvalue changed linearly with time in contrast to the original IST of Zakharov and Shabat.[4] We thought it would be really interesting to observe how a time-dependent eigenvalue problem could be fitted into the framework of Wahlquist and Estabrook.[1] Above all the differential geometric approach helps to justify the usual heuristic derivation of the IST. While the hierarchy of equations considered until now all sustain the solitary wave-type solutions due to a balance between the nonlinearity and dispersion, the present nonlinear equation possesses such solutions due to a balance between nonlinearity, dispersion, and nonuniformity of the medium. So quite a distinct physical phenomenon is governed by Eq. (1) below.

1. FORMULATION

The nonlinear Schrödinger equation under consideration describing the wave motion in a nonuniform medium reads

$$i\psi_t + \psi_{xx} + (-2\alpha x + |E|^2)\psi = 0. \tag{1}$$

To cast this into a set of different forms, we define $\psi_x = z$, $\bar{\psi}_x = \bar{z}$. It is then easy to observe that Eq. (1) and its complex conjugate is equivalent to

$$\begin{aligned}
\alpha_1 &= d\psi \wedge dt - z\, dx \wedge dt,\\
\bar{\alpha}_1 &= d\bar{\psi} \wedge dt - \bar{z}\, dx \wedge dt,\\
\alpha_2 &= -i\, d\psi \wedge dx + dz \wedge dt + (-2\alpha x + 2\bar{\psi}\psi)\psi\, dx \wedge dt,\\
\bar{\alpha}_2 &= i\, d\bar{\psi} \wedge dx + d\bar{z} \wedge dt + (-2\alpha x + 2\psi\bar{\psi})\bar{\psi}\, dx \wedge dt,
\end{aligned} \tag{2}$$

when proper sectioning is taken.

One can immediately see that this system of differential form is closed under exterior differentiation. That is, $d\alpha_i = \Sigma\eta_{ij} \wedge \alpha_j$, η_{ij} are a set of 1-forms. So we may now proceed for the determination of the nonexact differential forms or Pfaffian forms,

$$\omega^k = dy^k + F^k(\psi,\bar{\psi},z,\bar{z},x,t)\, dx + G^k(\psi,\bar{\psi},z,\bar{z},x,t)\, dt, \tag{3}$$

by demanding that

$$d\omega^k = \Sigma f_i\alpha_i + \Sigma g_i\bar{\alpha}_i + \Sigma\eta_{ij} \wedge \omega^j, \tag{4}$$

so that $d\omega^k$ is in the extended ideal. Equation (4) yields

$$\begin{aligned}
G_x - F_t &= zG_\psi + \bar{z}G_{\bar{\psi}} - iF_\psi(-2\alpha x + \bar{\psi}\psi^2)\\
&\quad + iF_{\bar{\psi}}(-2\alpha x\bar{\psi} + 2\psi\bar{\psi}^2) + G^iF^k_{y_i} - F^iG^k_{y_i},
\end{aligned} \tag{5}$$

along with

$$\begin{aligned}
iF_\psi &= G_z,\\
-iF_{\bar{\psi}} &= G_{\bar{z}}.
\end{aligned} \tag{6}$$

In order to proceed further with our calculation we need the information regarding the dependence of F and G upon the primitive variables. At this point it should be noted that the presence of explicit space dependence in the r.h.s. of Eq. (5) enforces an explicit (x,t) dependence on F and G. For simplicity we assume $F_x = 0$, so that we immediately deduce the following

$$\begin{aligned}
&G_{xx} = G_{\psi z} = G_{\bar{\psi}z} = 0,\\
&G_{zz} = G_{\bar{z}\bar{z}} = 0, \quad G_{\psi\psi\psi} = 0,\\
&G_{ttt} = 0, \quad F_{\psi\psi} = 0 = F_{\bar{\psi}\psi}, \quad \text{etc.},
\end{aligned} \tag{7}$$

which in turn implies

$$\begin{aligned}
F &= tx_0 + \psi x_1 + \bar{\psi}\bar{x}_1,\\
G &= t^2x_5 + \bar{\psi}\psi x_{10} + xx_{11}\\
&\quad + \psi t x_8 + \bar{x}_8\bar{\psi}t + zx_1 + \bar{z}\bar{x}_1.
\end{aligned} \tag{8}$$

In this context some comments should be kept in mind. From our above analysis it is quite clear that we could always keep terms of the form $\bar{\psi}\psi$ in F, and $\psi^2,\bar{\psi}^2$ and many other higher order terms in G. But we have tried here to retain the minimal structure of the functions F and G.

2. LIE ALGEBRA AND THE REPRESENTATION OF THE IST

Substituting these expressions of F and G in (5) and equating the coefficients of differential 2-forms we obtain

$$[x_0,x_1] = -x_8, \quad [x_1,\bar{x}_1] = -x_{10}, \quad [x_0,x_5] = 0,$$

 0022-2488/80/061416-02$1.00

$$[x_0,x_{11}]=0,\quad [x_1,x_{11}]=-2i\alpha x_1,\quad [\bar{x}_1,x_{11}]=2i\alpha\bar{x}_1,$$
$$[x_1,x_8]=0,\quad [\bar{x}_1,\bar{x}_8]=0,$$
$$[\bar{x}_1,x_{10}]=ix_1,\quad [\bar{x}_1,x_{10}]=-i\bar{x}_1, \tag{9}$$
$$[x_0,x_8]+[x_1,x_5]=0,$$
$$[x_0,\bar{x}_8]+[x_1,\bar{x}_8]=0,$$
and
$$[x_0,x_{10}]+[\bar{x}_1,x_8]+[x_1,\bar{x}_8]=0,$$
along with $x_0=x_{11},\quad x_5=\sigma x_{10}$.

It is seen that the last relation in Eq. (9) is helpful for obtaining the closure of the algebra under the above commutation relations and Jacobi identities. The algebra depicted in Eq. (9) can be represented in the following way over two prolongation variables (y_1,y_2)[5]:

$$x_6=ix_1=-(2\alpha i)^{-1}x_8=iy_2\frac{\partial}{\partial y_1},$$
$$\bar{x}_6=i\bar{x}_1=(2\alpha i)^{-1}\bar{x}_8=iy_1\frac{\partial}{\partial y_2}, \tag{10}$$
$$x_5=2\alpha x_0=-2i\alpha^2\left(y_1\frac{\partial}{\partial y_1}-y_2\frac{\partial}{\partial y_2}\right),$$
$$x_{10}=i\left(y_1\frac{\partial}{\partial y_1}-y_2\frac{\partial}{\partial y_2}\right),\quad x_{11}=-\alpha x_{10}.$$

If we now substitute these representations in (3), we obtain the two Pfaffian forms given by

$$\omega^1=dy^1+(-it\alpha y_1+y_2\psi)\,dx$$
$$+(-2it^2\alpha^2y_1+iy_1\bar{\psi}\psi-xi\alpha y_1$$
$$+2\psi t\alpha y_2+y_2z)\,dt, \tag{11}$$
$$\omega^2=dy^2+(it\alpha y_2-y_1\bar{\psi})\,dx$$
$$+(2it^2\alpha^2y_2-\bar{\psi}\psi iy_2+xi\alpha y_2-2\bar{\psi}t\alpha y_1-\bar{z}y_1)\,dt,$$

which on proper sectioning yields the inverse scattering equations of Chen and Liu. The most interesting point to be noted is that the space derivative equations can be thrown into an isospectral form if and only if we identify αt to be equal to some quantity ζ, which serves as the eigenvalue of the system. Thus, the equations of IST as obtain in Ref. 3 are obtained, with the condition $d\zeta/dt=\alpha$ or $\zeta=\alpha t$, so that a time-varying eigenvalue takes care of the explicit space-dependent term in the original nonlinear equation. Lastly, it is interesting to note that the present equation can be connected to the usual nonlinear Schrödinger equation via a transformation of both dependent and independent variable transformation, given by

$$\psi(x,t)=\phi(x,t)\exp(i\alpha xt+i\beta t^3),$$
$$\text{along with } \xi=x+ft^2, \tag{12}$$
$$\tau=t,$$

for suitable choice of the constants α',β,α,f.

The consequence of such a transformation on the structure of F and G can be immediately ascertained if we note that these transformations induce a change of variable in the space of primitive variables also. Thus, it is not very difficult to visualize the forms of new F and G. Lastly, it can be mentioned that Eq. (12) yields a consistency check on our calculation of F and G. Other interesting features concerning the above equation will be communicated shortly.

[1]F.B. Estabrook and H.D. Wahlquist, J. Math. Phys. **17**, 1293 (1976).
[2]"Bäcklund Transformation", in *Lecture Notes In Mathematics* (Springer-Verlag, Berlin, Yr.), Vol. 515, pp. 136–61.
[3]H.H. Chen and C.S. Liu, Phys. Rev. Lett. **37**, 693 (1976).
[4]V.E. Zakharov and A.B. Shabat, JETP Sov. Phys. **34**, 62 (1972).
[5]H.W. Guggenheimer, *Differential Geometry* (McGraw-Hill, New York, 1958), pp. 125–45.

A homogeneous Hilbert problem for the Kinnersley–Chitre transformations of electrovac spacetimes [a)]

Isidore Hauser and Frederick J. Ernst
Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616

(Received 9 January 1980; accepted for publication 8 February 1980)

The homogeneous Hilbert problem which we recently formulated for Kinnersley–Chitre transformations of vacuum spacetimes is here generalized to handle transformations of electrovac spacetimes. This provides in particular a simple derivation of our previously published integral equation.

1. INTRODUCTION

In a series of papers[1-3] we have been laying the groundwork for attempting to prove the validity of several conjectures which have been made concerning Kinnersley–Chitre (KC) transformations of vacuum and electrovac spacetimes.[4] With this objective in mind we introduced in Ref. 3 a homogeneous Hilbert problem (HHP), each solution of which gives the result of applying an element of the KC representation of the Geroch group K.[5] The main objective of the present paper is the generalization of that HHP so that each solution will give the result of applying an element u of the KC representation of the Kinnersley group K',[6] i.e., the electrovac generalization of K. In addition, we shall provide a simple derivation of the electrovac integral equation which we described in Ref. 2, where we promised that such a derivation would be forthcoming.

We shall let V' denote the set of all electrovac spacetimes and associated Maxwell fields for which there exist coordinates x^1, x^2, x^3, x^4 such that the line element has the form (signature $= +2$)

$$g_{ij}\delta x^i\delta x^j + g_{ab}\delta x^a\delta x^b \quad (i,j=1,2),(a,b=3,4), \tag{1.1}$$

where g_{ij} and g_{ab} depend at most on x^1, x^2, and where the 2×2 matrix h whose elements are

$$h_{ab} := g_{ab} \quad (a,b=3,4) \tag{1.2}$$

obeys the condition that $d(\det h)$ is not zero and is not a null 1-form, while the Maxwell 2-form $\mathscr{F}$ has vanishing Lie derivatives with respect to the Killing vector fields $\mathbf{X}_3$ and $\mathbf{X}_4$, and $(X_3X_4)\ulcorner\mathscr{F}=0$, where X_a is the covector (1-form) of $\mathbf{X}_a$.[7]

As we showed in Sec. 2 of Ref. 2, a given member of V' determines up to a gauge transformation a 3×3 matrix potential $F_0(\mathbf{x},t)$, which depends not only upon $\mathbf{x}=(x^1,x^2)$ but also upon a complex parameter t. For proving some theorems we have found it to be convenient to restrict the gauge of the F-potential. Among other things we shall show that the gauge can be selected[8] so that for fixed $\mathbf{x}$, $F_0(\mathbf{x},t)$ is holomorphic in a neighborhood of $t=0$, and

$$F_0(\mathbf{x},t)\begin{pmatrix} t & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$$

is holomorphic in a neighborhood of $t=\infty$ (including ∞).

To each smooth contour L symmetric with respect to the real axis and surrounding the origin in the complex plane there corresponds a representation K'_L of the group K'. K'_L is the set of all 3×3 complex matrix analytic functions $u(t)$ such that

$$\det u(t)=1,\quad u(t)^{\mathsf T}\mathfrak{G}(t)u(t)=\mathfrak{G}(t), \tag{1.3}$$

$$\mathfrak{G}(t) := \begin{pmatrix} 0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 0 & -it/2\end{pmatrix},\quad u(t)^\dagger := \text{h.c. of } u(t^*), \tag{1.4}$$

and such that $u(t)$ is holomorphic on L and

$$\begin{pmatrix} t^{-1} & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} u(t)\begin{pmatrix} t & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$$

is holomorphic at $t=\infty$.

To effect the transformation of any given potential $F_0(\mathbf{x},t)$ for any given $u(t)$ in K'_L, we restrict $\mathbf{x}$ to a compact region U_c of the real plane such that $F_0(\mathbf{x},t)$ is holomorphic on $L+L_+$ for every $\mathbf{x}$ in U_c. Here L_+ denotes the open set inside L. The solution of the HHP which we shall formulate in this paper yields an output potential $F(\mathbf{x},t)$ which automatically satisfies all of the relations that a bona fide F-potential must satisfy. From this $F(\mathbf{x},t)$ a new member of V' can be constructed.

2. THE F-POTENTIAL

In Ref. 2 we introduced a 3×3 matrix generalization $F^{(1)}(\mathbf{x})$ of the complex Ernst potentials $\mathscr{E}$ and Φ associated with a member of V'. Knowledge of this field permits one to construct a certain 1-form function $\Gamma(\mathbf{x},t)$ of the complex parameter t. The differential equation

$$dF(t)=\Gamma(t)\Omega F(t), \tag{2.1}$$

where

$$\Omega := \begin{pmatrix} 0 & i & 0\\ -i & 0 & 0\\ 0 & 0 & 1\end{pmatrix}, \tag{2.2}$$

was shown to be completely integrable, and it was also shown that the subsidiary conditions

$$F(0)=\Omega,\quad \dot F(0)=F^{(1)}, \tag{2.3}$$

$$F(t)^\dagger\mathscr{H}(t)F(t)=i\mathfrak{G}(t), \tag{2.4}$$

could be imposed, where

$$\mathscr{H}(t) := i\mathfrak{G}(t) - it\,[\mathfrak{G}(0)F^{(1)}\Omega + \Omega F^{(1)\dagger}\mathfrak{G}(0)]. \tag{2.5}$$

[a)]Supported in part by National Science Foundation grants PHY 75-08750 and PHY 79-08627.

The 1-form $\Gamma(t)$ was defined by

$$\Gamma(t) := tA(t)^{-1}\, dF^{(1)}, \qquad (2.6)$$

where

$$A(t) := 1 - 2t(z \pm \rho *), \qquad (2.7)$$

the symbol * denoting the duality operator introduced in Ref. 3. In this paper we shall simultaneously treat the cases $\mathrm{sgn}(\det h) = -1$ and $\mathrm{sgn}(\det h) = +1$. Whenever we use the notation $\pm$ or $\mp$ the upper sign will be the one appropriate for stationary axially symmetric fields. The field ρ will be defined as usual by

$$\rho := |\det h|^{1/2}, \qquad (2.8)$$

and the field z will be defined by

$$z := \tfrac{1}{2}\mathrm{Tr}[F^{(1)}\Omega]. \qquad (2.9)$$

If, as in Ref. 2, one expresses $F^{(1)}$ in the form

$$F^{(1)} = \begin{pmatrix} H & \phi \\ 2iL & 2iK \end{pmatrix}, \qquad (2.10)$$

where

$$dL = \phi^{\dagger}\epsilon\, dH, \quad dK = \phi^{\dagger}\epsilon\, d\phi, \qquad (2.11)$$

with

$$\epsilon := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad (2.12)$$

then the definition (2.9) is equivalent to the relation

$$\tfrac{1}{2}(H - H^{T}) = (K + iz)\epsilon. \qquad (2.13)$$

One should also note that the additive constants in H can and will be chosen so that

$$\mathrm{Re}H = -h - \mathrm{Re}(\phi\phi^{\dagger}) + (\mathrm{Re}K)\epsilon. \qquad (2.14)$$

Furthermore, Eqs. (2.13) and (2.14) imply that

$$\tfrac{1}{2}(H + H^{\dagger}) = -h - \phi\phi^{\dagger} + iz\epsilon. \qquad (2.15)$$

From Eqs. (2.8) and (2.13) and the "self-duality conditions"

$$-\rho^{-1}h\epsilon * dH = i\, dH, \qquad (2.16a)$$

$$-\rho^{-1}h\epsilon * d\phi = i\, d\phi, \qquad (2.16b)$$

it is readily established by the same technique that was used in Ref. 3 for the vacuum case that here again

$$*dz = -d\rho. \qquad (2.17)$$

In the present paper we shall constrain the choice of $F(t)$ a little more than we did in Ref. 2 by imposing the additional condition

$$\det F(t) = -\lambda(t)^{-1}, \qquad (2.18)$$

where

$$\lambda(t) := [(1 - 2tz)^2 \pm (2t\rho)^2]^{1/2}. \qquad (2.19)$$

Showing that it is possible to do this necessitates showing that

$$d[\lambda(t)\det F(t)] = 0. \qquad (2.20)$$

However, there is a theorem[9] that says that if

$$dF(t) = \Gamma(t)\Omega F(t),$$

then

$$d[\det F(t)] = \mathrm{Tr}[\Gamma(t)\Omega]\det F(t).$$

In our case, Eqs. (2.6) and (2.9) give

$$\begin{aligned}\mathrm{Tr}[\Gamma(t)\Omega] &= tA(t)^{-1}\mathrm{Tr}[dF^{(1)}\Omega] \\ &= 2tA(t)^{-1}\, dz = -\lambda^{-1}\, d\lambda,\end{aligned}$$

the last step following from the definitions (2.7) and (2.19). Therefore, Eq. (2.20) follows immediately.

The condition (2.18) completes the subsidiary conditions to which our $F(\mathbf{x},t)$ will be subjected. In Sec. 3 we shall show that the solution can always be selected so that the only singularities are at the zeros of $\lambda(\mathbf{x},t)$ and of $\lambda(\mathbf{x}_0,t)$, where $\mathbf{x}_0$ is a particular nonaxial point at which H is stipulated to be Hermitian and ϕ, L, and K are stipulated to vanish. This choice of $H(\mathbf{x}_0)$, $\phi(\mathbf{x}_0)$, $L(\mathbf{x}_0)$, and $K(\mathbf{x}_0)$ involves no loss of generality, for it can always be achieved by means of a gauge transformation.

3. GAUGE TRANSFORMATIONS

The residual arbitrariness of the 3×3 complex potential $F^{(1)}(\mathbf{x})$ may be described as follows:

(1) We shall usually select x^1, x^2 so that

$$g_{12} = 0, \quad g_{11} = \pm g_{22} = \exp(2\Gamma). \qquad (3.1)$$

There remain the conformal coordinate transformations which preserve this form of metric.

(2) There also remain the transformations

$$\rho \to (\exp b)\rho, \quad z \to (\exp b)z + c, \qquad (3.2)$$

where b, c are any real constants.

(3) Just as in the vacuum case treated in Ref. 3 we again have the transformation

$$H \to H + iB, \qquad (3.3)$$

where B is any 2×2 real symmetric constant matrix.

(4) Again X can be subjected to the SL(2,R) transformations

$$X \to SX, \quad \det S = 1, \quad dS = 0, \qquad (3.4)$$

which induce the mapping

$$H \to SHS^{T}. \qquad (3.5)$$

(5) A 2×1 complex constant matrix can be added to $\phi(\mathbf{x})$.

(6) A 1×2 complex constant matrix can be added to $L(\mathbf{x})$.

(7) A complex constant can be added to $K(\mathbf{x})$.

It should be noted that in practice the imaginary part of the arbitrary constant in (7) is usually chosen so that

$$K - K^{*} = \phi^{\dagger}\epsilon\phi. \qquad (3.6)$$

We shall now set about choosing $F(\mathbf{x},t)$ at one particular point $\mathbf{x}_0$ off the axis. The chosen $F(\mathbf{x}_0,t)$ must satisfy the conditions (2.3), (2.4), and (2.18). The task is simplified by employing the gauge transformations (5)–(7) to make $\phi(\mathbf{x}_0)$, $L(\mathbf{x}_0)$, and $K(\mathbf{x}_0)$ vanish, for then the problem reduces to that which we faced in the vacuum case treated in Ref. 3. Thus, we employ the gauge transformation (3) to make $H(\mathbf{x}_0)$ Hermitian, and we note that

$$\mathscr{H}(\mathbf{x}_0,t) = \begin{pmatrix} i\epsilon A(\mathbf{x}_0,t) & 0 \\ 0 & \frac{1}{2}t \end{pmatrix}, \tag{3.7}$$

where

$$A(t) := I - 2t(-h + iz\epsilon)i\epsilon. \tag{3.8}$$

Proceeding as in Ref. 3 we find that $F(\mathbf{x}_0,t)$ can be chosen so that

$$F(x_0,t) = \begin{pmatrix} A(\mathbf{x}_0,t)^{-1/2} i\epsilon w(\mathbf{x}_0,t) & 0 \\ 0 & 1 \end{pmatrix}, \tag{3.9}$$

where

$$A(t)^{-1/2} = \frac{[1 - 4tz + \lambda(t)]I + 2t(-h + iz\epsilon)i\epsilon}{\lambda(t)[2(1 - 2tz + \lambda(t))]^{1/2}}, \tag{3.10}$$

and

$$w(t) = \begin{pmatrix} 2^{1/2}[1 - 2tz + \lambda(t)]^{-1/2} & 0 \\ 0 & 2^{-1/2}[1 - 2tz + \lambda(t)]^{1/2} \end{pmatrix}. \tag{3.11}$$

Threfore, it can be seen that our choice of $F(\mathbf{x}_0,t)$ has no singularities except for a branch point of index $-1/2$ at the zeros of $\lambda(\mathbf{x}_0,t)$ and that

$$F(\mathbf{x}_0,t)\begin{pmatrix} t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is holomorphic in a neighborhood of $t = \infty$ and has an inverse there.

The solution of Eq. (2.1) may be expressed in the form

$$F(\mathbf{x},t) = \mathscr{F}(\mathbf{x},\mathbf{x}_0,t)F(\mathbf{x}_0,t), \tag{3.12}$$

where $\mathscr{F}(\mathbf{x},\mathbf{x}_0,t)$ is that particular solution of Eq. (2.1) which satisfies

$$\mathscr{F}(\mathbf{x}_0,\mathbf{x}_0,t) = I \tag{3.13}$$

for all t. Proceeding as in Ref. 3 we again find that the only singularities of $\mathscr{F}(\mathbf{x},\mathbf{x}_0,t)$ occur at the zeros of $\lambda(\mathbf{x},t)$ and $\lambda(\mathbf{x}_0,t)$. Furthermore, introducing null coordinates x_A as in Ref. 3, we can establish that

$$\frac{\partial F^{(1)T}}{\partial x_A}\mathfrak{G}(0)\frac{\partial F^{(1)}}{\partial x_A} = 0, \tag{3.14}$$

which shows immediately that

$$\det\frac{\partial F^{(1)}}{\partial x_A} = 0.$$

Thus, the product of the eigenvalues of $[\partial F^{(1)}/\partial x_A]\Omega$ vanishes. One can also show that the trace of $[\partial F^{(1)}/\partial x_A]\Omega$ and the trace of $\{[\partial F^{(1)}/\partial x_A]\Omega\}^2$ are both equal to unity. This shows that two of the eigenvalues must vanish, and one must be equal to unity. Consequently, here, as in the vacuum case, we conclude that since

$$\frac{\partial F(t)}{\partial x_A} = -\tfrac{1}{2}(r_A - \tau)^{-1}\left(\frac{\partial F^{(1)}}{\partial x_A}\Omega\right)F(t) \tag{3.15}$$

(using the same notation as in Ref. 3), the branch points at $r_A = \tau$ will be of index $-\frac{1}{2}$ while those at $r_{0A} = \tau$ will be of index $\frac{1}{2}$.

Having concluded these preliminaries, we are now ready to discuss the generalization of the HHP to electrovac fields.

4. THE HOMOGENEOUS HILBERT PROBLEM

As in Ref. 3, we select an arbitrary compact region U_c in the domain of $F_0^{(1)}$. Then there exists at least one smooth contour L about the origin in the complex t-plane such that $F_0(t)$ is holomorphic on $L + L_+$. Choose any L in this category, and let $u(t)$ be any member of K'_L. Then define

$$G(t) := F_0(t)u(t)F_0(t)^{-1}. \tag{4.1}$$

The HHP consists of finding functions $X_-(t)$ and $X_+(t)$, holomorphic in $L + L_-$ and $L + L_+$, respectively, where L_- is the complement of $L + L_+$, such that for all s on L

$$X_-(s) = X_+(s)G(s), \quad X_+(0) = I. \tag{4.2}$$

[We accept the same working hypotheses as in the vacuum case, e.g., that the component indices of the HHP for $G(s)$ vanish.]

Once one has effected the solution of the HHP for $G(s)$, one defines

$$F^{(1)} := F_0^{(1)} + \dot{X}_+(0)\Omega, \tag{4.3a}$$

$$\mathscr{H}(t) := i\mathfrak{G}(t) - it[\mathfrak{G}(0)F^{(1)}\Omega + \Omega F^{(1)\dagger}\mathfrak{G}(0)], \tag{4.3b}$$

$$\Gamma(t) := t\Lambda(t)^{-1}dF^{(1)}. \tag{4.3c}$$

Proceeding exactly as in the vacuum case, we can show that in the electrovac case

$$\det X(t) = 1, \tag{4.4a}$$

$$X(t)^\dagger\mathscr{H}(t)X(t) = \mathscr{H}_0(t), \tag{4.4b}$$

$$\det\mathscr{H}(t) = \det\mathscr{H}_0(t) = -\tfrac{1}{2}t\lambda(t)^2, \tag{4.4c}$$

$$dX(t) = \Gamma(t)\Omega X(t) - X(t)\Gamma_0(t)\Omega, \tag{4.4d}$$

and

$$\mathscr{H}(t)\,dX(t) + it[X(t)^\dagger]^{-1}\mathfrak{G}(0)\,dF_0^{(1)}\,\Omega = it\mathfrak{G}(0)dF^{(1)}\,\Omega X(t), \tag{4.4e}$$

where $X(t)$ is the sectionally holomorphic function

$$X(t) := \begin{cases} X_+(t) & \text{if } t \text{ is in } L + L_+, \\ X_-(t) & \text{if } t \text{ is in } L + L_-. \end{cases} \tag{4.5}$$

Using Eqs. (4.4a)–(4.4e) one can then establish that the potential defined by

$$F(t) := X_+(t)F_0(t) \tag{4.6}$$

satisfies

$$dF(t) = \Gamma(t)\Omega F(t), \tag{4.7a}$$

$$\mathscr{H}(t)\Gamma(t) = it\ \mathfrak{G}(0)\,dF^{(1)}, \tag{4.7b}$$

$$F(0) = \Omega, \quad \dot{F}(0) = F^{(1)}, \tag{4.7c}$$

$$\det F(t) = -\lambda(t)^{-1}, \tag{4.7d}$$

and

$$F(t)^\dagger\mathscr{H}(t)F(t) = i\mathfrak{G}(t). \tag{4.7e}$$

Thus, $F(t)$ satisfies all of the defining equations for the F-potential corresponding to a given H and ϕ.

If one expresses $F^{(1)}$ in the usual form

$$F^{(1)} = \begin{pmatrix} H & \phi \\ 2iL & 2iK \end{pmatrix}, \tag{4.8}$$

then from Eq. (4.7b) one concludes that L and K satisfy

$$dL = \phi^\dagger\epsilon\,dH, \quad dK = \phi^\dagger\epsilon\,d\phi, \tag{4.9}$$

and that H and ϕ satisfy the self-duality conditions

$$h\epsilon\, dH = i\rho{*}dH, \quad h\epsilon d\phi = i\rho{*}d\phi, \tag{4.10}$$

where

$$h := -\operatorname{Re}[H + \phi\phi^\dagger - K\epsilon]. \tag{4.11}$$

From Eq. (4.4a) it follows that

$$\operatorname{Tr}\dot{X}_+(0) = 0, \tag{4.12}$$

and, therefore, that

$$\operatorname{Tr}(F^{(1)}\Omega) = \operatorname{Tr}(F_0^{(1)}\Omega) = 2z. \tag{4.13}$$

From Eqs. (4.11) and (4.13) we then conclude that

$$h^T = h. \tag{4.14}$$

Finally, to establish that $h\epsilon h = -\rho^2\epsilon$ or

$$|\det h| = \rho, \tag{4.15}$$

we appeal to Eq. (4.4c), using the definition (4.3b) and writing $F^{(1)}$ in the form (4.8). The 3×3 determinant is easily reduced to a 2×2 determinant

$$\det A(t) = \lambda(t)^2, \tag{4.16}$$

where

$$A(t) := (1 - 2tz)I + 2ith\epsilon. \tag{4.17}$$

One then employs the equation

$$A(t)^T\epsilon A(t) = \epsilon \det A(t), \tag{4.18}$$

as in the vacuum case, to establish (4.15).

Eqs. (4.9), (4.10), (4.13), (4.14), and (4.15) show that $F^{(1)}$ fulfills the definition of an $F^{(1)}$-potential for a spacetime in V'.

5. DERIVATION OF THE ELECTROVAC INTEGRAL EQUATION

One method of solving the HHP for $G(s)$ involves solving the integral equation which we described in Ref. 2. The derivation of that integral equation is trivial, for it is an immediate consequence of the HHP that we have formulated in this paper.

Since $X_-(t)$ is holomorphic in $L + L_-$ (including $t = \infty$ according to our working hypothesis), we obviously have for t inside L

$$\frac{1}{2\pi i}\int_L ds\,\frac{X_-(s)}{s(s-t)} = 0. \tag{5.1}$$

However, by Eq. (4.2) this can be expressed in the form

$$\frac{1}{2\pi i}\int_L ds\,\frac{X_+(s)G(s)}{s(s-t)} = 0. \tag{5.2}$$

Hence, writing

$$X_+(t) = I + tf(t) \tag{5.3}$$

and

$$G(t) = I + K(t), \tag{5.4}$$

where $f(t)$ is holomorphic in $L + L_+$, we have

$$\frac{1}{2\pi i}\int_L ds\,\frac{[I + sf(s)][I + K(s)]}{s(s-t)} = 0. \tag{5.5}$$

This is easily seen to be equivalent to

$$f(t) + \frac{1}{2\pi i}\int_L ds\,\frac{[f(s) + s^{-1}I]K(s)}{s-t} = 0, \tag{5.6}$$

the integral equation which we described in Ref. 2.

6. PERSPECTIVES

In this section we should like to mention several areas where we feel that more work will be required in order to keep the development of the electrovac theory abreast of that of the vacuum theory.

Recently we succeeded in showing[10] that the vacuum-to-vacuum KC transformations corresponding to $u(t)$ of the form

$$u^{(1)}(t) = \begin{pmatrix} 1 & t\alpha(t) \\ 0 & 1 \end{pmatrix} \tag{6.1}$$

or

$$u^{(2)}(t) = \begin{pmatrix} 1 & 0 \\ -t^{-1}\beta(t) & 1 \end{pmatrix}, \tag{6.2}$$

where $\alpha(t)$ and $\beta(t)$ are real[11] functions holomorphic on L, can be effected by solving a (regular) Fredholm integral equation of the second kind. For example, for $u^{(2)}(t)$ the integral equation assumes the form

$$F_{44}(t) + \frac{1}{2\pi i}\int_L ds\,\frac{tF_{44}(s)\beta(s)U(s,t)}{s^2(s-t)} = F_{44}^{\text{seed}}(t). \tag{6.3}$$

Here $F_{44}(t)$ is the lower right element of $F(t)$, and

$$U(s,t) := [\epsilon F^{\text{seed}}(s)^{-1}F^{\text{seed}}(t)]_{44}. \tag{6.4}$$

The regularity of the integral equation follows from the fact that $U(s,s) = 0$ for all s.

Once Eq. (6.3) is solved for $F_{44}(t)$, the complex Ernst potential

$$\mathscr{E} = \dot{F}_{44}(0) \tag{6.5}$$

may be evaluated immediately. The complete metric can be obtained from $\mathscr{E}$ by using well-known methods. Alternatively one can calculate $h = -\operatorname{Re}H$, for there exist equations resembling Eq. (6.3) which may be used to determine the full matrix potential $F(t)$, and from this it is trivial to obtain $H = \dot{F}(0)$.

Since any vacuum-to-vacuum $u(t)$ can be factored into a small number of factors[3] of the $u^{(1)}$ and $u^{(2)}$ type, one can concentrate upon the analysis of these nonmatrix Fredholm equations, thus avoiding the need to analyze directly the more complicated matrix integral Eq. (5.6). The problem of finding an equally fruitful factorization of the electrovac-to-electrovac $u(t)$ deserves attention.

In the vacuum case we have also found[10] that if a given stationary axially symmetric solution can be obtained from some static Weyl metric using a single transformation of type $u^{(1)}$ or from some static Weyl metric using a single transformation of type $u^{(2)}$, both the static Weyl metric and the matrix $u(t)$ are completely determined by the value on the axis of the complex Ernst potential $\mathscr{E}$ of the given stationary axially symmetric field. For example, in the case of $u^{(1)}$ one has the relation

$$\mathscr{E}(z) = e^{2\psi(z)} + i\alpha(1/2z), \tag{6.6}$$

where $\psi(z)$ is the t-independent Weyl potential evaluated on the axis $\rho = 0$, and $\alpha(1/2z)$ is the function $\alpha(t)$ in Eq. (6.1) evaluated at $t = 1/2z$.

One wonders whether or not in the case of a stationary axially symmetric electrovac solution, knowledge of the two complex Ernst potentials $\mathscr{E}$ and Φ on the axis will determine a static Weyl metric and some transformation $u(t)$ which may similarly be effected by solving a (regular) Fredholm integral equation of the second kind. In any event we consider this to be a worthwhile subject for study.

Finally, among electrovac solutions the Plebański–Demiański Petrov type D solution[12] would be a particularly interesting solution to subject to KC transformations, both because of the intrinsic elegance of that solution and because it encompasses so many physically important solutions. At the present time several of our students are calculating the potential $F(t)$ for the Plebański–Demiański solution. With this potential in hand, we expect to be able to apply the present solution-generating theory to derive new and potentially interesting solutions of the Einstein–Maxwell field equations. Study of the Plebański–Demiański solution may also provide some hint concerning how to extend the theory to solutions with nonvanishing cosmological constant.

[1]I. Hauser and F.J. Ernst, Phys. Rev. D **20**, 362 (1979). [There is a sign error in the second term in Eq. (3.2).]

[2]I. Hauser and F.J. Ernst, Phys. Rev. D **20**, 1783 (1979). [There is the same sign error in the second term in Eq. (3.3). We should also like to point out that one can easily generalize the calculation performed in Sec. IV to include not only poles on the real axis in Eq. (4.1) but also pairs of poles at complex conjugate points.]

[3]I. Hauser and F.J. Ernst, J. Math. Phys. **21**, 1126 (1980).

[4]W. Kinnersley and D.M. Chitre, J. Math. Phys. **18**, 1538 (1977); **19**, 1926 (1978); **19**, 2037 (1978).

[5]R. Geroch, J. Math. Phys. **12**, 918 (1971); **13**, 394 (1972).

[6]W. Kinnersley, J. Math. Phys. **14**, 651 (1973); **18**, 1529 (1977).

[7]Our notations and conventions concerning differential forms, Grassmann products, duality operations, etc., are described in an appendix to the paper by I. Hauser and F.J. Ernst, J. Math. Phys. **19**, 1316 (1978).

[8]We stress that we do not advocate always using the particular gauge introduced here and in Ref. 3.

[9]This theorem can be found, for example, in S. Lefschetz, *Differential Equations: Geometric Theory*, 2nd ed. (Dover, New York, 1977), p. 60.

[10]A manuscript concerning these developments in the vacuum theory is now being prepared.

[11]By "real" we mean that $\alpha(t)^* = \alpha(t^*)$ and $\beta(t)^* = \beta(t^*)$.

[12]J.F. Plebański and M. Demiański, Ann. Phys. (N.Y.) **98**, 98 (1976). (Until we learn how to cope with a nonvanishing cosmological constant we shall have to set $\lambda = 0$).

Space–times with distribution valued curvature tensors [a)]

A. H. Taub
Mathematics Department, University of California, Berkeley, California 94720

(Received 27 August 1979; accepted for publication 26 November 1979)

A space–time in which in an admissible coordinate system the metric tensor is continuous but has a finite jump in its first and second derivatives across a submanifold will have a curvature tensor containing a Dirac delta function. The support of this distribution may be of three, two, or one dimension or may even consist of a single event. Lichnerowicz's formalism for dealing with such tensors is modified so as to obtain a formalism in which the Bianchi identities are satisfied in the sense of distributions. The resulting formalism is then applied to the discussion of the Einstein field equations for problems in which the source of the gravitational field is given by a distribution valued stress-energy tensor. Gravitational shocks are also discussed and their theory is compared with that of high-frequency gravitational waves given by Y. Choquet-Bruhat. By considering a class of line sources as obtainable from cylindrical shells by a limiting process, as was proposed by Israel, one may use the distribution formalism developed for hypersurfaces to treat line sources. The line source model proposed by Israel to represent the Kerr metric in the neighborhood of its singular disk is shown to lead to a gravitational mass and angular momentum inconsistent with those of the latter metric. It is proposed to remove this difficulty by changing the assumptions made by Israel concerning the nature of the space–time inside the cylindrical shell which is the support of the distribution in the curvature tensor. The details of the effect of this change are not given in this paper.

1. INTRODUCTION

Although many physical systems undergo very rapid transitions of their state of motion, their states are not likely to be discontinuous functions of events in space–time or to have discontinuities in their low order derivatives. Still there are many examples in which a mathematical description of the system which is based on distribution valued states of the system give an accurate picture of some important aspects of the physical problem and such a description is more amenable to treatment than is the treatment which contains a smooth description of the physical states.

A very illuminating example occurs in classical hydrodynamics. If heat conduction and viscosity are ignored, then that theory involves shock waves, i.e., one must deal with weak solutions of the equations describing the conservation of mass, momentum, and energy. Such solutions describe transitions in the state of the medium in which entropy is changed (increased). The mathematical description of the transition is made by introducing a discontinuity in the variables describing the motion of the fluid. When heat conduction and viscosity are taken into account it is found that the entropy of the fluid changes by the amount given by the shock wave theory and that a rapid but continuous change in state takes place. Thus, gross features of the transition are given correctly by the simplified theory involving shock waves but some fine features are not treated at all; in particular, the structure of the transition region is ignored in such a theory.

One should expect that general relativistic hydrodynamics would be similar to classical hydrodynamics and the former's mathematical description by the use of distribution valued quantities would contain an accurate description of some aspects of the behavior of self-gravitating fluids. Thus, one is led to consider sources of gravitational fields which are described by distribution valued stress-energy tensors. In view of the Einstein field equations, this means that we should deal with space–times whose Ricci tensors are distribution valued. Since the conformal tensor (the Weyl tensor) is related to the Ricci tensor by means of the Bianchi identities one must expect that the entire Riemann–Christoffel curvature tensor should be distribution valued.

Such a curvature tensor will arise if one deals with a space–time in which in an admissible coordinate system the metric tensor is continuous but has discontinuous first and second derivatives due to finite jumps which occur across submanifolds of three, two, or one dimension or even at a single event. The theory of a gravitational shock wave is one in which there is a hypersurface Σ across which finite discontinuities in the first and second derivatives of the metric tensor occur. These are described by two second order tensors $b_{\mu\nu}$ and $\hat{b}_{\mu\nu}$ defined on Σ. These quantities may be shown to satisfy algebraic equations involving the vector l_μ, the normal to Σ, and differential equations describing their propagation in the direction of the normal to Σ. Thus, one is able to characterize the "singular" hypersurface Σ and describe its development in time. In this case one considers space–time to include a region containing a submanifold on which the curvature tensor has a delta function behavior and studies the behavior of the region. One would expect to treat the other space–time with distribution valued curvature tensors in a similar manner and thus be able to discuss the behavior of shock waves, shells of matter, the history of line sources, singular world lines, and singular events.

We shall mainly be concerned with space–times whose curvature tensors contain Dirac delta functions with sup-

[a)]This paper is based on lectures given at the College de France, Paris, France, during the period October 4–18, 1977.

 0022-2488/80/061423-09$1.00

ports on submanifolds or even isolated events. When one recalls that the curvature tensor is linear in the second derivatives of the metric tensor and quadratic in the first derivatives, one sees that for a space–time in which the first and second derivatives of the metric tensor have a finite jump across a submanifold, then its curvature tensor will contain a Dirac delta function with support on the submanifold. The jump in the first derivative is describable by a Heaviside function which will enter the curvature tensor quadratically. Fortunately, the product of such distributions is quite tractable.

Lichnerowicz[1] has given a discussion of hydrodynamic and gravitational shock wave problems by using curvature tensors for space–times of the type described in the preceding paragraph. In the sequel we shall apply his formalism with a slight modification. The purpose of this modification is to have a formalism in which the Bianchi identities are satisfied in the sense of distributions by the curvature tensor. These identities and the Stokes' theorems will be the main tools we shall use in analyzing various problems.

We shall also compare the derivation of the equations satisfied by $b_{\mu\nu}$ and l_μ with the derivation of the similar equations that occur in the treatment of high-frequency gravitational waves given by Y. Choquet-Bruhat.[2] In the latter theory the role of Σ is played by the hypersurface of constant phase and the role of $b_{\mu\nu}$ by the slowly varying amplitude of the gravitational wave. The computational origin of the equations in both theories is very similar and supports what one would expect on intuitive grounds, namely that gravitational shocks, especially weak ones, should behave much the same as continuous solutions of the Einstein fields equations.

In addition, the paper will contain an application of the distribution valued curvature formalism to the theory of shells and line sources in general relativity.

We shall begin our discussion with Lichnerowicz's formalism for the case in which there is a hypersurface Σ across which the metric tensor has a discontinuous derivative. Let Σ be described by the equation

$$\varphi(x) = 0 \tag{1-1}$$

and have the normal vector

$$l_\mu = \varphi,_\mu = \frac{\partial \varphi}{\partial x^\mu}. \tag{1-2}$$

We shall assume that the hypersurface Σ divides a region Ω of space–time into two parts Ω^+ and Ω^- where $\varphi > 0$ and $\varphi < 0$, respectively. We denote by

$$[f] = (f+) - (f-), \tag{1-3}$$

where $f+(f-)$ is the limit of the function f in $\Omega^+(\Omega^-)$ as the point in Ω^+ approaches Σ. The tensor $g_{\mu\nu}$ will be assumed to be continuous across Σ. The finite discontinuities in the first and second derivatives of the metric tensor are then given by

$$[g_{\mu u,\sigma}] = l_\sigma b_{\mu u}, \tag{1-4}$$

$$[g_{\mu\nu,\sigma\tau}] = l_{\sigma,\tau} b_{\mu\nu} + l_\sigma b_{\mu\nu,\tau} + l_\tau b_{\mu\nu,\sigma} + l_\sigma l_\tau \hat{b}_{\mu\nu}. \tag{1-5}$$

The latter equations may be derived by noting that in the neighborhood of Σ we may write

$$g^{\pm}_{\mu\nu} = g^0_{\mu\nu} + \varphi g'^{\pm}_{\mu\nu} + \frac{\varphi^2}{2} g''^{\pm}_{\mu\nu} + \cdots.$$

It follows from Eq. (1-4) that

$$2[\Gamma^\alpha_{\beta\gamma}] = l_\beta b^\alpha_\gamma + l_\gamma b^\alpha_\beta - l^\alpha b_{\beta\gamma}, \tag{1-6}$$

where $\Gamma^\alpha_{\beta\gamma}$ are the Christoffel symbols computed from the $g_{\mu\nu}$. Since the Riemann–Christoffel curvature tensor is given by

$$2R_{\alpha\beta\mu\nu} = g_{\alpha\nu,\beta\mu} + g_{\beta\mu,\alpha\nu} - g_{\beta\nu,\alpha\mu} - g_{\alpha\mu,\beta\nu} - 2g_{\rho\sigma}\Gamma^\rho_{\alpha\mu}\Gamma^\sigma_{\beta\nu} + 2g_{\rho\sigma}\Gamma^\rho_{\alpha\nu}\Gamma^\sigma_{\beta\mu}$$

and since

$$g_{\rho\sigma}[\Gamma^\rho_{\alpha\mu}\Gamma^\sigma_{\beta\nu}] = g_{\rho\sigma}\{\bar{\Gamma}^\rho_{\alpha\mu}[\Gamma^\sigma_{\beta\nu}] + [\Gamma^\rho_{\alpha\mu}]\bar{\Gamma}^\sigma_{\beta\nu}\}, \tag{1-7}$$

where

$$\bar{\Gamma}^\rho_{\alpha\mu} = \tfrac{1}{2}\{\Gamma^+_{\rho\alpha\mu} + \Gamma^-_{\rho\alpha\mu}\}, \tag{1-8}$$

we may write

$$[R^{\ \rho}_{\sigma\ \mu\nu}] = [\Gamma^\rho_{\sigma\nu}]_{:\mu} - [\Gamma^\rho_{\sigma\mu}]_{:\nu} + l_\mu A^{\ \rho}_{\sigma\nu} - l_\nu A^{\ \rho}_{\sigma\mu}, \tag{1-9}$$

with

$$2A^{\ \rho}_{\sigma\mu} = b^\rho_{\sigma:\mu} + b^\rho_{\sigma:\mu} - b^{\ :\rho}_{\sigma\mu} + l_\mu \tilde{b}^\rho_\sigma + l_\sigma \tilde{b}^\rho_\mu - l^\rho \tilde{b}_{\sigma\mu} + c^\rho_\tau l^\tau b_{\sigma\mu}. \tag{1-10}$$

The colon denotes the covariant derivative with respect to the connection $\bar{\Gamma}$, and

$$\tilde{b}^\mu_\nu = \hat{b}^\mu_\nu + 2c^\mu_\rho b^\rho_\nu,$$

$$g_{\mu\nu:\tau} = c_{\mu\nu} l_\tau = \frac{\partial g^0_{\mu\nu}}{\partial x^\tau} - \frac{1}{2}\left\{\left(\frac{\partial g_{\mu\nu}}{\partial x^\tau}\right)^+ + \left(\frac{\partial g_{\mu\nu}}{\partial x^\tau}\right)^-\right\},$$

$$g^0_{\mu\nu} = g_{\mu\nu}\big|_{\varphi=0}. \tag{1-11}$$

It then follows that

$$[R_{\sigma\mu}] = [R^{\ \rho}_{\sigma\ \mu\rho}] = [\Gamma^\rho_{\sigma\rho}]_{:\mu} - [\Gamma^\rho_{\sigma\mu}]_{:\rho} + l_\mu A^{\ \rho}_{\sigma\rho} - l_\rho A^{\ \rho}_{\sigma\mu} \tag{1-12}$$

and that

$$[R] = (bl^\mu)_{:\mu} + l^\mu b_{:\mu} + l^\tau l_\tau(\hat{b} + 2b_{\beta\mu}c^{\beta\mu}) - (b^\mu_\nu)_{:\mu} - l^\mu b^{\ \nu}_{\mu:\nu} + bc^{\alpha\nu}l_\alpha l_\nu - l^\mu l_\nu(\hat{b}^\nu_\mu + 3c^\nu_\rho b^\rho_\mu), \tag{1-13}$$

$$b = b^\sigma_\sigma, \quad \hat{b} = \hat{b}^\sigma_\sigma.$$

Therefore,

$$2[G_{\mu\nu}] = (l^\rho b'_{\mu\nu})_{:\rho} + l^\rho b'_{\mu\nu:\rho} - (b'^{\rho}_{\mu} l_\rho)_{:\nu} - (b'^{\rho}_{\nu} l_\rho)_{:\mu} - l_\mu b'^{\rho}_{\nu:\rho} - l_\nu b'^{\rho}_{\mu:\rho} - l_\mu l_\rho \tilde{b}'^{\rho}_{\nu} - l_\nu l_\rho \tilde{b}'^{\rho}_{\mu} + l_\tau l^\tau \tilde{b}'_{\mu\nu} + g_{\mu\nu}\tilde{b}'_{\rho\sigma} l^\rho l^\sigma + g_{\mu\nu}((b'^{\rho}_{\sigma} l^\sigma)_{:\rho} + l^\sigma b'^{\rho}_{\sigma:\rho}) + b'_{\mu\nu} c_{\rho\sigma} l^\rho l^\sigma - 2l_\tau l^\tau c_{\mu\rho} b'^{\rho}_{\nu} + g_{\mu\nu} l^\sigma l_\rho c^\rho_\tau b'^{\tau}_{\sigma},$$

where

$$\tilde{b}'^{\rho}_{\sigma} = \tilde{b}^\rho_\sigma - \tfrac{1}{2}\tilde{b}\delta^\rho_\sigma, \quad b'^{\rho}_{\sigma} = b^\rho_\sigma - \tfrac{1}{2}b\delta^\rho_\sigma.$$

2. DISTRIBUTION VALUED CURVATURE TENSORS

Lichnerowicz defines in Ref. 1 a p-tensor distribution as follows: If T and U are two p tensors, one denotes by (T,U)

their scalar product at a point x of the space–time V_4. Let $\mathscr{D}^p(V_4)$ be the space p tensors with compact support and of a given differentiability class. If U is in $\mathscr{D}^p(V_4)$, one defines for all locally summable p tensors T

$$\langle T, U\rangle = \int_{V_4} (T, U)\sqrt{-g}\, d^4x\,. \qquad (2\text{-}1)$$

A p-tensor distribution T is a linear functional on $\mathscr{D}^p(V_4)$ defined by Eq. (2-1). One usually writes $\langle T,U\rangle = T[U]$, indicating the bilinearity of $T[U]$.

We shall depart slightly from the approach used by Lichnerowicz and introduce the function θ which has the value 1 on Ω^+, 1/2 on Σ, and 0 on Ω^-, i.e., $\theta = \theta(\varphi)$ and

$$\theta(\varphi) = \begin{cases} 1, & \varphi > 0, \\ \frac{1}{2}, & \varphi = 0, \\ 0, & \varphi < 0. \end{cases} \qquad (2\text{-}2)$$

Further, we have

$$\theta_{,\mu} = \frac{\partial\theta}{\partial x^\mu} = l_\mu \bar\delta\,, \qquad (2\text{-}3)$$

where $\bar\delta$ is the Dirac delta function distribution with support on Σ. Thus, for an arbitrary function f of compact support

$$\int_\Omega \bar\delta f\sqrt{-g}\, d^4x = \int_{\partial\Omega^-} f d_3 v = -\int_{\partial\Omega^+} f d_3 v\,, \qquad (2\text{-}4)$$

where d_3v is the invariant volume element induced on the hypersurface Σ.

If T is a vector field defined in Ω and is suitably smooth (say C^3) and if T and its derivatives have finite discontinuities across Σ, we may define distributions in terms of them as follows:

$$(T^\mu)^D = \theta T^{+\mu} + (1-\theta)T^{-\mu}\,,$$
$$(T^\mu_{;\nu})^D = \theta(T^\mu_{;\nu})^+ + (1-\theta)(T^\mu_{;\nu})^-\,, \qquad (2\text{-}5)$$

where the superscripts $\pm$ on a tensor field restrict that tensor field to the regions $\Omega^\pm$, respectively. Further,

$$(T^\mu_{;\nu})^\pm = (T^\mu_{,\nu} + T^\rho\Gamma^\mu_{\rho\nu})^\pm\,. \qquad (2\text{-}6)$$

Thus, for such tensor fields T, $T = T^+$ in Ω^+, $T = T^-$ in Ω_-, and $T = \bar T = \frac{1}{2}(T_+ + T_-)$ on Σ.

We observe that

$$[T^\rho]_{,\mu} - [T^\rho_{,\mu}] = \tilde T^\rho l_\nu\,, \qquad (2\text{-}7)$$

where $\tilde T^0$ is a vector field on Σ.

For distribution valued vector and tensor fields we define covariant differentiation by formulas analogous to

$$(T^{\mu D})_{;\nu} = (T^{\mu D})_{,\nu} + T^{\rho D}\Gamma^\mu_{\rho\nu}\,, \qquad (2\text{-}8)$$

where

$$\Gamma^\mu_{\rho\nu} = \theta\Gamma^{+\mu}_{\rho\nu} + (1-\theta)\Gamma^{-\mu}_{\rho\nu}\,. \qquad (2\text{-}9)$$

Equation (2-8) may be written as

$$(T^{\mu D})_{;\nu} = \bar\delta l_\nu[T^\mu] + (T^\mu_{;\nu})^\nu - \theta(1-\theta)\{[T^\rho][\Gamma^\mu_{\rho\nu}]\}\,. \qquad (2\text{-}10)$$

Applying the usual rules of covariant differentiation and using Eq. (2-7) we find that

$$(T^{\rho D})_{;\mu\nu} - (T^{\rho D})_{;\nu\mu} = -T^{\sigma D}Q^\rho_{\sigma\mu\nu}\,,$$

where

$$Q^\rho_{\sigma\mu\nu} = \Gamma^\rho_{\sigma\nu,\mu} - \Gamma^\rho_{\sigma\mu,\nu} + \Gamma^\tau_{\sigma\nu}\Gamma^\rho_{\tau\mu} - \Gamma^\tau_{\sigma\mu}\Gamma^\rho_{\tau\nu}\,,$$

i.e.,

$$Q^\rho_{\sigma\mu\nu} = \bar\delta H^\rho_{\sigma\mu\nu} + (R^\rho_{\sigma\mu\nu})^D - \theta(1-\theta)J^\rho_{\sigma\mu\nu}\,, \qquad (2\text{-}11)$$

with

$$2H^\rho_{\sigma\mu\nu} = b^\rho_\nu l_\sigma l_\mu - b^\rho_\mu l_\sigma l_\nu - b_{\sigma\nu}l^\rho l_\mu + b_{\sigma\mu}l^\rho l_\nu\,, \qquad (2\text{-}12)$$

$$J^\rho_{\sigma\mu\nu} = [\Gamma^\tau_{\sigma\nu}][\Gamma^\rho_{\tau\mu}] - [\Gamma^\tau_{\sigma\mu}][\Gamma^\rho_{\tau\nu}]\,. \qquad (2\text{-}13)$$

Hence,

$$Q_{\sigma\mu} = Q^\rho_{\sigma\mu\rho} = \bar\delta H_{\sigma\mu} + (R_{\sigma\mu})^D - \theta(1-\theta)J_{\sigma\mu}\,,$$

with

$$2H_{\sigma\mu} = l^\rho l_\rho b_{\sigma\mu} - l^\rho b'_{\sigma\rho}l_\mu - l^\rho b'_{\rho\mu}l_\sigma\,, \qquad (2\text{-}14)$$

$$J_{\sigma\mu} = [\Gamma^\tau_{\sigma\rho}][\Gamma^\rho_{\tau\mu}] - [\Gamma^\tau_{\sigma\mu}][\Gamma^\rho_{\tau\rho}]\,. \qquad (2\text{-}15)$$

In addition

$$2H = 2g^{\sigma\mu}H_{\sigma\mu} = l^\rho l_\rho b - 2\,l^\rho l^\sigma b'_{\rho\sigma}\,, \qquad (2\text{-}16)$$

$$J = g^{\sigma\mu}J_{\sigma\mu}\,,$$

$$\begin{aligned} 2(H_{\sigma\mu} - \tfrac{1}{2}g_{\sigma\mu}H) &= l^\rho l_\rho b'_{\sigma\mu} - l^\rho b'_{\rho\sigma}l_\mu \\ &\quad - l^\rho b'_{\rho\mu}l_\sigma + g_{\sigma\mu}b'_{\sigma\tau}l^\sigma l^\tau \\ &= 2H'_{\sigma\mu}\,. \end{aligned} \qquad (2\text{-}17)$$

3. THE BIANCHI IDENTITIES

If the metric is C^3, these identities may be written as

$$\tfrac{1}{2}\delta^{\alpha\beta\gamma}_{\lambda\mu\nu}R_{\rho\sigma\alpha\beta;\gamma} = R_{\rho\sigma\lambda\mu;\nu} + R_{\rho\sigma\mu\nu;\lambda} + R_{\rho\sigma\nu\lambda;\mu} = 0\,,$$

where $\delta^{\alpha\beta\gamma}_{\lambda\mu\nu}$ is the generalized Kronnecker delta. A generalization of these identities to distribution valued curvature tensors is

$$\tfrac{1}{2}\delta^{\alpha\beta\gamma}_{\lambda\mu\nu}Q_{\rho\sigma\alpha\beta;\gamma} = \theta(1-\theta)\mathscr{A}_{\rho\sigma\lambda\mu\nu}\,, \qquad (3\text{-}1)$$

where $\mathscr{A}_{\rho\sigma\lambda\mu\nu}$ is a tensor defined on Σ. Thus, the right-hand side of this equation vanishes everywhere except on Σ and

$$\langle T^{\sigma\lambda\mu\nu}_\rho, \tfrac{1}{2}\delta^{\alpha\beta\gamma}_{\lambda\mu\nu}Q^\rho_{\sigma\alpha\beta;\gamma}\rangle = 0 \qquad (3\text{-}2)$$

for arbitrary tensors $T^{\sigma\lambda\mu\nu}_\rho$ in $D^5(V_4)$.

Since

$$(Q^{\delta\gamma} - \tfrac{1}{2}g^{\delta\gamma}Q)_{;\gamma} = \tfrac{1}{4}(g^{\rho\beta}g^{\sigma\gamma}g^{\alpha\delta}\delta^{\lambda\mu\nu}_{\alpha\beta\gamma}Q_{\rho\sigma\mu\nu})_{;\lambda}$$

and since Eqs. (2-12) imply that

$$H_{\rho\sigma\mu\nu}l_\lambda\delta^{\mu\nu\lambda}_{\alpha\beta\gamma} = 0\,,$$

it may be shown that Eqs. (3-1) and (2-12) in turn imply that

$$(Q^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}Q)_{;\nu} = \theta(1-\theta)d^\mu\,, \qquad (3\text{-}3)$$

where d^μ is a vector defined on Σ. Thus, Eq. (3-2) imply that

$$\langle T_\mu, (Q^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}Q)_{;\nu}\rangle = 0 \qquad (3\text{-}4)$$

for arbitrary vectors in $D^1(V_4)$.

It is a consequence of Eq. (2-11) and the rules for covariant differentiation of distributions that

$$\begin{aligned} Q^\rho_{\sigma\mu\nu;\lambda} = {} & \bar\delta' l_\lambda H^\rho_{\sigma\mu\nu} + \bar\delta(H^\rho_{\sigma\mu\nu;\lambda} + l_\lambda[R^\rho_{\sigma\mu\nu}]) + (R^\rho_{\sigma\mu\nu;\lambda})^D \\ & - \theta(1-\theta)\{J^\rho_{\sigma\mu\nu;\lambda} + [R^\tau_{\sigma\mu\nu}][\Gamma^\rho_{\tau\lambda}] - [R^\rho_{\tau\mu\nu}] \\ & \times[\Gamma^\tau_{\sigma\nu}] - [R^\rho_{\sigma\tau\nu}][\Gamma^\tau_{\mu\lambda}] - [R^\rho_{\sigma\mu\tau}][\Gamma^\tau_{\nu\lambda}]\}. \end{aligned} \qquad (3\text{-}5)$$

It may be verified from Eqs. (2-12) and (1-9) and the fact that the Bianchi identities are in Ω^+ and Ω^- that

$$\begin{aligned} \epsilon^{\alpha\mu\nu\lambda}Q^\rho_{\sigma\mu\nu;\lambda} = {} & -\theta(1-\theta)\epsilon^{\alpha\mu\nu\lambda}\{J^\rho_{\sigma\mu\nu;\lambda} + [R^\tau_{\sigma\mu\nu}][\Gamma^\rho_{\tau\lambda}] \\ & - [R^\rho_{\tau\mu\nu}][\Gamma^\tau_{\sigma\lambda}]\}\,. \end{aligned}$$

It follows from this equation and Eqs. (2-13) and (1-9) that

$$\epsilon^{\alpha\mu\nu\lambda}Q^{\rho}_{\ \sigma\mu\nu;\lambda} = -2\theta(1-\theta)\,l_\mu\epsilon^{\alpha\mu\nu\lambda} \times\{A^{\ \tau}_{\sigma\nu}[\Gamma^{\rho}_{\tau\lambda}] - A^{\ \rho}_{\tau\nu}[\Gamma^{\tau}_{\sigma\lambda}]\}.$$

Thus, Eqs. (3-1) hold with

$$\mathscr{A}_{\rho\sigma\lambda\mu\nu} = -2\delta^{\alpha\delta\gamma}_{\lambda\mu\nu}l_\alpha \times\{A^{\ \tau}_{\sigma\delta}[\Gamma^{\beta}_{\tau\gamma}] - A^{\ \beta}_{\tau\gamma}[\Gamma^{\tau}_{\sigma\gamma}]\}g_{\beta\rho}.$$

4. GRAVITATIONAL SHOCK WAVES

We shall generalize the Einstein field equations by assuming that they involve distribution valued curvature and stress-energy tensors. Thus, we assume that the field equations are

$$Q_{\beta\lambda} - \frac{1}{2}g_{\beta\lambda}Q = \bar{\delta}\Big(H_{\beta\lambda} - \frac{H}{2}g_{\beta\lambda}\Big) + G^{D}_{\beta\lambda} - \theta(1-\theta)\Big(J_{\beta\lambda} - \frac{J}{2}g_{\beta\lambda}\Big) = -\kappa(\bar{\delta}\tau_{\beta\lambda} + T^{D}_{\beta\lambda} - \theta(1-\theta)\mathscr{T}_{\beta\sigma}), \tag{4-1}$$

where $T^{+}_{\beta\lambda}(T^{-}_{\beta\lambda})$ is the stress-energy tensor in $\Omega^{+}(\Omega^{-})$, $\tau_{\beta\lambda}$ and $\mathscr{T}_{\beta\lambda}$ describe stress-energy tensors associated with the hypersurface Σ. The above equations are equivalent to the equations

$$H_{\beta\lambda} - \frac{H}{2}g_{\beta\lambda} = -\kappa\tau_{\beta\lambda}, \tag{4-2}$$

$$G^{\pm}_{\beta\lambda} = -\kappa T^{\pm}_{\beta\lambda}, \tag{4-3}$$

and

$$J_{\beta\lambda} - \frac{J}{2}g_{\beta\lambda} = \kappa\mathscr{T}_{\beta\lambda} = [\Gamma^{\tau}_{\beta\rho}][\Gamma^{\rho}_{\tau\lambda}] - [\Gamma^{\tau}_{\beta\lambda}][\Gamma^{\rho}_{\tau\rho}] - \frac{1}{2}g_{\beta\lambda} \times\{[\Gamma^{\tau}_{\sigma\rho}][\Gamma^{\sigma}_{\tau\alpha}]g^{\rho\alpha} - \frac{b}{2}l_c g^{\sigma\alpha}[\Gamma^{\tau}_{\sigma\alpha}]\}. \tag{4-4}$$

We shall first discuss the case of a vacuum, i.e., the case for which

$$\tau_{\beta\lambda} = T_{\beta\lambda} = 0. \tag{4-5}$$

Equations (4-2) imply that

$$H_{\beta\lambda} = bl_\beta l_\lambda - l_\rho b^{\rho}_{\beta}l_\lambda - l_\rho b^{\rho}_{\lambda}l_\beta + l^\rho l_\rho b_{\beta\lambda} = 0. \tag{4-6}$$

These equations have as a consequence the statement that either

$$l^\sigma l_\sigma = 0 \tag{4-7}$$

and

$$l^\sigma b'_{\sigma\tau} = l^\sigma\Big(b_{\sigma\tau} - \frac{b}{2}g_{\sigma\tau}\Big) = 0, \tag{4-8}$$

or

$$l^\sigma l_\sigma \neq 0 \tag{4-9}$$

and

$$b_{\sigma\tau} = l_\sigma t_\tau + l_\tau t_\sigma \tag{4-10}$$

for an arbitrary vector field t_σ, for if we assume that the inequality (4-9) holds, it follows from Eqs. (4-6) that

$$p^{\sigma\tau}b_{\sigma\tau}l_\beta l_\lambda + l_\alpha l^\alpha p^{\sigma}_{\beta}p^{\tau}_{\lambda}b_{\sigma\tau} = 0,$$

with

$$p_{\sigma\tau} = g_{\sigma\tau} - l_\sigma l_\tau / l_\rho l^\rho.$$

Hence, we must have

$$p^{\sigma\tau}b_{\sigma\tau} = 0, \quad p^{\sigma}_{\lambda}p^{\tau}_{\beta}b_{\sigma\tau} = 0,$$

from which Eqs. (4-10) follow.

The solutions to Eq. (4-8) are given by

$$b_{\sigma\tau} = l_\sigma t_\tau + l_\tau t_\sigma + fm_\sigma m_\tau + \bar{f}\bar{m}_\sigma\bar{m}_\tau, \tag{4-11}$$

where t_σ is an arbitrary vector field, and m_σ is a complex null vector which together with l_σ and a real null vector n_σ form a null tetrad such that

$$g_{\mu\nu} = -l_\mu n_\nu - l_\nu n_\mu + m_\mu\bar{m}_\nu + \bar{m}_\mu m_\nu.$$

As was shown in Ref. 2, it is no restriction to assume that $t_\sigma = 0$ in Eqs. (4-10) and (4-11) for under the continuous transformation of coordinates

$$x^{*\mu} = x^\mu + \frac{\varphi^2}{2}t^\mu, \quad \varphi > 0,$$

$$x^{*\mu} = x^\mu, \qquad \varphi \leqslant 0,$$

$$b_{\mu\nu} \to b^{*}_{\mu\nu} = b_{\mu\nu} - l_\nu t_\mu - l_\mu t_\nu.$$

Equations (4-3) and (4-5) imply that

$$2[R_{\mu\nu}] = 0.$$

As a consequence of Eqs. (4-7), (4-8), and (1-11) this equation becomes

$$2l^\rho b_{\mu\nu;\rho} + l^{\rho}_{;\rho}b_{\mu\nu} - l_\nu\bar{\Phi}_\mu - l_\mu\Phi_\nu = -b_{\mu\nu}c^{\rho\sigma}l_\rho l_\sigma, \tag{4-12}$$

where

$$\Phi_\mu = b'^{\rho}_{\mu:\rho} + l_\rho\Big(\hat{b}^{\rho}_{\mu} - \frac{\hat{b}}{2}\delta^{\rho}_{\mu}\Big).$$

The colon again denotes the covariant derivative with respect to $\bar{\Gamma}$. It should be noted that as a consequence of Eqs. (4-7) and (4-8), we have

$$l^\rho[\Gamma^{\sigma}_{\tau\rho}] = 0.$$

Hence,

$$l^\rho\bar{\Gamma}^{\sigma}_{\tau\rho} = l^\rho\Gamma^{+\sigma}_{\tau\rho} = l^\rho\Gamma^{-\sigma}_{\tau\rho}$$

on Σ and the differentiation which occurs in Eq. (4-12) may be taken with respect to any of the connections Γ^{+}, Γ^{-}, or $\bar{\Gamma}$.

Equation (4-12) may be viewed as propagation equations for $b_{\mu\nu}$ in the direction of l^ρ. They do however involve the $\hat{b}_{\mu\nu}$ which are determined by the jump in the second derivative of the metric tensor across Σ. However, if we multiply these equations by $b'^{\mu\nu}$ and sum, we find that as a consequence of Eqs. (4-8) that

$$(\tau l^\mu)_{;\mu} = -\tau c^{\rho\sigma}l_\rho l_\sigma, \tag{4-13'}$$

where

$$\tau = b^{\mu\nu}b'_{\mu\nu} = b^{\mu\nu}\Big(b_{\mu\nu} - \frac{b}{2}g_{\mu\nu}\Big) = 2f\bar{f}. \tag{4-14}$$

When Eq. (4.7) holds in a neighborhood of Σ so that

$$\{(l_\mu l_\nu g^{\mu\nu})_{,\rho}\}^{+} = \{(l_\mu l_\nu g^{\mu\nu})_{,\rho}\}^{-}.$$

We have

$$c^{\mu\nu} l_\mu l_\nu = 0$$

and Eq. (4-13) becomes

$$(\tau l^\mu)_{;\mu} = 0\,. \tag{4-13}$$

It then follows from Eqs. (1-6) and (2-13) that

$$\begin{aligned} 4\left(J_{\sigma\mu} - \frac{J}{2} g_{\sigma\mu}\right) &= -4\kappa \mathscr{T}_{\sigma\mu} \\ &= \tau\left(l_\sigma l_\mu + \frac{l^\alpha l_\alpha}{2} g_{\sigma\mu}\right) + 2 l^\rho b'_{\rho\mu} l^\tau b'_{\tau\sigma} \\ &\quad - 2(l_\alpha l^\alpha) b'_{\sigma\rho} b^\rho_\mu - g_{\sigma\mu} l^\rho b'^{\nu}_{\rho} b'_{\tau\nu} l^\tau\,. \end{aligned} \tag{4-15}$$

When Eqs. (4-7) and (4-8) hold it follows from the above that

$$\kappa \mathscr{T}_{\sigma\mu} = -\frac{\tau}{4} l_\sigma l_\mu\,,$$

where τ is given by Eq. (4-14). The equation

$$\mathscr{T}^{\sigma\mu}_{\ \ ;\mu} = 0\,,$$

which in view of Eq. (4-1) is equivalent to the contracted Bianchi identities is also equivalent to Eq. (4-13) since

$$l_{\mu;\sigma} l^\sigma = l^\sigma l_{\sigma;\mu} = 0$$

because $l^\sigma l_\sigma = 0$ and $l_{\sigma;\mu} = l_{\mu;\sigma}$.

On substituting Eqs. (4-11) into the definition of $H_{\rho\sigma\mu\nu}$ one obtains

$$2H_{\rho\sigma\mu\nu} = -f P_{\rho\sigma} P_{\mu\nu} - \bar{f} \bar{P}_{\rho\sigma} \bar{P}_{\mu\nu}\,, \tag{4-16}$$

where

$$P_{\rho\sigma} = m_\rho l_\sigma - m_\sigma l_\rho = P^{\cup}_{\rho\sigma} = \frac{i}{2} \eta_{\rho\sigma\alpha\beta} P^{\alpha\beta}\,, \tag{4-17}$$

where for an arbitrary antisymmetric tensor $f_{\mu\nu}$, $f^{\cup}_{\mu\nu}$ denotes its dual defined as in the last of Eqs. (4-17) and $H^{\vee}_{\rho\sigma\mu\nu}$ is defined as

$$^{\cup}H^{\cup}_{\rho\sigma\mu\nu} = -\tfrac{1}{4} \eta_{\rho\sigma\alpha\beta} H^{\alpha\beta\gamma\delta} \eta_{\mu\nu\gamma\delta}\,.$$

$^{\cup}R^{\cup}_{\rho\sigma\mu\nu}$ is defined similarly. Thus,

$$2H_{\rho\sigma\mu\nu} = 2\,^{\cup}H^{\cup}_{\rho\sigma\mu\nu} \tag{4-18}$$

and

$$H_{\rho\sigma\mu\nu} l^\nu = 0\,,$$

where

$$\epsilon^{\mu\nu\alpha\beta} = \sqrt{-g}\,\eta^{\mu\nu\alpha\beta} = \frac{1}{\sqrt{-g}} \eta_{\mu\nu\alpha\beta} = \epsilon_{\mu\nu\alpha\beta}$$

and the ϵ's are the Levi–Civita alternating tensor densities, i.e., $H_{\rho\sigma\mu\nu}$ is an algebraically special tensor with the symmetry properties of a Riemannian curvature tensor whose Ricci tensor vanishes and whose conformal (Weyl) tensor is of Petrov type N.

It follows from Eqs. (1-9) and (2-11) that

$$[^{\cup}R^{\cup\rho\sigma\mu\nu}] l_\nu + {}^{\cup}H^{\cup\rho\sigma\mu\nu}_{\ ;\nu} = 0\,. \tag{4-19}$$

Thus, for a vacuum solution, i.e., a gravitational shock, for which $b_{\sigma\tau}$ is given by Eqs. (4-11) we find that the necessary and sufficient condition that the jump in the conformal tensor is of type N is that

$$(f P^{\mu\nu})_{;\nu} = f P^{\mu\nu} B_\nu\,, \tag{4-20}$$

where

$$B_\nu = n_\sigma l^\sigma_{\ ;\nu} - \bar{m}_\sigma m^\sigma_{\ ;\nu}\,.$$

This result follows from Eqs. (4-16), (4-17), and (4-19) along with the properties of the null tetrad. It is a further consequence of Eq. (4-20) that the null vector l_σ is shearfree. These results were obtained by Penrose in Ref. 3.

5. HIGH FREQUENCY GRAVITATIONAL WAVES AND SHOCKS

The discussion of gravitational shocks given in the preceding section characterizes them in terms of the tensors $b_{\mu\nu}$ and $\tilde{b}_{\mu\nu}$ defined on the shock hypersurface Σ and which describe the discontinuities in the first and second derivatives of the metric tensor $g_{\mu\nu}$ across Σ. These quantities satisfy the algebraic equations (4-8) and (4-9) which involve the normal l_μ to Σ and the propagation equations (4-12).

Very similar equations occur in the treatment of high frequency gravitational waves given by Y. Choquet-Bruhat.[2] In that theory it is assumed that the metric tensor may be expressed as

$$g_{\sigma\nu} = g^{(0)}_{\mu\nu}(x) + \frac{1}{\omega} g^{(1)}_{\mu\nu}(x;\xi) + \frac{1}{\omega^2} g^{(2)}_{\mu\nu}(x;\xi) + \cdots\,, \tag{5-1}$$

where

$$\omega \gg 1\,, \quad \xi = \omega\varphi(x)\,.$$

The role of Σ is played by the hypersurfaces of constant phase, the hypersurfaces φ constant. The role of $b_{\mu\nu}$ is played by the slowly varying amplitude of the gravitational wave described by $g^{(1)}_{\mu\nu}(x;\xi)$. The slowly varying amplitude part of $g^{(2)}_{\mu\nu}$ enters into a propagation equation for $g^{(1)}_{\mu\nu}$ in a manner similar to the way in which $\tilde{b}_{\mu\nu}$ enters into Eq. (4-12).

The equation given by Choquet-Bruhat which describe the behavior of high frequency gravitational waves are very similar to those given above in the treatment of gravitational shocks. On intuitive grounds one would expect that gravitational shocks would behave much the same as high frequency continuous solutions of the Einstein field equations and this is indeed the case as may be seen by comparing the results of the preceding sections with those given in Ref. 2.

An explanation of this fact is to be found in the fact that the two theories may be considered different versions of the theory that emerges from the single variational principle that determines the Einstein field equations, namely, the variational principle

$$\delta \int R \sqrt{-g}\, d^4x = 0\,. \tag{5-2}$$

If in this integral we replace R by

$$Q = \bar{\delta} H + R^D - \theta(1-\theta) J$$

and vary the $g_{\mu\nu}$, we obtain as field equations those equations characterizing gravitational shocks.

If instead of proceeding in this fashion we approximate this integral by the method use by Mac Callum and Taub[4] in applying the averaged Lagrangian technique, we find that we obtain the equations satisfied by the slowly varying amplitude of a high frequency gravitational wave. Thus, both of gravitational shock theory and the high-frequency wave theory have the same origin, namely, the variational principle (5-2), and it is therefore not surprising that the equations and results of the two theories are so similar.

6. SHELLS OF MATTER

The equations governing the behavior of shells of matter given by Papapetrou and Hamoui[5] and by Israël[6] follow from Eqs. (4-1) under the assumptions that

$$\tau_{\mu\nu} \neq 0$$

and

$$l^\rho l_\rho \neq 0 . \tag{6-1}$$

Both of these papers further require that $(T_{\mu\nu})^D = 0$. Equation (6-1) follows from the requirement that the hypersurface Σ is to be generated by the world lines of "particles" which constitute the shell.

It is a consequence of Eqs. (4-2) that

$$-2\,\kappa\tau_{\mu\nu} = l^\rho l_\rho b'_{\sigma\tau} p^\sigma_\mu p^\tau_\nu + p_{\mu\nu} b'_{\sigma\tau} l^\sigma l^\tau$$

or equivalently

$$-2\,\kappa\tau_{\mu\nu} = l_\rho l^\rho (b_{\sigma\tau} p^\sigma_\mu p^\tau_\nu - p_{\mu\nu} p^{\sigma\tau} b_{\sigma\tau}) \tag{6-2}$$

and hence

$$\tau_{\mu\nu} l^\nu = 0 . \tag{6-3}$$

The $b_{\sigma\tau}$ may be interpreted in terms of the jump in the second fundamental form of Σ, i.e., $K_{\mu\nu}$. When Σ is considered a hypersurface in space–time, we have

$$2[K_{\mu\nu}] = 2[l_{\sigma;\tau}] p^\sigma_\mu p^\tau_\nu = -2\,[l_\rho \Gamma^\rho_{\sigma\tau}] p^\sigma_\mu p^\tau_\nu ,$$

i.e.,

$$2[K_{\mu\nu}] = l_\rho l^\rho b_{\sigma\tau} p^\sigma_\mu p^\tau_\nu$$

and hence

$$[K_{\mu\nu} - p_{\mu\nu} K] = -\kappa\tau_{\mu\nu} . \tag{6-4}$$

Equations (4-3), namely, the equations

$$G^{\pm}_{\mu\nu} = -\kappa T^{\pm}_{\mu\nu} , \tag{6-5}$$

imply that

$$[G_{\mu\nu}] = -\kappa[T_{\mu\nu}] . \tag{6-6}$$

Hence,

$$[G_{\mu\nu}] l^\mu l^\nu = -\kappa[T_{\mu\nu}] l^\mu l^\nu = -\kappa\tau^{\mu\nu} l_{\mu:\nu} , \tag{6-7}$$

as follows from Eq. (1-14). From the definition of $l_{\mu:\nu}$, namely,

$$l_{\mu:\nu} = l_{\mu,\nu} - l_\rho \bar{\Gamma}^\rho_{\mu\nu} = \tfrac{1}{2}(l_{\mu,\nu} - l_\rho \Gamma^\rho_{\mu\nu})^+ + \tfrac{1}{2}(l_{\mu,\nu} - l_\rho \Gamma^\rho_{\mu\nu})^- ,$$

and from Eq. (6-3) which implies that

$$\tau_{\sigma\tau} p^\sigma_\mu p^\tau_\nu = \tau_{\mu\nu} ,$$

it follows that Eqs. (6-7) may be written as

$$[G_{\mu\nu} l^\mu l^\nu] = \frac{-\kappa}{2} \tau^{\mu\nu} (K^+_{\mu\nu} + K^-_{\mu\nu}) , \tag{6-8}$$

where

$$K^{\pm}_{\mu\nu} = (l_{\alpha,\beta} - l_\rho \Gamma^{\pm\rho}_{\alpha\beta}) p^\alpha_\mu p^\beta_\nu . \tag{6-9}$$

It is a consequence of the contracted Bianchi identities [Eqs. (3-4)], that

$$l_\nu \tau^{\mu\nu} = 0 , \tag{6-10}$$

$$\kappa\tau^{\mu\nu}_{;\nu} = -\kappa l_\nu [T^{\mu\nu}] = l_\nu [G^{\mu\nu}] , \tag{6-11}$$

and

$$(T^{\mu\nu}_{;\nu})^{\pm} = 0 . \tag{6-12}$$

Equations (6-10) and (6-11) arise from the propagation equations for $b_{\mu\nu}$, namely, Eqs. (1-14) and the field equations.

7. LINE SOURCES

Distribution valued curvature tensors with Dirac delta function behavior on two-dimensional manifolds in space–time arise in the discussion of some of the line sources treated by Israël[7] and in the treatment of conical singularities given by Sokolov and Starobinsky[8].

Israël assumes that one may introduce a coordinate system in space time in which the metric takes the form

$$ds^2 = d\rho^2 + g_{ab}(\rho, x^a) dx^a dx^b \tag{7-1}$$

and in which the line source is represented by the equation

$$\rho = 0 . \tag{7-2}$$

This equation does not describe a hypersurface in space–time but rather the history of a line source in space. A class of these sources are taken to be derivable from cylindrical shells of matter by a limiting process, i.e., one treats $\rho = \epsilon$ by the theory of shells described earlier and takes the limit as $\epsilon \to 0$.

The metrics describing the "Weyl struts" are shown by Israël to have a singularity in the metric similar to that of a two-dimensional cone. Sokolov and Starobinsky use the two-dimensional version of the Gauss–Bonnet formula to show these conical singularities may be described by a distribution valued curvature tensor, i.e., they essentially use Stokes' formula to evaluate the singular curvature tensor.

The Weyl canonical form of the static cylindrically symmetric metric is

$$ds^2 = e^{2(\nu - V)}(d\rho^2 + dz^2) + \rho^2 e^{-2V} d\varphi^2 - e^{2V} dt^2 . \tag{7-3}$$

The Einstein vacuum equations require that for $\rho \neq 0$

$$\nabla^2 V = V_{\rho\rho} + \frac{1}{\rho} V_\rho + V_{zz} = 0 ,$$

$$\nu_\rho = \rho(V^2_\rho - V^2_z) , \ \nu_z = 2\rho V_\rho V_z . \tag{7-4}$$

The axis

$$\rho = 0$$

is a two-dimensional manifold, describing the history of the line source where the curvature tensor is singular. One recalls that in the case of a two space consisting of a conical surface with a metric tensor

$$ds^2 = d\rho^2 + a^2\rho^2 d\varphi^2 ,$$

the Gauss–Bonnet formula tells us that the integral of the Gaussian curvature K over a region Q is

$$\int\!\!\int_Q K a\rho \, d\rho \, d\varphi = 2\pi - \int_{\partial Q} k_g ds ,$$

where k_g is the geodesic curvature of the curve bounding Q. This formula enables us to evaluate K for curves surrounding the point $\rho = 0$ and thus we may write

$$R^{12}_{12} = R^1_1 = R^2_2 = \frac{2\pi(1-a)}{a}\delta_2(\rho), \tag{7-5}$$

where $R^{\alpha\beta}_{\gamma\delta}$ $(\alpha, \beta, \gamma, \delta = 1, 2)$ is the curvature tensor of the two space and R^α_β is its Ricci tensor. The function $\delta_2(\rho)$ is defined by the requirement that

$$\int_0^\infty \int_0^{2\pi} \delta_2(\rho)\rho\, d\rho\, d\varphi = 1 .$$

If one calculates the curvature tensor for the metric given by Eq. (7-3), one again finds that $\rho = 0$ describes a singularity of this tensor. The singularity occurs in the components of the two form $R^{(1)}{}_{(2)}$ when one uses the one forms

$$\omega^0 = e^V dt, \qquad \omega^1 = e^{\nu - V} d\rho,$$
$$\omega^2 = \rho\, e^{-V} d\varphi, \quad \omega^3 = e^{\nu - V} dz.$$

Formulas similar to those occurring in Eqs. (7-5) express the singularity of the curvature tensor in terms of Dirac delta functions with support on the two surface $\rho = 0$, the history of the Weyl strut. The stress-energy tensor of this source of the gravitational field is given by

$$\tau^0_0 = \tau^3_3 = \frac{2\pi}{\kappa}(e^{\nu(0,z)} - 1)\delta_2(\rho)e^{-2(\nu - V)},$$
$$\tau^1_1 = \tau^2_2 = 0,$$

where the indices 0, 1, 2, and 3 correspond to t, ρ, φ, and z, respectively.

8. THE KERR SINGULARITY

Israël's treatment of the Kerr solution in Ref. 7 involves the observation that under the transformation

$$a + \varphi \cos\psi = \sqrt{r^2 + a^2}\sin\theta,$$
$$\rho \sin\psi = r\cos\theta,$$

the Kerr metric

$$ds^2 = \Sigma\left(\frac{dr^2}{\Delta} + d\theta^2\right) + (r^2 + a^2 + 2mar\Sigma^{-1}\sin^2\theta)\sin^2\theta\, d\varphi^2$$
$$- 4mar\Sigma^{-1}\sin^2\theta d\varphi\, dt - (1 - 2mr\Sigma^{-1})dt^2, \tag{8-1}$$

with

$$\Sigma = r^2 + a^2\cos^2\theta, \; \Delta = r^2 - 2mr + a^2,$$

becomes for small ρ $(r \approx 0, \theta \approx \pi/2)$

$$ds^2 = d\rho^2 + \rho^2 d\psi^2 + a^2 d\varphi^2 - dt^2 + \tfrac{1}{2}V(\rho, \psi)(ad\varphi - dt)^2, \tag{8-2}$$

where

$$V(\rho, \psi) = \frac{4m}{\sqrt{2a\rho}}\cos\psi/2, \tag{8-3}$$

i.e., the singularity in the Kerr metric, the two-surface $r = 0$, $\theta = \pi/2$ is that of the line source given by the metric (8-2), where V is a particular solution of the equation

$$V_{\rho\rho} + \frac{1}{\rho}V_\rho + \frac{1}{\rho^2}V_{\psi\psi} = 0.$$

Under the transformation

$$z = \frac{1}{\sqrt{2}}(a\varphi - t), \; t' = \frac{1}{\sqrt{2}}(a\varphi + t),$$
$$\varphi' = \psi.$$

The line element given by Eqs. (8-2) becomes

$$ds^2 = d\rho^2 + \rho^2(d\varphi')^2 + V(\rho, \varphi')dz^2 + 2dz\, dt'. \tag{8-4}$$

This is the line element that Israël uses to characterize a nonsimple line source.

The space–time described by the metric (8-4) (with the primes removed) will be required to satisfy the vacuum field equations except on the two-surface $\rho = \rho_0$. The metric will be assumed to be continuous across this manifold but will undergo a jump in its derivative. Thus, we must have for $\rho \neq \rho_0$

$$\nabla^2 V = V_{\rho\rho} + \frac{1}{\rho}V_\rho + \frac{1}{\rho^2}V_{\varphi\varphi} = 0,$$

as follows from the requirement that $G_{\mu\nu} = 0$, and across $\rho = \rho_0$

$$\left[\frac{\partial g_{\mu\nu}}{\partial x^\rho}\right] = [V_\rho]\delta^2_\mu\delta^2_\nu\delta^1_\sigma,$$

where we have used the notation

$$x^1 = \rho, x^2 = z, x^3 = \varphi, x^4 = t,$$

thus,

$$l_\mu = \delta^1_\sigma,$$
$$g_{\mu\nu}l^\mu l^\nu = g^{\mu\nu}l_\mu l_\nu = 1,$$

and

$$b_{\mu\nu} = [V_\rho]\delta^2_\mu\delta^2_\nu = [V_\rho]u_\mu u_\nu,$$
$$u_\mu = \delta^2_\mu.$$

Therefore,

$$b = [V_\rho]g^{22} = 0,$$

since

$$(g^\pm_{\mu\nu}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & V^\pm & 0 & 1 \\ 0 & 0 & \rho^2 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

$$\sqrt{-g} = \rho,$$

$$(g^{\pm\mu\nu}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1/\rho^2 & 0 \\ 0 & 1 & 0 & -V^\pm \end{pmatrix}.$$

It then follows from Eq. (6-2) that

$$-2\kappa\,\tau_{\mu\nu} = [V_\rho]u_\mu u_\nu, \tag{8-5}$$

where

$$g^{\mu\nu}u_\mu u_\nu = 0$$

and

$$u_\mu l^\mu = 0.$$

It is a consequence of Eq. (4-15) that

$$\mathscr{T}_{\mu\nu} = 0.$$

Israël constructs a thin shell source by taking

$$V = C\cos n\varphi \begin{cases} \rho^{-n}, & \rho > \rho_0, \\ \left(\dfrac{\rho}{\rho_0^2}\right)^n, & \rho \leqslant \rho_0. \end{cases}$$

Then

$$[V_\rho] = -2nC \cos n\varphi \rho_0^{-(n+1)}$$

and we may write

$$\kappa\tau_{\mu\nu} = \frac{C}{\rho_0} \cos n\varphi U_\mu U_\nu , \tag{8-6}$$

with

$$U_\mu = \sqrt{n}\rho_0^{-n/2} u_\mu = \sqrt{n}\rho_0^{-n/2}\delta^2_\mu . \tag{8-7}$$

Israël states: "For all finite ρ_0, and hence also in the limit $\rho_0 \to 0$ the source is composed of dust-like material with a 2^n-pole mass distribution streaming along the axis with the speed of light."

The Kerr metric in the neighborhood of its singular disk, where $r = 0$ and $\theta = \pi/2$, is taken by Israel to be a special case of the above with $n = 1/2$ and $C = 4m/\sqrt{2a}$. Thus, his treatment of the Kerr metric assumes that in the neighborhood of this disk, the metric is the limit of the form given by Eq. (8-2) whose source is a cylindrical shell of radius ρ_0 with vacuum regions inside and outside of the shell. The limiting process consists of letting ρ_0 approach zero. This type of metric is that of a space–time with a distribution valued curvature tensor which depends on the size of the discontinuity in the normal derivative of the metric tensor on the subspace $\rho = \rho_0$, namely, on $[V_\rho]$. This quantity also enters into the formulas for the gravitational mass and angular momentum of metric which is to represent the Kerr metric, i.e., the metric given by Eq. (8-2) for $\rho/a \ll 1$, with a discontinuous V_ρ at $\rho = \rho_0$, and given by Eq. (8-1) for large ρ.

We now turn to a discussion of the mass and angular momentum of Isräel's representation of the Kerr metric.

For a space–time with a metric tensor whose derivatives are discontinuous on a hypersurface one may define Killing vector fields to be those which satisfy the equations

$$(\xi_{\mu;\nu})^D + (\xi_{\nu;\mu})^D = 0 .$$

It is a consequence of these equations that

$$((\xi^{\rho;\mu})^D)_{;\rho} = -\xi^\rho Q^\mu_\rho \tag{8-8}$$

and that

$$\int_{V_4} (\xi^\rho Q^\mu_\rho)_{;\mu} \sqrt{-g}\, d^4x = \oint_{\partial V_4} \xi^\rho Q^\mu_\rho n_\mu d_3v = 0 , \tag{8-9}$$

where n_μ is the normal to the hypersurface enclosing the region V_4 of space–time. Equation (8-9) is a consequence of Eq. (8-8) and the fact that $(\xi^{\rho;\mu})^D$ is antisymmetric.

It follows from Eq. (8-8) that

$$\int_{V_3} \xi^\rho Q^\mu_\rho n_\mu d_3v = -\int_{V_3} (\xi^{\mu;\nu})^D_{;\nu} n_\mu d_3v . \tag{8-10}$$

Let ∂V_3 be the boundary of a three dimensional region in a hypersurface with unit normal n_μ. We shall deal with the hypersurface V_3 defined by

$$x^4 = t = \text{const}$$

and denote the interior of the cylindrical region $\rho = \rho_0$ by I and its exterior by E. We shall assume that E is a vacuum region. From Stokes' theorem we have

$$\int_{V_3} ((\xi^{\nu;\mu})^D)_{;\nu} n_\mu d_3v = -\int_{\partial V_3} ((\xi_{\sigma;\tau})^D)^* d\tau^{\sigma\tau} ,$$

where

$$((\xi_{\sigma;\tau})^D)^* = \tfrac{1}{2}\sqrt{-g}\,\epsilon_{\sigma\tau\mu\nu}(\xi^{\mu;\nu})^D ,$$

$$d\tau_{\alpha\tau} = \frac{\partial x^\sigma}{\partial y^i}\frac{\partial x^\tau}{\partial y^i}\epsilon^{i\nu} dy^1 dy^2 ,$$

and the equations describing the two-surface ∂V_3 bounding V_3 are

$$x^\mu = x^\mu(y^1, y^2) .$$

It then follows from Eqs. (8-11) and (2-11) and another application of Stokes' theorem that

$$\int_{V_3} \xi^\rho Q^\mu_\rho n_\mu d_3v = \int_I \xi^\rho Q^\mu_\rho n_\mu d_3v + \frac{1}{2}\int_{\partial I} [(\xi_{\sigma;\tau})^*] d\tau^{\sigma\tau} = \frac{1}{2}\int_{\Sigma_\infty} (\xi_{\sigma;\tau})^* d\tau^{\sigma\tau} , \tag{8-11}$$

where Σ_∞ is the two surface given in the coordinate system in which Eq. (8-1) obtains by the conditions $t =$ constant and $r = \infty$. We shall evaluate Eq. (8-11) for two Killing vector fields, namely, the vector fields having the components

$$\xi^\mu_b = \delta^\mu_b \quad (b = 0, 3)$$

in this coordinate system.

It may be readily verified that

$$\tfrac{1}{2}\int_{\Sigma_\infty} (\xi_{b\sigma;\tau})^* d\tau^{\sigma\tau} = 4\pi m(\delta^0_b - 2a\delta^3_b) , \tag{8-12}$$

i.e., that m and ma are, respectively, the mass and angular momentum of the Kerr solution.

We shall use the coordinate system in which Eq. (8-2) holds in the discussion of the left hand side of Eq. (8-11). We write

$$x^0 = t, x^1 = \rho, x^2 = \psi, x^3 = \phi .$$

The variables ψ and ϕ may be taken to be the parameters on the surface $\rho =$ constant, $t =$ constant. In this coordinate system

$$b_{\mu\nu} = \tfrac{1}{2}[V_\rho] u_\mu u_\nu ,$$

where V is given by Eq. (8-3) and

$$u_\mu = \delta^0_\mu - a\delta^3_\mu, \; u^\mu = -\frac{1}{a}(a\delta^\mu_0 + \delta^\mu_3) .$$

Further,

$$l_\mu = \delta^1_\mu, \; l^\mu = \delta^\mu_1 .$$

It may be verified from Eqs. (2-11) that

$$u^\nu Q^\mu_\nu n_\mu = -\frac{1}{a}(a\xi^\nu_0 + \xi^\nu_3) Q^\mu_\nu n_\mu = 0,$$

when n_μ is the normal to the hypersurface $t =$ constant. However, Eqs. (8-11) and (8-12) imply that for a metric approaching the Kerr metric for large ρ we must have

$$\int_{V_3} u^\nu Q^\mu_\nu n_\mu d_3v = 4\pi m . \tag{8-13}$$

Hence, the assumptions made by Isräel in his discussion of the Kerr metric do not lead to correct values for both the mass and angular momentum of the Kerr metric.

The requirement that the metric in region I be of the form given by Eq. (8-2) is not imposed by the other assumptions made by Isräel. One could assume that the metric is of the form

$$ds^2 = d\rho^2 + \rho^2 d\psi^2 + g_{AB} dx^A dx^B \quad (A, B) = (0, 3),$$

where the g_{AB} are functions of ρ and ψ which reduce this line element to that given by Eq. (8-2) when $\rho = \rho_0$, and which may have derivatives with respect to ρ whose values at $\rho = \rho_0$ disagree with those of the coefficients of the latter line element. Thus, by changing the assumption regarding the metric in the region I it may be possible to satisfy Eq. (8-13) and also obtain the correct value for the mass to be associated with the metric. The details of this change will not be treated here.

The technique of treating a space–time with a two-dimensional subspace which is the support of a distribution valued curvature tensor as the limit of one with cylindrical shells suggests that space–times with a world line as a support of such a curvature tensor be treated as the limit of one with spherical shells. Such a treatment will be discussed elsewhere.

[1]A. Lichnerowicz, C.R. Acad. Sci. **273**, 528 (1971); Instituto Nazionale di Alta Matematica, Symposia Matematica, Volume XII (1973), p. 93.
[2]Y. Choquet-Bruhat, Commun. Math. Phys. **12**, 16 (1969).
[3]R. Penrose, *The Geometry of Impulsive Gravitational Waves*; *General Relativity*, edited by L.O'Raifeartaigh (Clarendon, Oxford, 1972), pp. 101–118.
[4]M.A.H. MacCallum and A.H. Taub, Commun. Math. Phys. **30**, 153 (1973).
[5]A. Papapetrou and A. Hamoui, Ann. Inst. Henri Poincaré **9**, 179 (1968).
[6]W. Israël, Nuovo Cimento **44**, 1 (1966).
[7]W. Israël, Phys. Rev. D **15**, 935 (1977).
[8]D.D. Sokolov and A.A. Starobinsky, Dokl. Akad. Nauk SSSR **234**, 1043 (1977).

Gravitation as an internal gauge theory of the Poincaré group

C. P. Luehr
Department of Mathematics, University of Florida, Gainesville, Florida 32611
M. Rosenbaum P. [a]
Centro de Estudios Nucleares, Universidad Nacional, Autonoma de Mexico, Circuito Exterior, C. U., México 20, D.F., México

(Received 21 November 1979; accepted for publication 11 January 1980)

We describe a gauge theory of gravitation based on the Poincaré group treated as an internal symmetry group acting on five-dimensional fiber spaces. One special feature of the theory is that the metric structure of the manifold which serves as the base space appears as a natural consequence of the formalism and is not imposed *a priori*.

The theory encompasses general relativity and the Einstein–Cartan theory as particular cases, as well as other gravitational theories with torsion which have been recently proposed.

I. INTRODUCTION

Whenever one has a field theory which is globally invariant under a Lie group $\mathscr{G}$ of transformations one can introduce compensating (gauge) fields in order to extend this global invariance to a local one. The result is a theory of the interaction of the original and the gauge fields. The free theory of the gauge fields is usually called the gauge theory of the group $\mathscr{G}$.

Gauge theories play a fundamental role in the description of the basic interactions of nature.[1]

Electromagnetism and gravitation are the two basic long range interactions of nature which have been studied for a long time. Maxwell's and Einstein's classical theories seem to describe them reasonably well (within certain ranges). Maxwell's electromagnetism is well understood[2] as a gauge theory of the invariance group U(1) [of charged (complex) field theories] and although general relativity has all the features of gauge theories (arbitrary functions appearing in the description of the fields and the existence of constraints, i.e., nondynamical equations of motion) there is still some controversy about its proper treatment as the gauge theory of a group.[3-10] It is interesting then, to study in detail the basic ideas that give rise to the gaugelike treatment of general relativity.[3,4,6]

The work of Kibble[6] has been the starting point of many of the current papers on the subject. In reviewing Ref. 6 carefully, we find that some of the apparently well-established results cannot be properly justified following the Utimaya-Kibble approach. In fact, several authors[4,6,7,9] agree in stating that general relativity is a special case (for vanishing torsion) of the gauge theory of the Poincaré group. A closer look at the approach reveals that the "group" in question is not really the Poincaré group and as a matter of fact, it may not even be a Lie group.

In order to elucidate this point, we briefly summarize the essential features of gauge field theories.

Let $\mathscr{M}$ be a manifold, $\mathscr{T}_q$ be the tangent vector space at an arbitrary point $q \in \mathscr{M}$, and $\mathscr{T}'_q$ be the dual space to $\mathscr{T}_q$.

[a] Supported in part by International Scientific Exchange Program Grants, National Science Foundation OIP75-09783AO1 and Consejo Nacional de Ciencia y Technología No. 955.

General theory with global group: Let $\mathscr{V}$ be a vector space, $\mathscr{V}'$ the dual space to $\mathscr{V}$, and $\mathscr{G}$ be a group of linear transformations on $\mathscr{V}$. Besides being a vector space, $\mathscr{V}$ can have additional structure, e.g. an inner product, which is assumed to be invariant under the action of the group.

Each derivation operator X on $\mathscr{M}$ can act on any $\mathscr{V}$-valued vector field $\mathbf{v}$ to give another $\mathscr{V}$-valued vector field $X\mathbf{v}$ defined uniquely by the equation

$$\mathbf{c}' \circ (X\mathbf{v}) = X(\mathbf{c}' \circ \mathbf{v})$$

for every $\mathbf{c}'$ (constant dual vector) in $\mathscr{V}'$. Note that $\mathbf{c}' \circ \mathbf{v}$ is a scalar field whose value at q is $\mathbf{c}' \circ \mathbf{v}(q)$, where "$\circ$" denotes the action of the dual vector $\mathbf{c}'$ on the vector $\mathbf{v}(q)$ to give a scalar.

Under suitable conditions, e.g. if $\mathscr{M}$ has a metric structure, we can set up a Lagrangian for a $\mathscr{V}$-valued vector field which is invariant under the global group $\mathscr{G}$.

General theory with local groups. For each point $q \in \mathscr{M}$ let there be a separate vector space $\mathscr{V}_q$ with a structure isomorphic to that of $\mathscr{V}$, and a group $\mathscr{G}_q$ of linear transformations on $\mathscr{V}_q$ isomorphic to $\mathscr{G}$.

Let $\mathbf{w}$ be any $\mathscr{V}_q$-valued vector field on $\mathscr{M}$, i.e., its value $\mathbf{w}(q)$ at q is in $\mathscr{V}_q$. Assume that we have a covariant derivative operator D_X corresponding to each derivation operator X. D_X acts on each $\mathscr{V}_q$-valued vector field $\mathbf{w}$ to give another $\mathscr{V}_q$-valued vector field $D_X\mathbf{w}$.

Assume that D_X has certain properties relating it to the structure of the spaces $\mathscr{V}_q$ and the groups $\mathscr{G}_q$. These include the following:

(1) Given any two points $q, p \in \mathscr{M}$ and a curve joining them, the structure of the vector space $\mathscr{V}_q$ goes into the corresponding structure of the vector space $\mathscr{V}_p$ under parallel transport along the curve by means of the operator D_X.

(2) Given any $q \in \mathscr{M}$ and a closed curve in $\mathscr{M}$ starting and ending at q, parallel transport of vectors around this curve from q to q is a linear transformation on $\mathscr{V}_q$ that belongs to $\mathscr{G}_q$.

(3) Given any two points $q, p \in \mathscr{M}$ and a curve joining them, parallel transport along that curve of linear transformations on $\mathscr{V}_q$ into linear transformations on $\mathscr{V}_p$ generates an isomorphism of $\mathscr{G}_q$ onto $\mathscr{G}_p$.

 0022-2488/80/061432-07$1.00

If $\mathscr{G}_q$ coincides with the set of all linear transformations on $\mathscr{V}_q$ that preserves the structure of $\mathscr{V}_q$, then property (1) implies properties (2) and (3).

Under suitable conditions, e.g. if $\mathscr{M}$ has a metric structure, we can then set up a Lagrangian for a $\mathscr{V}_q$-valued vector field and corresponding gauge fields which is invariant under the local groups $\mathscr{G}_q$.

Minimal coupling principle: The general prescription of gauge theory consists of replacing $\mathscr{V}$-valued vector fields and derivation operators X respectively by $\mathscr{V}_q$-valued vector fields and covariant derivative operators D_X in the Lagrangian. To determine the equations of motion of the D_X, which occur in the theory as additional fields (the gauge fields), a free Langrangian term must be included.

Details of this procedure for extending internal global symmetries to local ones, thus leading to an interacting gauge theory of fields can be found in the work of Utimaya[3] and others.[10]

A direct application of the formalism to external symmetry groups[3,4,6–9] poses, however, some basic problems which may be best seen by recalling the essential steps followed in the above cited papers for the Poincaré group.

Global and local Poincaré invariance: The local variation of a field ψ^A (where A labels the components of the field) under infinitesimal Poincaré transformations in Minkowski space is given by[4,6]

$$\delta_* \psi^A = -\epsilon^i \partial_i \psi^A + \tfrac{1}{2}\epsilon^{ij}[\delta^A_B(x_i\partial_j - x_j\partial_i) + S_{ij}{}^A_B]\psi^B \qquad (1.1)$$

and

$$\delta x^i = \epsilon^i_j x^j + \epsilon^i , \qquad (1.2)$$

where ϵ^i and $\epsilon^{ij} = -\epsilon^{ji}$ are ten infinitesimal constant parameters characterizing the infinitesimal Poincaré transformation and $S_{ij}{}^A_B = -S_{ji}{}^A_B$ are the generators of the Lorentz group appropriate to the generic field.

As is well known, the generators ∂_i and $J_{ij} = x_i\partial_j - x_j\partial_i + S_{ij}$ satisfy the Poincaré algebra

$$[\partial_i,\partial_j] = 0 , \qquad (1.3a)$$

$$[J_{ij},\partial_k] = \eta_{ki}\partial_j - \eta_{kj}\partial_i , \qquad (1.3b)$$

$$[J_{ij},J_{kl}] = \eta_{jk}J_{il} - \eta_{il}J_{jk} + \eta_{jl}J_{ik} + \eta_{ik}J_{jl} . \qquad (1.3c)$$

To label the group generators, we introduce an orthonormal coordinate basis such that

$$\mathbf{e}_\alpha \cdot \mathbf{e}_\beta = \eta_{\alpha\beta} = \eta^{\alpha\beta} = \mathrm{diag}(-1,1,1,1) , \qquad (1.4)$$

and define for any vector $\mathbf{V}$

$$V_\alpha \equiv \mathbf{e}_\alpha \cdot \mathbf{V},$$

and similarly for tensors. Note that in flat spacetime the coordinates of all such bases are related by $x^\alpha = \lambda_i^{\ \alpha} x^i + b^\alpha$, where $\lambda_i^{\ \alpha}$ and b^α are constants. Hence, we write equation (1.1) as

$$\delta_* \psi^A = -\epsilon^\alpha \partial_\alpha \psi^A + \tfrac{1}{2}\epsilon^{\alpha\beta}[(x_\alpha\partial_\beta - x_\beta\partial_\alpha)\delta^A_B + S_{\alpha\beta}{}^A_B]\psi^B, \qquad (1.5)$$

where

$$\epsilon^\alpha \equiv \lambda_i^{\ \alpha}\epsilon^i + \epsilon^{ij}\lambda_i^{\ \alpha}\lambda_j^{\ \beta} b_\beta \quad \text{and} \quad \epsilon^{\alpha\beta} \equiv \lambda_i^{\ \alpha}\lambda_j^{\ \beta}\epsilon^{ij} .$$

A special relativistic field theory is constructed by giving a Lagrangian density $\mathscr{L}(\psi^A,\partial_i\psi^A)$ which behaves as a scalar under Poincaré transformations. The density $\mathscr{L}$ is said to be invariant under a transformation defined by $\delta\psi^A = \delta_*\psi^A - \delta x^\mu \partial_\mu \psi^A$ and $\delta x^\mu = \xi^\mu(x)$ if

$$\delta\mathscr{L} \equiv \frac{\partial\mathscr{L}}{\partial\psi^A}\delta\psi^A + \frac{\partial\mathscr{L}}{\partial(\partial_\mu\psi^A)}\delta(\partial_\mu\psi^A) + \mathscr{L}\partial_\mu\xi^\mu = 0 , \qquad (1.6)$$

which implies that the current

$$J^\mu = \frac{\partial\mathscr{L}}{\partial(\partial_\mu\psi^A)}\delta_*\psi^A + \mathscr{L}\xi^\mu \qquad (1.7)$$

is conserved.

Special relativistic theories are invariant under the Poincaré group transformation. Therefore, ten currents (the 4-momentum vector and total angular momentum tensor) are conserved. Extending the global symmetry to a local one in accordance with Utimaya's scheme involves making the ten constant parameters ϵ^α and $\epsilon^{\alpha\beta}$ arbitrary functions of spacetime, i.e., writing

$$\delta_*\psi^A = -\epsilon^\alpha(x)\partial_\alpha\psi^A + \tfrac{1}{2}\epsilon^{\alpha\beta}(x)[(x_\alpha\partial_\beta - x_\beta\partial_\alpha)\delta^A_B + S_{\alpha\beta}{}^A_B]\psi^B . \qquad (1.8)$$

Note, however, that because of the explicit appearance of spacetime coordinates in the orbital part of the angular momentum term in Eq. (1.8), the prescription is inadequate for arriving at a generally covariant theory.

To obviate this problem, Kibble[6] proposed rearranging Eq. (1.8) into the form

$$\delta_*\psi^A = -\xi^\alpha(x)\partial_\alpha\psi^A + \tfrac{1}{2}\epsilon^{\alpha\beta}(x)S_{\alpha\beta}{}^A_B\psi^B , \qquad (1.9)$$

where

$$\xi^\alpha(x) = \epsilon^\alpha(x) + \epsilon^{\alpha\beta}(x)x_\beta . \qquad (1.10)$$

This approach, which might be justifiable from a mathematical point of view, is conceptually problematic. In fact, if we rearrange terms in (1.5) first, we get

$$\delta_*\psi^A = -(\epsilon^\alpha + \epsilon^{\alpha\beta}x_\beta)\partial_\alpha\psi^A + \tfrac{1}{2}\epsilon^{\alpha\beta}S_{\alpha\beta}{}^A_B\psi^B . \qquad (1.11)$$

"Gauging" the parameters in (1.11) would then lead to (1.9). But, because of the explicit coordinate dependence in the parameter of the generator of translations, the transformations being gauged are not a Lie group. Moreover, if we make the parameters in (1.9) coordinate-independent, we do not retrieve the original Eq. (1.5) nor the invariance group of the theory we started with, as also becomes apparent when we consider what infinitesimal global transformations are required in order that Utimaya's procedure will result in the Eq. (1.9).

In order to avoid these problems and arrive at an unambiguous Poincaré gauge theory of gravitation, we develop in this paper a formalism based on treating a five-dimensional faithful representation of the Poincaré group as an internal group and using fiber bundle techniques as a natural framework for a geometric and coordinate-free discussion of the theory.

In Sec. II we present the essential features of the Poincaré group as a group of linear transformations on a five-dimensional space. In Sec. III we identify this space with a typical fiber of a vector bundle, in which the base manifold has no assumed metric structure, and it is shown how the local Poincaré group leads to a way of uniquely imposing a Minkowski metric structure on this space. Section IV contains a summary and discussion of our basic results.

II. FIVE-DIMENSIONAL FAITHFUL REPRESENTATION OF THE POINCARÉ GROUP

Let $\mathscr{E} = (\mathscr{E}, f'_o, \cdot)$ be a space with the following structure:

(1) $\mathscr{E}$ is a five-dimensional real vector space,

(2) A priviledged element $\mathbf{f}'_o \neq \mathbf{0}' \in \mathscr{E}'$ is given,

(3) A Minkowski inner product $\mathbf{u}\cdot\mathbf{v}$ with the signature $(- + + +)$ is given on the subspace $\mathscr{H}$ of $\mathscr{E}$ where $\mathscr{H} = \{\mathbf{w} | \mathbf{w} \in \mathscr{E}, \mathbf{f}'_o \circ \mathbf{w} = 0\}$. With this structure, $\mathscr{E}$ is a faithful representation space of the Poincaré group $\mathscr{P}$.

The hyperplane $\mathscr{K} = \{\mathbf{w} | \mathbf{w} \in \mathscr{E}, \mathbf{f}'_o \circ \mathbf{w} = 1\}$, which is parallel to $\mathscr{H}$, can be interpreted as the flat spacetime manifold of special relativity. Each point of spacetime is represented by a vector in $\mathscr{K}$, and we will define the action of the Poincaré group on $\mathscr{E}$ in such a way that $\mathscr{K}$ remains invariant. Any vector $\mathbf{k}_o \in \mathscr{K}$ can be chosen as an "origin vector" to represent a choice of an origin in spacetime. Those Poincaré transformations under which $\mathbf{k}_o$ is invariant will be called Lorentz transformations with respect to $\mathbf{k}_o$.

The representation of the Poincaré group in $\mathscr{E}$ will consist of all linear transformations $\mathbf{P} \in \mathscr{E} \otimes \mathscr{E}'$ such that

$$\mathbf{f}'_o \circ \mathbf{P} = \mathbf{f}'_o , \tag{2.1}$$

$$(\mathbf{P}\circ\mathbf{u})\cdot(\mathbf{P}\circ\mathbf{v}) = \mathbf{u}\cdot\mathbf{v} , \tag{2.2}$$

for all $\mathbf{u}, \mathbf{v} \in \mathscr{H}$.

Note that, as a consequence of (2.1), $\mathbf{u} \in \mathscr{H}$ implies that $\mathbf{P}\circ\mathbf{u} \in \mathscr{H}$; thus $\mathscr{H}$ is invariant under $\mathbf{P}$. This result is needed to make (2.2) consistent, since the dot product is defined only for elements of $\mathscr{H}$. We also have that $\mathbf{u} \in \mathscr{K}$ implies $\mathbf{P}\circ\mathbf{u} \in \mathscr{K}$, i.e., $\mathscr{K}$ is invariant under $\mathbf{P}$.

A Lorentz transformation with respect to $\mathbf{k}_o$ as origin vector in $\mathscr{K}$, is accomplished by any Poincaré transformation $\mathbf{L}$ satisfying the additional property

$$\mathbf{L}\circ\mathbf{k}_o = \mathbf{k}_o . \tag{2.3}$$

To construct a translation on $\mathscr{K}$ by a vector $\mathbf{t} \in \mathscr{H}$, we define a unit transformation $\mathbf{E}_o \in \mathscr{E} \otimes \mathscr{E}'$ such that $\mathbf{E}_o \circ \mathbf{u} = \mathbf{u}$ for all $\mathbf{u} \in \mathscr{E}$, and let

$$\mathbf{T} = \mathbf{E}_o + \mathbf{t} \otimes \mathbf{f}'_o . \tag{2.4}$$

For all $\mathbf{h} \in \mathscr{H}$, we have

$$\mathbf{T}\circ\mathbf{h} = \mathbf{h} , \tag{2.5}$$

and for all $\mathbf{u} \in \mathscr{K}$, we have

$$\mathbf{T}\circ\mathbf{u} = \mathbf{u} + \mathbf{t} \in \mathscr{K} . \tag{2.6}$$

Moreover, any vector $\mathbf{u} \in \mathscr{E}$ may be uniquely decomposed into the form

$$\mathbf{u} = \alpha\mathbf{k}_o + \mathbf{h}, \tag{2.7}$$

where $\mathbf{h} \in \mathscr{H}$. Then, for any Lorentz transformation $\mathbf{L}$, we have

$$\mathbf{L}\circ\mathbf{u} = \mathbf{L}\circ(\alpha\mathbf{k}_o + \mathbf{h}) = \alpha\mathbf{k}_o + \mathbf{L}\circ\mathbf{h} . \tag{2.8}$$

But an arbitrary Poincaré transformation can always be expressed as $\mathbf{P} = \mathbf{T}\circ\mathbf{L}$ where $\mathbf{T}$ and $\mathbf{L}$ are given by Eqs. (2.4) and (2.8). Hence

$$\begin{aligned}\mathbf{P}\circ\mathbf{u} &= \mathbf{T}\circ\mathbf{L}\circ\mathbf{u} = \mathbf{T}\circ\mathbf{L}\circ(\alpha\mathbf{k}_o + \mathbf{h}) \\ &= \mathbf{T}\circ(\alpha\mathbf{k}_o + \mathbf{L}\circ\mathbf{h}) = \alpha(\mathbf{k}_o + \mathbf{t}) + \mathbf{L}\circ\mathbf{h} .\end{aligned} \tag{2.9}$$

If $\alpha = 1$, then $u \in \mathscr{K}$, and the above equation becomes

$$\mathbf{P}\circ\mathbf{u} = \mathbf{P}\circ(\mathbf{k}_o + \mathbf{h}) = \mathbf{k}_o + \mathbf{L}\circ\mathbf{h} + \mathbf{t} . \tag{2.10}$$

III. GAUGE THEORY OF THE POINCARÉ GROUP

Making use of the structures introduced in the previous section, with $\mathscr{E} = (\mathscr{E}, \mathbf{f}'_o, \cdot)$ serving as a representation space in terms of which the Poincaré group may be treated as an internal symmetry group, we can now formulate a gauge field theory of gravitation which is locally invariant under $\mathscr{P}$ and which does not present the problems, discussed in the introduction, of other formalisms appearing in the literature.

For this purpose, we construct the vector bundle[10,11] $(\mathscr{E}(\mathscr{M}), \mathscr{M}, \mathscr{E}, \pi, \mathscr{P}, \phi)$, where $\mathscr{E}(\mathscr{M})$ denotes the bundle space, the base space $\mathscr{M}$ is a four-dimensional manifold, $\mathscr{E}$ is the typical fiber (or standard fiber space) which we identify with the five-dimensional space with the structure given in the previous section, π is the surjective projection of $\mathscr{E}(\mathscr{M})$ onto $\mathscr{M}$ and $\mathscr{P}$ is the structural group of the bundle. The bundle satisfies, in addition, the condition of local triviality, which implies that there exists a covering of $\mathscr{M}$ by neighborhoods $\{U_j\}$ and a homeomorphism ϕ_{U_j} of $\pi^{-1}(U_j)$ onto the topological product $U_j \times \mathscr{E}$ such that, for $\rho = (q, \mathbf{v}_q) \in \mathscr{E}(\mathscr{M})$ with $q \in U_j$ and $\mathbf{v}_q \in \pi^{-1}(q) = \mathscr{E}_q$, we have $\phi_{U_j}(\rho) = (q, \bar{\phi}_{U_j}(\mathbf{v}_q))$ with $\bar{\phi}_{U_j}$ denoting a homeomorphism of $\mathscr{E}_q$ onto $\mathscr{E}$. Also, for each q, as a part of the structure of $\mathscr{E}_q$, we have $\mathbf{f}'(q) \neq \mathbf{0}'$ and the Minkowski inner product on $\mathscr{H}_q = \{\mathbf{w}_q | \mathbf{w}_q \in \mathscr{E}_q, \mathbf{f}'(q)\circ\mathbf{w}_q = 0\}$. A choice of an origin at each q is specificied by a vector field $\mathbf{k}$ whose value at q is a vector $\mathbf{k}(q) \in \mathscr{K}_q = \{\mathbf{w}_q | \mathbf{w}_q \in \mathscr{E}_q, \mathbf{f}'(q)\circ\mathbf{w}_q = 1\}$. Thus the action of the structure group in each fiber $\mathscr{E}_q$ is represented by the group of linear transformations $\mathscr{P}_q$ which preserve the structure of $\mathscr{E}_q$.

Let $\mathbf{v}$ be a morphism $\mathbf{v}: \mathscr{M} \to \mathscr{E}(\mathscr{M})$ such that $\pi\mathbf{v} = Id_{\mathscr{M}}$, then $\mathbf{v}$ is a cross section of our vector bundle and it defines a vector field in $\mathscr{E}(\mathscr{M})$, i.e., it associates a vector $\mathbf{v}(q)$ in $\mathscr{E}_q$ with each point q in $\mathscr{M}$.

Denote by $\Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$ the space of all smooth cross sections of $\mathscr{E}(\mathscr{M})$, so $\mathbf{v} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$.

A connection $\mathbf{D}$ on $\mathscr{E}(\mathscr{M})$ is a linear differential operator from the space of sections of $\mathscr{E}(\mathscr{M})$ to sections of the bundle $\mathscr{T}'(\mathscr{M}) \otimes \mathscr{E}(\mathscr{M})$, where $\mathscr{T}'(\mathscr{M})$ is the dual tangent bundle over $\mathscr{M}$, i.e.,

$$\mathbf{D}: \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M})) \to \Gamma(\mathscr{M}, \mathscr{T}'(\mathscr{M}) \otimes \mathscr{E}(\mathscr{M})) . \tag{3.1}$$

If we define

$$\mathbf{x}\circ\mathbf{D} \otimes \mathbf{v} \equiv D_X \mathbf{v}, \tag{3.2}$$

where

$$\mathbf{x} \equiv X \in \Gamma(\mathscr{M}, \mathscr{T}(\mathscr{M})) \quad \text{and} \quad D_X \mathbf{v} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M})) ,$$

then D_X satisfies the following axioms:

$$D_X(\mathbf{v} + \mathbf{w}) = D_X\mathbf{v} + D_X\mathbf{w} , \tag{3.3a}$$

$$D_X(g\mathbf{v}) = (Xg)\mathbf{v} + gD_X\mathbf{v}\,, \tag{3.3b}$$

$$D_{X+Y}\mathbf{v} = D_X\mathbf{v} + D_Y\mathbf{v}\,, \tag{3.3c}$$

$$D_{gX}\mathbf{v} = gD_X\mathbf{v}\,, \tag{3.3d}$$

where $g \in \Gamma(\mathscr{M}, \mathbb{R})$ is a smooth scalar field.

Since $\mathbf{f}'(q)$ is part of the structure of each fiber $\mathscr{E}_q$ which is to be preserved under parallel transport, we require that

$$D_X\mathbf{f}' = \mathbf{0}'\,, \tag{3.4}$$

which is equivalent to

$$\mathbf{f}' \circ (D_X\mathbf{v}) = X(\mathbf{f}' \circ \mathbf{v})\,. \tag{3.5}$$

Also, since the inner product in each $\mathscr{H}_q$ is part of the structure of $\mathscr{E}_q$ and thus is required to be preserved under the action of $\mathscr{P}_q$, the compatibility condition with the inner product

$$X(\mathbf{h}\cdot\mathbf{l}) = (D_X\mathbf{h})\cdot\mathbf{l} + \mathbf{h}\cdot(D_X\mathbf{l}) \tag{3.6}$$

must be satisfied. Here $\mathbf{h}, \mathbf{l}, \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$ and have their values $\mathbf{h}(q), \mathbf{l}(q) \in \mathscr{H}_q$.

At this point, the theory already differs from the typical internal gauge theory in one respect. The manifold $\mathscr{M}$ is not given any metric structure. In the typical gauge theory, the metric is essential in order to construct a Lagrangian. However, with the theory we are now constructing, it will be possible to write down a Lagrangian without initially assuming a metric structure on $\mathscr{M}$.

The presence of the origin vector field $\mathbf{k}$ is another way that the theory differs from the typical gauge theory. We may regard each $\mathbf{k}(q)$ as being the point at which each $\mathscr{H}_q$ is tied to the manifold. In the following discussion, we use this field $\mathbf{k}$ to define a unique map from the tangent space $\mathscr{T}_q$ onto $\mathscr{H}_q$. This map leads to a unique way of imposing a metric structure and connection on the tangent bundle $(\mathscr{T}(\mathscr{M}), \mathscr{M}, \mathbb{R}^4, \pi_{\mathscr{T}}, GL(4,\mathbb{R}))$, where $\pi_{\mathscr{T}}^{-1}(q) = \mathscr{T}_q$ is the fiber above q.

Induced structure on $\mathscr{T}(\mathscr{M})$ from $\mathscr{E}(\mathscr{M})$: Consider the "origin vector" field $\mathbf{k}$, and define the tensor field $\mathbf{J}$ with value $\mathbf{J}(q)$ at q in $\mathscr{T}'_q \otimes \mathscr{E}_q$ by

$$\mathbf{J} = \mathbf{D} \otimes \mathbf{k}\,. \tag{3.7}$$

Theorem 1: $\mathbf{J}(q) \in \mathscr{T}'_q \otimes \mathscr{H}_q \subset \mathscr{T}'_q \otimes \mathscr{E}_q\,.$ (3.8)

Proof: From the definition of covariant differentation of dual vector fields, we have

$$X(\mathbf{f}' \circ \mathbf{k}) = (D_X\mathbf{f}') \circ \mathbf{k} + \mathbf{f}' \circ (D_X\mathbf{k}) \tag{3.9}$$

for every X. Moreover, making use of (3.4) and the condition that $\mathbf{f}' \circ \mathbf{k} = 1$ for all q, we get

$$(D_X\mathbf{k}) \circ \mathbf{f}' = \mathbf{x} \circ (\mathbf{D} \otimes \mathbf{k}) \circ \mathbf{f}' = 0\,, \tag{3.10}$$

i.e., $(\mathbf{D} \otimes \mathbf{k})_q \in \mathscr{T}'_q \otimes \mathscr{H}_q$. Q.E.D.

Note that in order to get a nontrivial theory, we do not want $\mathbf{k}$ to be preserved under parallel transport. This is consistent with the fact that the selection of $\mathbf{k}(q)$ is not a part of the structure of $\mathscr{E}_q$ that is used in defining the group $\mathscr{P}_q$. We will show later on that a change in $\mathbf{k}$ leads to an equivalent theory.

At each q, $\mathbf{J}(q)$ maps $\mathscr{T}_q$ into $\mathscr{H}_q$ as follows:

$$\mathbf{x}_q \in \mathscr{T}_q \rightarrow \mathbf{x}_q \circ \mathbf{J}(q) \in \mathscr{H}_q\,. \tag{3.11}$$

If we assume $\mathbf{J}$ to be nonsingular for each q, then the map is a bijection and we can use it to impose the metric structure of $\mathscr{H}_q$ onto the tangent space $\mathscr{T}_q$, i.e., we can define $\mathbf{x}_q \cdot \mathbf{y}_q$ for $\mathbf{x}_q$, $\mathbf{y}_q$ in $\mathscr{T}_q$ by

$$\mathbf{x}_q \cdot \mathbf{y}_q = (\mathbf{x}_q \circ \mathbf{J}_q) \cdot (\mathbf{y}_q \circ \mathbf{J}_q)\,. \tag{3.12}$$

It follows from (3.12) that the unit tensors $\mathbf{I}_{\mathscr{H}}(q) \in \mathscr{H}_q \otimes \mathscr{H}_q$ and $\mathbf{I}_{\mathscr{T}}(q) \in \mathscr{T}_q \otimes \mathscr{T}_q$, defined by the equations $\mathbf{I}_{\mathscr{H}}(q) \cdot \mathbf{u}_q = \mathbf{u}_q$ and $\mathbf{I}_{\mathscr{T}}(q) \cdot \mathbf{x}_q = \mathbf{x}_q$ for $\mathbf{u}_q \in \mathscr{H}_q$ and $\mathbf{x}_q \in \mathscr{T}_q$, are related by

$$\mathbf{I}_{\mathscr{H}} = \tilde{\mathbf{J}} \circ \mathbf{I}_{\mathscr{T}} \circ \mathbf{J}\,. \tag{3.13}$$

Thus, even though we started with no assumption of any metric structure on $\mathscr{M}$, the local Poincaré gauge theory leads to a way of uniquely imposing a metric structure on $\mathscr{M}$.

Once we know how to map vectors with $\mathbf{J}$, the mapping of tensors is straightforward. For notational convenience in subsequent calculations, we define

$(\circ\mathbf{J}(q))_1, (\circ\mathbf{J}(q))_2, (\circ\mathbf{J}(q))_3$, etc., as linear maps acting on a tensor to the left as illustrated in the following special case:

$$(\mathbf{x}_q \otimes \mathbf{y}_q \otimes \mathbf{z}_q)(\circ\mathbf{J}(q))_1 = \mathbf{x}_q \otimes \mathbf{y}_q \otimes (\mathbf{z}_q \circ \mathbf{J}(q)) = (\mathbf{x}_q \otimes \mathbf{y}_q \otimes \mathbf{z}_q) \circ \mathbf{J}(q)\,, \tag{3.14a}$$

$$(\mathbf{x}_q \otimes \mathbf{y}_q \otimes \mathbf{z}_q)(\circ\mathbf{J}(q))_2 = \mathbf{x}_q \otimes (\mathbf{y}_q \circ \mathbf{J}(q)) \otimes \mathbf{z}_q\,, \tag{3.14b}$$

$$(\mathbf{x}_q \otimes \mathbf{y}_q \otimes \mathbf{z}_q)(\circ\mathbf{J}(q))_3 = (\mathbf{x}_q \circ \mathbf{J}(q)) \otimes \mathbf{y}_q \otimes \mathbf{z}_q\,, \tag{3.14c}$$

where $\mathbf{x}_q, \mathbf{y}_q, \mathbf{z}_q \in \mathscr{T}_q$. We also define $(\mathbf{J}(q)\circ)_1$, $(\mathbf{J}(q)\circ)_2$, $(\mathbf{J}(q)\circ)_3$, etc., as linear maps acting on a tensor to the right as illustrated in the example:

$$(\mathbf{J}(q)\circ)_1(\mathbf{l}'_q \otimes \mathbf{m}'_q \otimes \mathbf{n}'_q) = (\mathbf{J}(q) \circ \mathbf{l}'_q) \otimes \mathbf{m}'_q \otimes \mathbf{n}'_q = \mathbf{J}(q) \circ (\mathbf{l}'_q \otimes \mathbf{m}'_q \otimes \mathbf{n}'_q)\,, \tag{3.15a}$$

$$(\mathbf{J}(q)\circ)_2(\mathbf{l}'_q \otimes \mathbf{m}'_q \otimes \mathbf{n}'_q) = \mathbf{l}'_q \otimes (\mathbf{J}(q) \circ \mathbf{m}'_q) \otimes \mathbf{n}'_q\,, \tag{3.15b}$$

$$(\mathbf{J}(q)\circ)_3(\mathbf{l}'_q \otimes \mathbf{m}'_q \otimes \mathbf{n}'_q) = \mathbf{l}'_q \otimes \mathbf{m}'_q \otimes (\mathbf{J}(q) \circ \mathbf{n}'_q)\,, \tag{3.15c}$$

where

$$\mathbf{l}'_q, \mathbf{m}'_q, \mathbf{n}'_q \in \mathscr{H}'_q = \{\mathbf{g}'_q \mid \mathbf{g}'_q \in \mathscr{E}'_q, \mathbf{g}'_q \circ \mathbf{k}(q) = 0\}.$$

Mapping of connections with $\mathbf{J}$: Given the nonsingular $\mathscr{T}'_q \otimes \mathscr{H}_q$ valued tensor field $\mathbf{J}$, it follows that there exists a $\mathscr{H}'_q \otimes \mathscr{T}_q$ valued tensor field $\mathbf{F}$ which is the inverse of $\mathbf{J}$ in the sense that $\mathbf{z} \circ \mathbf{J} \circ \mathbf{F} = \mathbf{z}$ for every $\mathscr{T}_q$ valued vector field $\mathbf{z}$ and $\mathbf{h} \circ \mathbf{F} \circ \mathbf{J} = \mathbf{h}$ for every $\mathscr{H}_q$ valued vector field $\mathbf{h}$. Note that $\mathbf{J}$ maps each $\mathscr{T}_q$ valued vector field $\mathbf{z}$ onto a $\mathscr{H}_q$ valued vector field $\mathbf{z} \circ \mathbf{J}$, and that $\mathbf{F}$ maps each $\mathscr{H}_q$ valued vector field $\mathbf{h}$ onto a $\mathscr{T}_q$ valued vector field $\mathbf{h} \circ \mathbf{F}$. We also have

$$\mathbf{k} \circ \mathbf{F} = 0\,.$$

Let D_X be a connection on $\mathscr{E}_q$ valued vector fields such that parallel transport under this connection preserves the structure of the spaces $\mathscr{E}_q$.

Theorem 2: If $\mathbf{h}$ is an $\mathscr{H}_q$-valued vector field, then $D_X\mathbf{h}$ is also an $\mathscr{H}_q$-valued vector field.

Proof: From

$$X(\mathbf{f}' \circ \mathbf{h}) = (D_X\mathbf{f}') \circ \mathbf{h} + \mathbf{f}' \circ (D_X\mathbf{h}) \tag{3.16}$$

and Eq. (3.4), as well as the fact that $\mathbf{f}' \circ \mathbf{h} = 0$ for all q (since $\mathbf{h}(q)$ is in $\mathscr{H}_q$), we have

$$\mathbf{f}' \circ (D_X\mathbf{h}) = 0\,. \tag{3.17}$$

Q.E.D.

Hence $D_X \mathbf{h}$ is a $\mathscr{H}_q$-valued vector field.

This theorem implies that a connection D_X on $\mathscr{E}_q$-valued vector fields is also a connection on $\mathscr{H}_q$-valued vector fields when we restrict it to act on $\mathscr{H}_q$ valued vector fields.

Each connection on $\mathscr{E}_q$-valued vector fields can be mapped onto a connection ∇_X on the tangent bundle $\mathscr{T}(\mathscr{M})$ by means of the equation

$$\nabla_X \mathbf{z} = [D_X(\mathbf{z}\circ\mathbf{J})]\circ\mathbf{F} \tag{3.18}$$

for an arbitrary $\mathscr{T}_q$ valued vector field $\mathbf{z}$. This map $D_X \to \nabla_X$ is one to one if we restrict the domain of D_X to $\mathscr{H}_q$-valued vector fields. Equation (3.18) can be written in the equivalent form

$$(\nabla_X \mathbf{z})\circ\mathbf{J} = D_X(\mathbf{z}\circ\mathbf{J}) . \tag{3.19}$$

Theorem 3: A connection D_X compatible with the inner product in $\mathscr{H}(\mathscr{M})$ induces a connection ∇_X compatible with the inner product in $\mathscr{T}(\mathscr{M})$.

Proof: Let $\mathbf{y}$ and $\mathbf{z}$ be arbitrary $\mathscr{T}_q$-valued vector fields. Making use of (3.12) and (3.19) we get

$$\begin{aligned} X(\mathbf{y}\cdot\mathbf{z}) &= X[(\mathbf{y}\circ\mathbf{J})\cdot(\mathbf{z}\circ\mathbf{J})] = [D_X(\mathbf{y}\circ\mathbf{J})]\cdot(\mathbf{z}\circ\mathbf{J}) \\ &\quad + (\mathbf{y}\circ\mathbf{J})\cdot[D_X(\mathbf{z}\circ\mathbf{J})] \\ &= [(\nabla_X\mathbf{y})\circ\mathbf{J}]\cdot(\mathbf{z}\circ\mathbf{J}) + (\mathbf{y}\circ\mathbf{J})\cdot[(\nabla_X\mathbf{z})\circ\mathbf{J}] \\ &= (\nabla_X\mathbf{y})\cdot\mathbf{z} + \mathbf{y}\cdot(\nabla_X\mathbf{z}) . \end{aligned} \tag{3.20}$$

Q.E.D.

Curvature tensor: We define the curvature tensor $\mathbf{R}_{\mathscr{E}(\mathscr{M})}$ with values in $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{E}_q \otimes \mathscr{E}'_q$ by

$$\mathbf{xy} \overset{\circ}{\circ} \mathbf{R}_{\mathscr{E}(\mathscr{M})}\circ\mathbf{v} = (D_X D_Y - D_Y D_X - D_{[X,Y]})\mathbf{v} . \tag{3.21}$$

Making use of (2.7) we can write

$$\begin{aligned} \mathbf{xy} \overset{\circ}{\circ} \mathbf{R}_{\mathscr{E}(\mathscr{M})}\circ\mathbf{v} &= (D_X D_Y - D_Y D_X - D_{[X,Y]})\mathbf{h} \\ &\quad + \alpha(D_X D_Y - D_Y D_X - D_{[X,Y]})\mathbf{k} , \end{aligned} \tag{3.22}$$

where $\alpha \in \Gamma(\mathscr{M}, \mathbb{R})$ is a smooth scalar field and $\mathbf{h}(q) \in \mathscr{H}_q$. It follows that $\mathbf{R}_{\mathscr{E}(\mathscr{M})}(q) \in \mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{H}_q \otimes \mathscr{E}'_q$. Note now that because of (3.7) and (3.19),

$$\begin{aligned} \mathbf{xy} &\overset{\circ}{\circ} \mathbf{R}_{\mathscr{E}(\mathscr{M})}\circ\mathbf{k} \\ &= D_X(\mathbf{y}\circ\mathbf{D}\otimes\mathbf{k}) - D_Y(\mathbf{x}\circ\mathbf{D}\otimes\mathbf{k}) - [\mathbf{x},\mathbf{y}]\circ\mathbf{D}\otimes\mathbf{k} \\ &= D_X(\mathbf{y}\circ\mathbf{J}) - D_Y(\mathbf{x}\circ\mathbf{J}) - [\mathbf{x},\mathbf{y}]\circ\mathbf{J} \\ &= (\nabla_X\mathbf{y} - \nabla_Y\mathbf{x} - [\mathbf{x},\mathbf{y}])\circ\mathbf{J} = \mathbf{xy} \overset{\circ}{\circ} \mathbf{T}_{\mathscr{T}(\mathscr{M})}\circ\mathbf{J} , \end{aligned} \tag{3.23}$$

i.e.,

$$\mathbf{R}_{\mathscr{E}(\mathscr{M})}\circ\mathbf{k} = \mathbf{T}_{\mathscr{T}(\mathscr{M})}\circ\mathbf{J} \tag{3.24}$$

where $\mathbf{T}_{\mathscr{T}(\mathscr{M})}$ is the torsion tensor on the tangent bundle with values $\mathbf{T}_{\mathscr{T}(\mathscr{M})}(q) \in \mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{T}_q$.

In addition, from (3.11) and (3.19), we have

$$\begin{aligned} \mathbf{xy} \overset{\circ}{\circ} \mathbf{R}_{\mathscr{E}(\mathscr{M})}\circ\mathbf{h} &= \mathbf{xy} \overset{\circ}{\circ} \mathbf{R}_{\mathscr{E}(\mathscr{M})}\circ(\mathbf{z}\circ\mathbf{J}) \\ &= (D_X D_Y - D_Y D_X - D_{[X,Y]})(\mathbf{z}\circ\mathbf{J}) \\ &= [(\nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]})\mathbf{z}]\circ\mathbf{J} \\ &= \mathbf{xy} \overset{\circ}{\circ} (\mathbf{R}_{\mathscr{T}(\mathscr{M})}\circ\mathbf{z})\circ\mathbf{J} , \end{aligned} \tag{3.25}$$

i.e.,

$$\mathbf{R}_{\mathscr{E}(\mathscr{M})}\circ\tilde{\mathbf{J}} = \mathbf{R}_{\mathscr{T}(\mathscr{M})}(\circ\mathbf{J})_2 , \tag{3.26}$$

where $\mathbf{R}_{\mathscr{T}(\mathscr{M})}$ is the curvature tensor on the tangent bundle with values in $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{T}_q \otimes \mathscr{T}'_q$. Combining (3.24) and (3.26) we get

$$\mathbf{R}_{\mathscr{E}(\mathscr{M})} = \mathbf{R}_{\mathscr{T}(\mathscr{M})}(\circ\mathbf{J})_2(\circ\tilde{\mathbf{F}})_1 + \mathbf{T}_{\mathscr{T}(\mathscr{M})}\circ\mathbf{J}\otimes\mathbf{f}' , \tag{3.27}$$

which relates the curvature tensor on the fiber bundle to the curvature and torsion tensors on the tangent bundle.

Alternatively, Eq. (3.27) can be viewed as a unique way of decomposing $\mathbf{R}_{\mathscr{E}(\mathscr{M})}$ into

$$\mathbf{R}_{\mathscr{E}(\mathscr{M})} = \mathbf{R}_{\mathscr{H}(\mathscr{M})} + \mathbf{T}_{\mathscr{H}(\mathscr{M})} \otimes \mathbf{f}' , \tag{3.28}$$

where

$$\mathbf{T}_{\mathscr{H}(\mathscr{M})} \equiv \mathbf{R}_{\mathscr{E}(\mathscr{M})}\circ\mathbf{k} = \mathbf{T}_{\mathscr{T}(\mathscr{M})}\circ\mathbf{J} \tag{3.29}$$

and

$$\mathbf{R}_{\mathscr{H}(\mathscr{M})} \equiv \mathbf{R}_{\mathscr{E}(\mathscr{M})} - \mathbf{R}_{\mathscr{E}(\mathscr{M})}\circ(\mathbf{k}\otimes\mathbf{f}') = \mathbf{R}_{\mathscr{T}(\mathscr{M})}(\circ\mathbf{J})_2(\circ\tilde{\mathbf{F}})_1 \tag{3.30}$$

with

$$\mathbf{R}_{\mathscr{H}(\mathscr{M})}(q) \in \mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{H}_q \otimes \mathscr{H}'_q$$

and

$$\mathbf{T}_{\mathscr{H}(\mathscr{M})}(q) \in \mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{H}_q .$$

Performing contractions on $\mathbf{R}_{\mathscr{T}}$ gives the Ricci tensor $\mathrm{R}_{\mathscr{T}} \equiv C(13)\mathbf{R}_{\mathscr{T}} \in \mathscr{T}'_q \otimes \mathscr{T}'_q$ at q, and the curvature invariant $(\mathrm{R}_{\mathscr{T}})_s \equiv C(13;24)(\mathbf{R}_{\mathscr{T}}\circ\mathbf{I}_{\mathscr{T}}) \in \mathbb{R}$ at q, where $C(\ )$ denotes the contraction on the designated files in the parentheses. From $\mathbf{R}_{\mathscr{H}}$ we may also get a second order tensor $\mathrm{R}_{\mathscr{H}} \equiv C(13)\ (\mathbf{F}\circ\mathbf{R}_{\mathscr{H}}) \in \mathscr{T}'_q \otimes \mathscr{H}'_q$ at q, and a scalar $(\mathrm{R}_{\mathscr{H}})_s \equiv C(13;24)[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{R}_{\mathscr{H}}\circ\mathbf{I}_{\mathscr{H}}] \in \mathscr{R}$ at q. These quantities are related as follows:

$$\mathrm{R}_{\mathscr{T}} = \mathrm{R}_{\mathscr{H}}\circ\tilde{\mathbf{J}} , \tag{3.31}$$

$$(\mathrm{R}_{\mathscr{T}})_s = (\mathrm{R}_{\mathscr{H}})_s . \tag{3.32}$$

Local Poincaré transformations: Given a smooth cross section $\mathbf{v} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$, the structural group of our vector bundle associates a local Poincaré group $\mathscr{P}_q$ at each point on the base manifold such that the action of the group on $\mathbf{v}(q) \in \mathscr{E}_q$ is a linear transformation defined by Eqs. (2.1), (2.2), and (2.10) for each $q \in \mathscr{M}$.

A connection in $\Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$ transforms under the action of the group according to

$$\mathbf{P}\circ(D_X\mathbf{v}) = D_X{}^{(P)}(\mathbf{P}\circ\mathbf{v}) , \tag{3.33}$$

or, equivalently,

$$D_X{}^{(P)}\mathbf{v} = \mathbf{P}\circ D_X(\mathbf{P}^{-1}\circ\mathbf{v}) . \tag{3.34}$$

Moreover, by (2.9)

$$P^{-1}\circ\mathbf{v} = \mathbf{P}^{-1}\circ(\alpha\mathbf{k} + \mathbf{h}) = \alpha\mathbf{k} + \mathbf{L}^{-1}\circ(\mathbf{h} - \alpha\mathbf{t}) . \tag{3.35}$$

Hence

$$\begin{aligned} D_X{}^{(P)}\mathbf{v} &= \mathbf{P}\circ D_X[\alpha\mathbf{k} + \mathbf{L}^{-1}\circ(\mathbf{h} - \alpha\mathbf{t})] \\ &= \mathbf{L}\circ D_X[\alpha\mathbf{k} + \mathbf{L}^{-1}\circ(\mathbf{h} - \alpha\mathbf{t})] + (X\alpha)\mathbf{t} \\ &= \mathbf{L}\circ D_X[\mathbf{L}^{-1}\circ(\alpha\mathbf{k} + \mathbf{h} - \alpha\mathbf{t})] + (X\alpha)\mathbf{t} \\ &= \mathbf{L}\circ D_X(\mathbf{L}^{-1}\circ\mathbf{v}) - \alpha\mathbf{L}\circ D_X(\mathbf{L}^{-1}\circ\mathbf{t}) . \end{aligned} \tag{3.36}$$

Making use of (3.33) and (3.36), we have

$$\begin{aligned} \mathbf{P}\circ D_X\mathbf{k} &= D_X{}^{(P)}(\mathbf{P}\circ\mathbf{k}) = D_X{}^{(P)}(\mathbf{k} + \mathbf{t}) \\ &= \mathbf{L}\circ D_X[\mathbf{L}^{-1}\circ(\mathbf{k} + \mathbf{t})] - \mathbf{L}\circ D_X(\mathbf{L}^{-1}\circ\mathbf{t}) \\ &= (D_X\mathbf{k})\circ\tilde{\mathbf{L}} = (\mathbf{x}\circ\mathbf{J})\circ\tilde{\mathbf{L}} , \end{aligned}$$

i.e.,

$$\mathbf{P}\circ(\mathbf{x}\circ\mathbf{J}) = \mathbf{x}\circ\mathbf{J}\circ\tilde{\mathbf{L}} . \tag{3.37}$$

The action $r(\mathscr{P})$ on the tangent space $\mathscr{T}_q$ induced by a local Poincaré transformation on $\mathscr{E}_q$ is given by

$$r(P)\mathbf{y} \equiv [\mathbf{P}\circ(\mathbf{y}\circ\mathbf{J})]\circ\mathbf{F} = \mathbf{y}\circ\mathbf{J}\circ\tilde{\mathbf{L}}\circ\mathbf{F} \quad (3.38)$$

for $\mathbf{y}\in\mathscr{T}_q$. As a consequence of this equation and equation (3.12), we have

$$\begin{aligned}[r(P)\mathbf{x}]\cdot[r(P)\mathbf{y}] &= [(r(P)\mathbf{x})\circ\mathbf{J}]\cdot[(r(P)\mathbf{y})\circ\mathbf{J}]\\ &= (\mathbf{x}\circ\mathbf{J}\circ\tilde{\mathbf{L}})\cdot(\mathbf{y}\circ\mathbf{J}\circ\tilde{\mathbf{L}}) = [\mathbf{L}\circ(\mathbf{x}\circ\mathbf{J})]\cdot[\mathbf{L}\circ(\mathbf{y}\circ\mathbf{J})]\\ &= (\mathbf{x}\circ\mathbf{J})\cdot(\mathbf{y}\circ\mathbf{J}) = \mathbf{x}\cdot\mathbf{y}\,, \end{aligned} \quad (3.39)$$

i.e., the inner product in $\mathscr{T}_q$ is invariant under $r(P)$.

Free Lagrangians for the gauge fields: We can set up free Lagrangian densities $\mathscr{L}_{\mathscr{E}(\mathscr{M})}(\mathbf{R}_{\mathscr{E}(\mathscr{M})})$ for the gauge fields (i.e., the connection D_X) as scalar functionals of $\mathbf{R}_{\mathscr{E}(\mathscr{M})}$ which by construction will be locally Poincaré-invariant. Note that by virtue of Eq. (3.27) these Lagrangian densities will determine the allowed functional form in terms of the quantities $\mathbf{R}_{\mathscr{T}(\mathscr{M})}$ and $\mathbf{T}_{\mathscr{T}(\mathscr{M})}$ in the tangent bundle, so that a permissible Langrangian density $\mathscr{L}_{\mathscr{T}(\mathscr{M})}$ in $\mathscr{T}(\mathscr{M})$ will be given by

$$\begin{aligned}&\mathscr{L}_{\mathscr{T}(\mathscr{M})}(\mathbf{R}_{\mathscr{T}(\mathscr{M})},\mathbf{T}_{\mathscr{T}(\mathscr{M})})\\ &\quad = \mathscr{L}_{\mathscr{E}(\mathscr{M})}(\mathbf{R}_{\mathscr{T}(\mathscr{M})}(\circ\mathbf{J})_2(\circ\tilde{\mathbf{F}})_1 + \mathbf{T}_{\mathscr{T}(\mathscr{M})}\circ\mathbf{J}\otimes\mathbf{f}')\,. \end{aligned}\quad (3.40)$$

Volume element: The antisymmetric tensor element of volume on $\mathscr{M}$, denoted by $d\,\mathbf{\Omega}(q)\in\mathscr{T}_q\wedge\mathscr{T}_q\wedge\mathscr{T}_q\wedge\mathscr{T}_q$, is related to appropriately oriented coordinates x^0, x^1, x^2, x^3 on $\mathscr{M}$ by

$$d\,\mathbf{\Omega} = \mathbf{a}_0\wedge\mathbf{a}_1\wedge\mathbf{a}_2\wedge\mathbf{a}_3 dx^0dx^1dx^2dx^3\,, \quad (3.41)$$

where $a_\mu(q)\equiv\partial/\partial x^\mu$ is a natural basis for $\mathscr{T}_q$.

Using the inner product in $\mathscr{T}_q$, define the antisymmetric tensor field $\mathbf{\Lambda}$ with $\mathbf{\Lambda}(q)\in\mathscr{T}_q\wedge\mathscr{T}_q\wedge\mathscr{T}_q\wedge\mathscr{T}_q$ such that

$$\mathbf{\Lambda}::\mathbf{\Lambda} = -4!. \quad (3.42)$$

This defines $\mathbf{\Lambda}$ up to a sign. A scalar element of volume on $\mathscr{M}$ can be defined as

$$dV = d\Omega::\mathbf{\Lambda}\,. \quad (3.43)$$

Now using the inner product in $\mathscr{H}_q$, define the antisymmetric tensor field $\mathbf{M}$ with $\mathbf{M}(q)\in\mathscr{H}_q\wedge\mathscr{H}_q\wedge\mathscr{H}_q\wedge\mathscr{H}_q$ such that

$$\mathbf{M}::\mathbf{M} = -4!. \quad (3.44)$$

This defines $\mathbf{M}$ up to a sign. Another scalar element of volume on $\mathscr{M}$ can then be defined as

$$d\tau = [d\,\mathbf{\Omega}(\circ\mathbf{J})_4(\circ\mathbf{J})_3(\circ\mathbf{J})_2(\circ\mathbf{J})_1]::\mathbf{M}\,. \quad (3.45)$$

However $\mathbf{M}$ and $\mathbf{\Lambda}$ transform into each other according to

$$\mathbf{\Lambda} = \mathbf{M}(\circ\mathbf{F})_4(\circ\mathbf{F})_3(\circ\mathbf{F})_2(\circ\mathbf{F})_1 \quad (3.46)$$

and if the signs of $\mathbf{M}$ and $\mathbf{\Lambda}$ are appropriately chosen, it follows that

$$d\tau = dV\,. \quad (3.47)$$

Hence, the action for the gravitational field will be of the general form

$$\begin{aligned}L_0 &= \int \mathscr{L}_{\mathscr{E}(\mathscr{M})}(\mathbf{R}_{\mathscr{E}(\mathscr{M})})d\tau\\ &= \int \mathscr{L}_{\mathscr{T}(\mathscr{M})}(\mathbf{R}_{\mathscr{T}(\mathscr{M})},\mathbf{T}_{\mathscr{T}(\mathscr{M})})dV\,, \end{aligned}\quad (3.48)$$

where $\mathscr{L}_{\mathscr{E}(\mathscr{M})}$ and $\mathscr{L}_{\mathscr{T}(\mathscr{M})}$ are related by Eq. (3.40).

It is not the purpose of this paper to discuss in more detail possible explicit forms for the gravitational Lagrangian. Instead, we now turn to the remark made earlier in this section regarding the independence of our theory on the choice of the "origin" vector $\mathbf{k}$. To prove this statement, consider a different $\mathbf{k}$, obtained by the action of a translation on $\mathbf{k}$. From (2.5) and (2.6) we have

$$\mathbf{T}\circ\mathbf{k} = \mathbf{k} + \mathbf{t}\,,$$

$$\mathbf{T}\circ\mathbf{h} = \mathbf{h}, \quad \text{for } \mathbf{h}(q)\in\mathscr{H}_q\,.$$

The consequent change in D_X follows from (3.36) and is given by

$$D_X^{(T)}\mathbf{h} = D_X\mathbf{h}\,, \quad (3.49)$$

and

$$D_X^{(T)}\mathbf{k} = D_X(\mathbf{k} - \mathbf{t})\,, \quad (3.50)$$

or, making use of (2.6),

$$D_X^{(T)}(\mathbf{T}\circ\mathbf{k}) = D_X(\mathbf{k})\,. \quad (3.51)$$

Thus we see from (2.5) and (3.49) that a change in $\mathbf{k}$ leaves vectors and connections on $\mathscr{H}_q$ invariant, while from (3.51) we have

$$D^{(T)}\otimes(\mathbf{T}\circ\mathbf{k}) = \mathbf{D}\otimes\mathbf{k} = \mathbf{J}\,, \quad (3.52)$$

i.e., a change in $\mathbf{k}$ can be compensated for by a corresponding change in D_X such that the tensor $\mathbf{J}$, used for mapping vectors, tensors and connections into the tangent bundle, is unaffected. We are therefore led to a completely equivalent theory.

Variational Principles: Given a matter Langrangian in which the gauge fields are minimally coupled to the particle fields ψ^A by means of the covariant derivative operator D_X, and given a "free" Lagrangian constructed from these gauge fields according to the procedure outlined above, we can obtain field equations and conservation laws by means of a variational principle applied to the matter and gauge fields. We will now discuss the general features of this variational procedure for the Lagrangian in the context of the formalism so far developed.

The fundamental (gauge) quantity to be varied is D_X. Since any two linear connections may differ only by a linear transformation, we can write

$$\delta D_X = [(\delta\,\mathbf{B}_X)\circ]\,, \quad (3.53a)$$

or

$$(\delta D_X)\mathbf{v} = (\delta\,\mathbf{B}_X)\circ\mathbf{v} \quad (3.53b)$$

for each $\mathscr{E}_q$ valued vector field $\mathbf{v}$, where $\delta\,\mathbf{B}_X(q)\in\mathscr{E}_q\otimes\mathscr{E}'_q$ is linear in X. It also follows, in particular, that

$$\mathbf{x}\circ\delta\,\mathbf{J} = \delta\,(\mathbf{x}\circ\mathbf{J}) = \delta\,(D_X\mathbf{k}) = (\delta D_X)\mathbf{k} = \delta\,\mathbf{B}_X\circ\mathbf{k}\,, \quad (3.54)$$

or

$$\delta\,\mathbf{J} = (\delta\,\mathbf{B})\circ\mathbf{k}\,, \quad (3.55)$$

where

$$\delta\,\mathbf{B}(q)\in\mathscr{T}'_q\otimes\mathscr{E}_q\otimes\mathscr{E}'_q$$

is defined by

$$\delta \mathbf{B}_X = \mathbf{x} \circ \delta \mathbf{B} . \qquad (3.56)$$

Note that virtue of Eqs. (3.5) and (3.6) it follows that $\delta \mathbf{B}(q) \in \mathscr{T}'_q \otimes \mathscr{H}_q \otimes \mathscr{E}'_q$ and that $\delta \mathbf{B} \circ \mathbf{I}_{\mathscr{H}(\mathscr{M})}$ is antisymmetric in the second and third files. Consequently $\delta \mathbf{B}$ can be written in general as

$$\delta \mathbf{B} = \delta \mathbf{N} \otimes \mathbf{f}' + \delta \mathbf{W} . \qquad (3.57)$$

Here $\delta \mathbf{N}(q) \in \mathscr{T}'_q \otimes \mathscr{H}_q$ is completely arbitrary and $\delta \mathbf{W}(q) \in \mathscr{T}'_q \otimes \mathscr{H}_q \otimes \mathscr{H}'_q$ is arbitrary except that $\delta \mathbf{W} \circ \mathbf{I}_{\mathscr{H}(\mathscr{M})}$ is antisymmetric in the second and third files. It readily follows from Eq. (3.57) that

$$\delta \mathbf{J} = \delta \mathbf{N} . \qquad (3.58)$$

Making use of these results, we can now evaluate the corresponding variations in the tangent bundle induced by $\delta \mathbf{N}$ and $\delta \mathbf{W}$. By varying Eq. (3.12) we get, after some straightforward calculations,

$$\delta \mathbf{I}_{\mathscr{T}(\mathscr{M})} = -\mathbf{I}_{\mathscr{T}(\mathscr{M})} \circ \delta \mathbf{N} \circ \mathbf{F} - (\mathbf{I}_{\mathscr{T}(\mathscr{M})} \circ \delta \mathbf{N} \circ \mathbf{F})^{\sim} , \qquad (3.59)$$

which is symmetric but otherwise completely arbitrary. The variation of $\mathbf{T}_{\mathscr{T}(\mathscr{M})}$ is obtained by making use of Eqs. (3.21) and (3.24), and can be expressed in general as

$$\delta \mathbf{T}_{\mathscr{T}(\mathscr{M})} = \delta_1 \mathbf{T}_{\mathscr{T}(\mathscr{M})} + \delta_2 \mathbf{T}_{\mathscr{T}(\mathscr{M})} , \qquad (3.60)$$

where $\delta_1 \mathbf{T}_{\mathscr{T}(\mathscr{M})}$ is dependent only on $\delta \mathbf{N}$, and

$$\delta_2 \mathbf{T}_{\mathscr{T}(\mathscr{M})} = \{ [1 - (12)](23)[\delta \mathbf{W} \circ \tilde{\mathbf{J}}] \} \circ \mathbf{F} . \qquad (3.61)$$

[The symbol (kl) denotes the permutation of the k th and l th files.] Both $\delta_1 \mathbf{T}_{\mathscr{T}(\mathscr{M})}$ and $\delta_2 \mathbf{T}_{\mathscr{T}(\mathscr{M})}$ are antisymmetric in the first two files. Furthermore, by letting

$$\delta \mathbf{W} = \tfrac{1}{2}(\mathbf{J}\circ)_1 (\mathbf{I}_{\mathscr{H}(\mathscr{M})}\circ)_2 \{ [(23) - (13) - 1] \delta \mathbf{Q} \}, \qquad (3.62)$$

where $\delta \mathbf{Q}(q) \in \mathscr{H}'_q \otimes \mathscr{H}'_q \otimes \mathscr{H}'_q$ is completely arbitrary except for antisymmetry in the first two files, it can be shown that $\delta_2 \mathbf{T}_{\mathscr{T}(\mathscr{M})}$ is also arbitrary except for the previously mentioned antisymmetry property. Hence our theory leads to variations $\delta \mathbf{I}_{\mathscr{T}(\mathscr{M})}$ and $\delta \mathbf{T}_{\mathscr{T}(\mathscr{M})}$ which are independent of each other and arbitrary except for the conditions that $\delta \mathbf{I}_{\mathscr{T}(\mathscr{M})}$ is symmetric and $\delta \mathbf{T}_{\mathscr{T}(\mathscr{M})}$ is antisymmetric in the first two files.

Finally note that in addition to the variations induced in $\mathbf{I}_{\mathscr{T}(\mathscr{M})}$ and $\mathbf{T}_{\mathscr{T}(\mathscr{M})}$, the inner product and connection ∇_X on $\mathscr{T}(\mathscr{M})$ will have to vary in such a way that the inner product will continue to have $\mathbf{I}_{\mathscr{T}(\mathscr{M})}$ as a unit tensor and ∇_X will remain compatible with the inner product. All these entities, on the other hand, are the ones varied in the conventional approaches to gravitational theories.

IV. SUMMARY AND CONCLUSIONS

The central idea in this paper is the formulation of a theory of gravitation in which the Poincaré group is treated as an internal gauge group. However, the theory differs from internal group gauge theories in two significant aspects.

The first difference is that no metric structure or connection on the tangent bundle is assumed. In typical gauge theories the metric structure of the tangent bundle is given *a priori* together with a connection compatible with the metric.

The second way in which the procedure differs from the typical internal gauge theory is the inclusion of an "origin" vector field, which is interpreted as the points at which the fibers are tied to the manifold. The covariant gradient of this field gives a tensor field $\mathbf{J} = \mathbf{D} \otimes \mathbf{k}$, which can be utilized as a map by means of which structure can be mapped in a natural way onto the tangent bundle inducing in it a metric and connection. We have shown that the selection of the origin vector field $\mathbf{k}$ imposes no special restriction on the theory since a change in the choice of this field can be compensated by a corresponding change in the connection D_X such that a completely equivalent theory is obtained.

We have obtained the possible functional form of the Lagrangian permitted by the theory and by means of maps made possible by the tensor field $\mathbf{J}$, we have shown that the theory of gravitation obtained by treating the Poincaré group as an internal gauge group encompasses General Relativity and the Einstein-Cartan Theory as particular cases, as well as other gravitational theories with torsion which have been recently proposed.

ACKNOWLEDGMENT

We wish to thank S. Hojman and M.P. Ryan for fruitful discussions on the subject.

[1]S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967); A. Salam, in *Elementary Particle Theory: Relativistic Groups and Analyticity* (Nobel Simposium No. 8), edited by N. Svartholm (Almqvist and Wiksell, Stokholm, 1968), p. 367; J. Wess, B. Zumino, Nucl. Phys. B **70**, 39 (1974); R. Arnowitt, P. Nath, B. Zumino, Phys. Lett. B **56**, 81 (1975); P. Nath, R. Arnowitt, Phys. Lett. B **56**, 177 (1975); D.Z. Freedman, P. Van Nieuwenhuizen, S. Ferrara, Phys. Rev. D. **13**, 3214 (1976); S. Deser, B. Zumino, Phys. Lett. B **62**, 335 (1976).
[2]C.N. Yang, R.L. Mills, Phys. Rev. **96**, 191 (1954).
[3]R. Utimaya, Phys. Rev. **101**, 1597 (1956).
[4]F.W. Hehl, P. Van Der Heyde, G.D. Kerlick, J.M. Nester, Rev. Mod. Phys. **48**, 393 (1976).
[5]J.G. Taylor, Phys. Rev. D **18**, 3544 (1978).
[6]T.W.B. Kibble, J. Math. Phys. **2**, 212 (1961).
[7]K. Hayashi, A. Bregman, Ann. Phys. **75**, 562 (1973).
[8]H.P. Durr, Nuovo Cimento A **4**, 187 (1971).
[9]R. Utimaya, T. Fukuyama, Prog. Theor. Phys. **45**, 612 (1971).
[10]W. Drechsler, M.E. Mayer, *Fiber Bundle Techniques in Gauge Theories* (Lecture Notes in Physics Vol. 67, Springer, Berlin, 1977).
[11]Y. Choquet-Bruhat, C. De Witt, M. Dillard, *Analysis, Manifolds and Physics* (North-Holland Publishing Co., Amsterdam, 1977).

Gauss–Bonnet and Bianchi identities in Riemann–Cartan type gravitational theories

H. T. Nieh
Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11794

(Received 20 September 1979; accepted for publication 15 February 1980)

We present elementary derivations of the Gauss–Bonnet type and Bianchi type identities for Riemann–Cartan geometry. The identities are derived directly in terms of the vierbein field $e^a{}_\mu$ and the spin-connection field $\mathscr{V}^{ab}{}_\mu$, in the spirit of gauge theory, and are suitable for discussing Riemann–Cartan type gravitational theories.

1. INTRODUCTION

In the formulation of gravitational theories for half-integer spin fields, it is natural to introduce the vierbein field[1] $e^a{}_\mu$, or its generalization. There is a whole class of gravitational theories in which the vierbein field $e^a{}_\mu$ and the "Lorentz-spin connection" field $\mathscr{V}^{ab}{}_\mu$ are taken to be *independent* field variables.[2] The underlying geometry for such theories is the Riemann–Cartan geometry, in the sense that it has a metric tensor and torsion. We shall call this class of gravitational theories by the generic name of Riemann–Cartan gravitational theory.

The Bianchi identities for Riemann–Cartan geometry can be readily found in the literature.[3] However, the Gauss–Bonnet type identities in Riemann–Cartan geometry are not widely known; a simple and elementary derivation suitable for most physicists is simply lacking.

We shall present *simple* proofs of both the Gauss–Bonnet type and Bianchi identities for Riemann–Cartan geometry. These identities are derived directly in terms of the vierbein field $e^a{}_\mu$ and the spin-connection field $\mathscr{V}^{ab}{}_\mu$, *in the spirit of a gauge theory*, and are especially suitable for discussing the Riemann–Cartan type gravitational theories. The derivations are elementary and simple, and should be of pedagogic value.

In Sec. II, we first present the relationship between the field variables $e^a{}_\mu$ and $\mathscr{V}^{ab}{}_\mu$, on the one hand, and the geometrical quantities on the other. In Sec. III, we derive the Gauss–Bonnet type identities. The derivation is based on employing the Dirac matrices, and is very simple. In Sec. IV, the Binachi type identities are derived. The derivation is based on employing the de Sitter algebra.

II. RELATIONSHIP OF $e^a{}_\mu$ AND $\mathscr{V}^{ab}{}_\mu$ TO GEOMETRIC QUANTITIES[4]

The basic field variables in Riemann–Cartan type gravitational theories are the vierbein field $e^a{}_\mu$ and the Lorentz-spin connection field $\mathscr{V}^{ab}{}_\mu = -\mathscr{V}^{ba}{}_\mu$, where the Latin a, b are the anholonomic Lorentz indices and the Greek μ the holonomic coordinate index. The spin connection $\mathscr{V}^{ab}{}_\mu$ plays the role of the gauge potential for local Lorentz symmetry. With respect to the anholonomic Lorentz indices, covariant derivatives can be defined such as

$$\begin{aligned} \chi^a{}_{|\mu} &\equiv \chi^a{}_{,\mu} + \mathscr{V}^a{}_{b\mu}\chi^b, \\ \chi_{a|\mu} &\equiv \chi_{a,\mu} - \mathscr{V}^b{}_{a\mu}\chi_b, \end{aligned} \tag{1}$$

for the anholonomic Lorentz vectors χ^a and χ_a, where

$$\chi^a{}_{,\mu} = \frac{\partial \chi^a}{\partial x^\mu}.$$

The anholonomic Lorentz indices are raised and lowered by

$$\eta^{ab} = \eta_{ab} = (1, -1, -1, -1), \tag{2}$$

while the holonomic coordinate indices are raised and lowered by

$$g_{\mu\nu} = \eta_{ab} e^a{}_\mu e^b{}_\nu \tag{3}$$

and its inverse $g^{\mu\nu}$.

A connection for the holonomic coordinate indices has to be defined. The product

$$\chi_\mu = e^a{}_\mu \chi_a$$

transforms as an anholonomic scalar, and a holonomic vector. The product

$$\begin{aligned} e^a{}_\mu \chi_{a|\nu} &= e^a{}_\mu(\chi_{a,\nu} - \mathscr{V}^b{}_{a\nu}\chi_b) \\ &\equiv \chi_{\mu,\nu} - \Gamma^\lambda{}_{\mu\nu}\chi_\lambda \end{aligned} \tag{4}$$

should transform as an anholonomic scalar, and a holonomic tensor. The connection $\Gamma^\lambda{}_{\mu\nu}$ is thus identified to be

$$\Gamma^\lambda{}_{\mu\nu} = e_a{}^\lambda e^a{}_{\mu|\nu}, \tag{5}$$

where $e_a{}^\lambda$ is the inverse of $e^a{}_\lambda$, and

$$e^a{}_{\mu|\nu} = e^a{}_{\mu,\nu} + \mathscr{V}^a{}_{b\nu} e^b{}_\mu.$$

Using $\mathscr{V}^{ab}{}_\mu$ and $\Gamma^\lambda{}_{\mu\nu}$, we can define covariant derivatives with respect to both anholonomic and holonomic indices, such as

$$\begin{aligned} \chi_a{}^\lambda{}_{;\mu} &\equiv \chi_a{}^\lambda{}_{,\mu} - \mathscr{V}^b{}_{a\mu}\chi_b{}^\lambda + \Gamma^\lambda{}_{\nu\mu}\chi_a{}^\nu \\ &= \chi_a{}^\lambda{}_{|\mu} + \Gamma^\lambda{}_{\nu\mu}\chi_a{}^\nu; \\ \chi^a{}_{\nu;\mu} &\equiv \chi^a{}_{\nu,\mu} + \mathscr{V}^a{}_{b\mu}\chi^b{}_\nu - \Gamma^\lambda{}_{\nu\mu}\chi^a{}_\lambda \\ &= \chi^a{}_{\nu|\mu} - \Gamma^\lambda{}_{\nu\mu}\chi^a{}_\lambda. \end{aligned} \tag{6}$$

In can be easily verified that

$$e^a{}_{\mu;\nu} = 0, \quad e_a{}^\mu{}_{;\nu} = 0, \tag{7}$$

and consequently,

$$g^{\mu\nu}{}_{;\lambda} = 0, \quad g_{\mu\nu;\lambda} = 0. \tag{8}$$

 0022-2488/80/061439-03$1.00

The connection $\Gamma^{\lambda}{}_{\mu\nu}$ as defined by (5) is in general not symmetric,

$$\Gamma^{\lambda}{}_{\mu\nu} \neq \Gamma^{\lambda}{}_{\nu\mu},$$

giving rise to torsion, which we denote by

$$C^{\lambda}{}_{\mu\nu} \equiv \Gamma^{\lambda}{}_{\mu\nu} - \Gamma^{\lambda}{}_{\nu\mu}. \tag{9}$$

In the presence of torsion, the relations (8) imply a relation that is a generalization of the usual expression for the Christoffel connection. Using

$$g_{\lambda\mu;\nu} + g_{\nu\lambda;\mu} + g_{\mu\nu;\lambda} = 0,$$

one can obtain

$$\begin{aligned}\Gamma^{\lambda}{}_{\mu\nu} &= \tfrac{1}{2}(\Gamma^{\lambda}{}_{\mu\nu} + \Gamma^{\lambda}{}_{\nu\mu}) + \tfrac{1}{2}(\Gamma^{\lambda}{}_{\mu\nu} - \Gamma^{\lambda}{}_{\nu\mu}) \\ &= \tfrac{1}{2}g^{\lambda\rho}(g_{\rho\mu,\nu} + g_{\nu\rho,\mu} - g_{\mu\nu,\rho}) \\ &\quad + \tfrac{1}{2}(C_{\mu\nu}{}^{\lambda} + C_{\nu\mu}{}^{\lambda} + C^{\lambda}{}_{\mu\nu}).\end{aligned} \tag{10}$$

In the absence of torsion, this relation reduces to the usual Christoffel expression, with the geometry being Riemannian.

It is clear that the geometry is a Riemann–Cartan geometry, when torsion is not zero.

III. GAUSS–BONNET TYPE IDENTITIES

In Riemannian geometry, there are the following identities:

$$\sqrt{-g}\,\epsilon_{\alpha\beta\gamma\delta}\epsilon^{\mu\nu\lambda\rho}\mathscr{R}^{\alpha\beta}{}_{\mu\nu}\mathscr{R}^{\gamma\delta}{}_{\lambda\rho} = \text{total derivative}, \tag{11}$$

$$\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\mathscr{R}^{\alpha\beta}{}_{\mu\nu}R_{\alpha\beta\lambda\rho} = \text{total derivative}, \tag{12}$$

where $\epsilon_{\alpha\beta\gamma\delta}$ is the totally antisymmetric tensor, with

$$\epsilon_{0123} = -\sqrt{-g}. \tag{13}$$

For Riemann–Cartan geometry, discussions of these identities are not readily accessible. A simple and elementary discussion suitable for most physicists is lacking in the literature. We present here an elementary derivation, showing that the identities (11) and (12) also hold in Riemann–Cartan geometry.

Define

$$\mathscr{V}_{\mu} \equiv \tfrac{1}{4}\sigma_{ab}\mathscr{V}^{ab}{}_{\mu}, \tag{14}$$

where[5]

$$\sigma_{ab} = (i/2)[\gamma_a,\gamma_b],$$
$$\{\gamma_a,\gamma_b\} = 2\eta_{ab}. \tag{15}$$

We further define $R^{ab}{}_{\mu\nu}$ and $\bar{R}_{\mu\nu}$

$$\begin{aligned}\bar{R}_{\mu\nu} &= \tfrac{1}{4}\sigma_{ab}R^{ab}{}_{\mu\nu} \\ &\equiv \partial_{\mu}\mathscr{V}_{\nu} - \partial_{\nu}\mathscr{V}_{\mu} + i[\mathscr{V}_{\mu},\mathscr{V}_{\nu}],\end{aligned} \tag{16}$$

resulting in

$$R^{ab}{}_{\mu\nu} = \mathscr{V}^{ab}{}_{\mu,\nu} - \mathscr{V}^{ab}{}_{\nu,\mu} - \eta_{cd}(\mathscr{V}^{ac}{}_{\mu}\mathscr{V}^{db}{}_{\nu} - \mathscr{V}^{ac}{}_{\nu}\mathscr{V}^{db}{}_{\mu}). \tag{17}$$

It can be verified straightforwardly that

$$\begin{aligned}R^{\lambda\rho}{}_{\mu\nu} &\equiv e_a{}^{\lambda}e_b{}^{\rho}R^{ab}{}_{\mu\nu} \\ &= \Gamma^{\lambda\rho}{}_{\mu,\nu} - \Gamma^{\lambda\rho}{}_{\nu,\mu} - g_{\alpha\beta}(\Gamma^{\lambda\alpha}{}_{\mu}\Gamma^{\beta\rho}{}_{\nu} - \Gamma^{\lambda\alpha}{}_{\nu}\Gamma^{\beta\rho}{}_{\mu}),\end{aligned} \tag{18}$$

where $\Gamma^{\lambda}{}_{\mu\nu}$ is given by (5). This relation shows that the combination $e_a{}^{\lambda}e_b{}^{\rho}R^{ab}{}_{\mu\nu}$ is indeed the correct identification for the Riemann tensor (in Riemann–Cartan geometry).

Let us denote by η_{abcd} the totally antisymmetric Minkowski tensor, with

$$\eta_{0123} = -1. \tag{19}$$

There is the relationship

$$\eta_{abcd}e^{a}{}_{\mu}e^{b}{}_{\nu}e^{c}{}_{\lambda}e^{d}{}_{\rho} = \epsilon_{\mu\nu\lambda\rho}, \tag{20}$$

on account of

$$(\det e^{a}{}_{\mu})^2 = -\det g_{\mu\nu} = -g. \tag{21}$$

Using (18) and (20), we can express the Gauss–Bonnet product in the form

$$\begin{aligned}&\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\epsilon_{\alpha\beta\gamma\delta}R^{\alpha\beta}{}_{\mu\nu}R^{\gamma\delta}{}_{\lambda\rho} \\ &\quad= \sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\eta_{abcd}R^{ab}{}_{\mu\nu}R^{cd}{}_{\lambda\rho} \\ &\quad= 4i\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\mathrm{Tr}[\gamma_5\bar{R}_{\mu\nu}\bar{R}_{\lambda\rho}],\end{aligned} \tag{22}$$

where $\bar{R}_{\mu\nu}$ is defined in (16), and use has been made of

$$\mathrm{Tr}[\gamma_5\sigma_{ab}\sigma_{cd}] = -4i\eta_{abcd}.$$

Substituting (16) into (22), we obtain

$$\begin{aligned}&\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\epsilon_{\alpha\beta\gamma\delta}R^{\alpha\beta}{}_{\mu\nu}R^{\gamma\delta}{}_{\lambda\rho} \\ &\quad= 16i\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\mathrm{Tr}\{\gamma_5[(\partial_{\mu}\mathscr{V}_{\nu})(\partial_{\lambda}\mathscr{V}_{\rho}) \\ &\qquad + (\partial_{\mu}\mathscr{V}_{\nu})\mathscr{V}_{\lambda}\mathscr{V}_{\rho} + \mathscr{V}_{\mu}\mathscr{V}_{\nu}(\partial_{\lambda}\mathscr{V}_{\rho})]\},\end{aligned}$$

which, on account of

$$[\gamma_5,\mathscr{V}_{\mu}] = 0, \tag{23}$$

becomes

$$\begin{aligned}&16i\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\mathrm{Tr}\{\gamma_5[(\partial_{\mu}\mathscr{V}_{\nu})(\partial_{\lambda}\mathscr{V}_{\rho}) \\ &\quad + (2/3)\partial_{\mu}(\mathscr{V}_{\nu}\mathscr{V}_{\lambda}\mathscr{V}_{\rho})]\}.\end{aligned}$$

Since $(\sqrt{-g})\epsilon^{\mu\nu\lambda\rho}$ is a constant, we thus have[6]

$$\begin{aligned}&\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\epsilon_{\alpha\beta\gamma\delta}R^{\alpha\beta}{}_{\mu\nu}R^{\gamma\delta}{}_{\lambda\rho} \\ &\quad= \partial_{\mu}\{16i\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\mathrm{Tr}[\gamma_5(\mathscr{V}_{\nu}\partial_{\lambda}\mathscr{V}_{\rho} \\ &\qquad + (2/3)\mathscr{V}_{\nu}\mathscr{V}_{\lambda}\mathscr{V}_{\rho})]\}.\end{aligned} \tag{24}$$

One key point in the derivation is that γ_5 *commutes* with $\mathscr{V}_{\mu} = \tfrac{1}{4}\sigma_{ab}\mathscr{V}^{ab}{}_{\mu}$. There is another 4×4 matrix having this property. It is the unit matrix. This leads to the derivation of the other identity

$$\begin{aligned}&\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}R^{\alpha\beta}{}_{\mu\nu}R_{\alpha\beta\lambda\rho} = 2\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\mathrm{Tr}[\bar{R}_{\mu\nu}\bar{R}_{\lambda\rho}] \\ &= \partial_{\mu}\{8\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\mathrm{Tr}[\mathscr{V}_{\nu}\partial_{\lambda}\mathscr{V}_{\rho} + (2/3)\mathscr{V}_{\nu}\mathscr{V}_{\lambda}\mathscr{V}_{\rho}]\}.\end{aligned} \tag{25}$$

Relations (24) and (25) are the Gauss–Bonnet identities which, as we have seen, hold in Riemann–Cartan geometry as well as in Riemann geometry.

IV. BIANCHI TYPE IDENTITIES

The Bianchi identity for Riemann–Cartan geometry is known in the literature. It is a generalization of the usual Bianchi identity for the Riemann–Christoffel tensor, to take into account of torsion. We shall present a derivation of this identity directly in terms of the field variables $e^a{}_\mu$ and $\mathscr{V}^{ab}{}_\mu$, in the spirit of a gauge theory. We shall also derive a corresponding identity for the torsion tensor, which reduces to the cyclicity identity for the Riemann–Christoffel tensor when torsion is set equal to zero. The results contain nothing new. However, the derivations, which are based on using Lorentz and de Sitter algebras as artifice, may be of interest.

As is clear from (1), $\mathscr{V}^{ab}{}_\mu$ is the gauge potential for local Lorentz transformations, with the corresponding Yang–Mills field strength $R^{ab}{}_{\mu\nu}$ given by (17). It can be verified straightforwardly that

$$R^{ab}{}_{\mu\nu|\lambda} + R^{ab}{}_{\nu\lambda|\mu} + R^{ab}{}_{\lambda\mu|\nu} = 0, \quad (26)$$

where the gauge-covariant derivative is defined in (1). The identity is the Bianchi identity corresponding to a Lorentz algebra. In terms of the more general covariant derivatives defined in (6), it becomes

$$R^{ab}{}_{\mu\nu;\lambda} + R^{ab}{}_{\nu\lambda;\mu} + R^{ab}{}_{\lambda\mu;\nu} = C^\rho{}_{\mu\nu} R^{ab}{}_{\rho\lambda} + C^\rho{}_{\nu\lambda} R^{ab}{}_{\rho\mu} + C^\rho{}_{\lambda\mu} R^{ab}{}_{\rho\nu}, \quad (27)$$

where $C^\rho{}_{\mu\nu}$ is the torsion tensor defined by (9). On account of (7), the identity (27) can also be written in the form

$$R^{\alpha\beta}{}_{\mu\nu;\lambda} + R^{\alpha\beta}{}_{\nu\lambda;\mu} + R^{\alpha\beta}{}_{\lambda\mu;\nu} = C^\rho{}_{\mu\nu} R^{\alpha\beta}{}_{\rho\lambda} + C^\rho{}_{\nu\lambda} R^{\alpha\beta}{}_{\rho\mu} + C^\rho{}_{\lambda\mu} R^{\alpha\beta}{}_{\rho\nu}, \quad (28)$$

where $R^{\alpha\beta}{}_{\mu\nu}$ is defined by (18). The identity in the form (28) is the generalized Bianchi identity for Riemann–Cartan geometry.

The field variables $\mathscr{V}^{ab}{}_\mu$ and $e^a{}_\mu$ can be grouped together to form $\mathscr{V}^{AB}{}_\mu = -\mathscr{V}^{BA}{}_\mu$ (with $A = 0,1,2,3,5$):

$$\mathscr{V}^{AB}{}_\mu : \begin{cases} \mathscr{V}^{ab}{}_\mu, \\ \mathscr{V}^{a5}{}_\mu = e^a{}_\mu. \end{cases} \quad (29)$$

We can artificially consider the de Sitter algebra

$$i[X_{AB}, X_{CD}] = \eta_{AC}X_{BD} - \eta_{AD}X_{BC} + \eta_{BD}X_{AC} - \eta_{BC}X_{AD}, \quad (30)$$

where

$$\eta_{AB} = (1, -1, -1, -1, \pm 1). \quad (31)$$

Define

$$\widetilde{\mathscr{V}}_\mu \equiv \tfrac{1}{4} X_{AB} \mathscr{V}^{AB}{}_\mu, \quad (32)$$

$$\tfrac{1}{4} X_{AB} \tilde{R}^{AB}{}_{\mu\nu} \equiv \widetilde{\mathscr{V}}_{\mu,\nu} - \widetilde{\mathscr{V}}_{\nu,\mu} + i[\widetilde{\mathscr{V}}_\mu, \widetilde{\mathscr{V}}_\nu], \quad (33)$$

which yield

$$\tilde{R}^{AB}{}_{\mu\nu} = \mathscr{V}^{AB}{}_{\mu,\nu} - \mathscr{V}^{AB}{}_{\nu,\mu} + \eta_{CD} \times (\mathscr{V}^{AC}{}_\mu \mathscr{V}^{BD}{}_\nu - \mathscr{V}^{BC}{}_\mu \mathscr{V}^{AD}{}_\nu). \quad (34)$$

It can be straightforwardly verified that

$$\tilde{R}^{ab}{}_{\mu\nu} = R^{ab}{}_{\mu\nu} + \eta_{55}(e^a{}_\mu e^b{}_\nu - e^b{}_\mu e^a{}_\nu), \qquad \tilde{R}^{a5}{}_{\mu\nu} = e^a{}_{\mu|\nu} - e^a{}_{\nu|\mu} = C^a{}_{\mu\nu}, \quad (35)$$

where

$$C^a{}_{\mu\nu} = e^a{}_\lambda C^\lambda{}_{\mu\nu}.$$

Defining the covariant derivatives corresponding to the de Sitter algebra, such as

$$\tilde{R}^{AB}{}_{\mu\nu\|\lambda} \equiv \tilde{R}^{AB}{}_{\mu\nu,\lambda} + \eta_{CD}(\mathscr{V}^{AC}{}_\lambda \tilde{R}^{DB}{}_{\mu\nu} - \mathscr{V}^{BC}{}_\lambda \tilde{R}^{DA}{}_{\mu\nu}), \quad (36)$$

we can write down the Bianchi identity corresponding to the de Sitter algebra:

$$\tilde{R}^{AB}{}_{\mu\nu\|\lambda} + \tilde{R}^{AB}{}_{\nu\lambda\|\mu} + \tilde{R}^{AB}{}_{\lambda\mu\|\nu} = 0. \quad (37)$$

There are two sets of identities, one corresponding to setting $A = a$ and $B = b$, and the other $A = a$ and $B = 5$ in (37). It can be easily checked that

$$\tilde{R}^{ab}{}_{\mu\nu\|\lambda} + \tilde{R}^{ab}{}_{\nu\lambda\|\mu} + \tilde{R}^{ab}{}_{\lambda\mu\|\nu} = 0$$

is equivalent to (26), again yielding (27). The other set

$$\tilde{R}^{a5}{}_{\mu\nu\|\lambda} + \tilde{R}^{a5}{}_{\nu\lambda\|\mu} + \tilde{R}^{a5}{}_{\lambda\mu\|\nu} = 0,$$

on account of the definitions of the two kinds of covariant derivatives,

$$\tilde{R}^{a5}{}_{\mu\nu\|\lambda} = \tilde{R}^{a5}{}_{\mu\nu|\lambda} + \tilde{R}^a{}_{b\mu\nu} e^b{}_\lambda$$

can be written in the form

$$\tilde{R}^{a5}{}_{\mu\nu|\lambda} + \tilde{R}^{a5}{}_{\nu\lambda|\mu} + \tilde{R}^{a5}{}_{\lambda\mu|\nu} + \tilde{R}^a{}_{b\mu\nu} e^b{}_\lambda + \tilde{R}^a{}_{b\nu\lambda} e^b{}_\mu + \tilde{R}^a{}_{b\lambda\mu} e^b{}_\nu = 0. \quad (38)$$

In terms of the general covariant derivative defined in (6) this identity can be expressed as

$$C^a{}_{\mu\nu;\lambda} + C^a{}_{\nu\lambda;\mu} + C^a{}_{\lambda\mu;\nu} + R^a{}_{b\mu\nu} e^b{}_\lambda + R^a{}_{b\nu\lambda} e^b{}_\mu + R^a{}_{b\lambda\mu} e^b{}_\nu = C^\rho{}_{\mu\nu} C^a{}_{\rho\lambda} + C^\rho{}_{\nu\lambda} C^a{}_{\rho\mu} + C^\rho{}_{\lambda\mu} C^a{}_{\rho\nu}, \quad (39)$$

where use has been made of (35). Because of (7), we can write (39) in the form

$$C^\alpha{}_{\mu\nu;\lambda} + C^\alpha{}_{\nu\lambda;\mu} + C^\alpha{}_{\lambda\mu;\nu} + R^\alpha{}_{\mu\nu\lambda} + R^\alpha{}_{\nu\lambda\mu} + R^\alpha{}_{\lambda\mu\nu} = C^\rho{}_{\mu\nu} C^\alpha{}_{\rho\lambda} + C^\rho{}_{\nu\lambda} C^\alpha{}_{\rho\mu} + C^\rho{}_{\lambda\mu} C^\alpha{}_{\rho\nu}. \quad (40)$$

We note that this identity reduces to the cyclicity relation for the Riemann-Christoffel tensor, when the torsion tensor $C^\lambda{}_{\mu\nu}$ is zero.

We have thus seen that the two identities (27) and (40) are the Bianchi identities corresponding to the de Sitter algebra.

ACKNOWLEDGMENT

This work is partially supported by the National Science Foundation under Grant No. PHY 79-06376.

[1] H. Weyl, Z. Phys. **56**, 330 (1929).

[2] H. Weyl, Phys. Rev. **77**, 699 (1950); T.W.B. Kibble, J. Math. Phys. **2**, 212 (1961); D.W. Sciama, in *Recent Developments in General Relativity* (Pergamon, New York, 1962). A general review is given by F.W. Hehl, P. von der Heyde, and G.D. Kerlick, Rev. Mod. Phys. **48**, 393 (1976). Recent works include F. Mansouri and L.N. Chang, Phys. Rev. D **13**, 3192 (1976); F.W. Hehl *et al.*, Phys. Lett. B **78**, 102 (1978); H.T. Nieh, Stony Brook preprint ITP-SB-79-76 (1979).

[3] B.S. DeWitt, *The Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965); T.W.B. Kibble, in Ref. 2.

[4] A major portion of this section, which we include for self-containment, can be found in T.W.B. Kibble, in Ref. 2.

[5] For the Dirac matrices, we follow the convention of J.D. Bjorken and S.D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).

[6] The Gauss–Bonnet identity (24) has been used, e.g., in S.W. MacDowell and F. Mansouri, Phys. Lett. **38**, 739 (1977).

Matching of the plane symmetric static and homogeneous vacuum solutions of the Einstein field equations

George T. Carlson, Jr. [a] and John L. Safko
Department of Physics and Astronomy, University of South Carolina, Columbia, South Carolina 29208

(Received 19 March 1979; accepted for publication 26 November 1979)

The Taub plane symmetric static and homogeneous vacuum solutions are matched on a natural hypersurface. The space–times obtained in this way have distribution valued curvature tensors along the joining hypersurfaces. Our treatment of this problem follows Taub's presentation of space–times with distribution valued curvature tensors. We find that the surfaces of the join may be interpreted as thin null pressureless fluid shocks. The nature of these surfaces are further investigated by examining the behavior of geodesics crossing the surfaces.

I. INTRODUCTION

In Sec. II we give Taub's[1] definition of plane-symmetric space–times, and the plane-symmetric static and homogeneous vacuum solutions. In Sec. III we show how space–times may be produced by matching the static and homogeneous vacuum solutions along a hypersurface. The space–times produced in this way have distribution-valued energy-momentum tensors along the joining hypersurfaces. Taub[2] has presented a formalism for dealing with such hypersurfaces. The energy-momentum tensors along the joining hypersurfaces discussed in Sec. III are calculated in Sec. IV. Geodesics are the subject of Sec. V. It is shown how to continue geodesics across distribution-valued hypersurfaces. The first integrals are given for the static and homogeneous vacuum solutions. We specialize to the case where the motion is restricted to a z axis. Finally, we display some typical geodesics crossing the joining hypersurfaces.

In the discussion and conclusion we interpret the joining hypersurface as due to a null pressureless fluid. We also show how space–times such as those discussed in Sec. III may arise in a more natural way. The field equations may have characteristic surfaces upon which the equation cannot be integrated. If the metric is extended onto such a hypersurface it may be interpretable in a distributed sense.

II. BACKGROUND

A. Definition of plane-symmetric space–times

Consider a space–time (M,g). Let $(x,y,z,t) \longleftrightarrow (x^1,x^2,x^3,x^4)$ be local coordinates in a neighborhood of a point $P\in M$ The mapping

$$\begin{aligned}&\bar{x} = x\cos(\theta) + y\sin(\theta) + a\,,\\ &\bar{y} = x\sin(\theta) + y\cos(\theta) + b\,,\\ &0<\theta<2\pi,\ -\infty<a,b<\infty\,,\\ &\bar{z}=z,\ \bar{t}=t\,,\end{aligned} \tag{1}$$

is used to define plane symmetry. (M,g) is said to be plane symmetric if local coordinates (x,y,z,t) exist such that:

(i) Equation (1) is a local coordinate representation of a Lie-transformation group on M;

(ii) the infinitesimal generators of Eq. (1) ($k_{(a)}, k_{(b)}$, and $k_{(\theta)}$) are spacelike (cf. Taub[1]).

If (M,g) is plane symmetric, then Killings's equations for

$$\begin{aligned}&k_{(a)} = k_1\frac{\partial}{\partial x},\quad k_{(b)} = k_2\frac{\partial}{\partial y},\\ &k_{(\theta)} = k_3\Big(y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\Big),\end{aligned} \tag{2}$$

where k_1, k_2, and k_3 are constant nonzero scalars, imply that we may choose local coordinates (x,y,z,t) such that

$$ds^2 = A\,(dx^2+dy^2) + B\,dz^2 + 2C\,dz\,dt + D\,dt^2\,, \tag{3}$$

where A, B, C, and D are functions of z and t only. Conditions (i) and (ii) imply

$$-\infty < x, y < \infty;\ A>0\,. \tag{4}$$

In order that Eq. (3) define a Lorentz metric, B, C, and D must satisfy the Lorentz signature requirements

$$\begin{aligned}&B+D+[(B-D)^2+4C^2]^{1/2}>0\,,\\ &B+D-[(B-D)^2+4C^2]^{1/2}<0\,.\end{aligned} \tag{5}$$

The inequalities (4) and (5) may in general put restrictions on the range of the coordinates z and t.

If A, B, C, and D are analytic functions of z and t, there exists a coordinate transformation of the form

$$\begin{aligned}&x'=x,\quad y'=y,\\ &z'=F(z,t),\\ &t'=G(z,t),\end{aligned} \tag{6}$$

such that (3) takes the Taub[1] canonical form

$$ds^2 = A\,(dx^2+dy^2) + B\,(dz^2-dt^2)\,. \tag{7}$$

Carlson and Safko[3] have discussed canonical forms for C^k plane-symmetric metrics.

B. Taub static and homogeneous vacuum solutions

The Einstein equations $R_{\mu\nu}=0$ for (7) admit two nonflat solutions

$$g_{(s)}: ds^2_{(s)} = \Omega_0^2\,[\,|z|(dx^2+dy^2) + |z|^{-1/2}(dz^2-dt^2)]\,, \tag{8a}$$

[a] Current address: Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221.

 0022-2488/80/061442-07$1.00

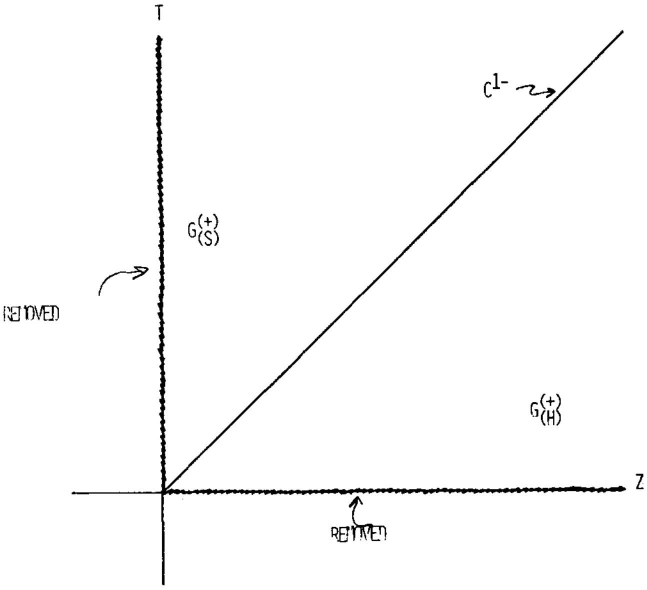

FIG. 1. The space–time $(M^{(+)}_{(i)}, g^{(+)}_{(i)})$. $g^{(+)}_{(i)}$ is Lipschitz continuous on $z = t$. The space-time cannot be extended out of the first quadrant of the (z,t) plane.

$$g_{(h)}: ds^2_{(h)} = \Omega_0^2 [\,|t\,|(dx^2 + dy^2) + |t\,|^{-1/2}(dz^2 - dt^2)] \,. \tag{8b}$$

Taub[1] derived the static solution (subscript s) and Davis and Ray[4] pointed out that the homogeneous solution (subscript h), which is a special case of the Kasner metric, also follows from solving $R_{\mu\nu} = 0$ for (7).

If in Eq. (8) the coordinates z and t are allowed to take on all values $-\infty < z, t < \infty$, then $g_{(s)}$ is singular on $z = 0$ and $g_{(h)}$ is singular on $t = 0$. So what we really have are four space–times

$$(M^{(\pm)}_{(s,h)}, g^{(\pm)}_{(s,h)}),$$

where

$$ds^{(\pm)2}_{(s)} = \Omega_0^2 [(\pm z)(dx^2 + dy^2) + (\pm z)^{-1/2}(dz^2 - dt^2)], \tag{9}$$

$$M^{(\pm)}_{(s)} = \left\{(x,y,z,t) \in R^4 / \begin{matrix} z > 0(+) \\ z < 0(-) \end{matrix}\right\},$$

$$ds^{(\pm)2}_{(h)} = \Omega_0^2 [(\pm t)(dx^2 + dy^2) + (\pm t)^{-1/2}(dz^2 - dt^2)], \tag{10}$$

$$M^{(\pm)}_{(h)} = \left\{(x,y,z,t) \in R^4 / \begin{matrix} t > 0(+) \\ t < 0(-) \end{matrix}\right\},$$

with an obvious use of notation.

Note that

$$(M^{(+)}_{(s,h)}, g^{(+)}_{(s,h)}) \simeq (M^{(-)}_{(s,h)}, g^{(-)}_{(s,h)}).$$

These space–times are locally analytic, nondegenerate, nonsingular, and inextendable By inextendable we mean that the space–times cannot be extended at their singularities which show up as singularities in $R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta}$.

III. MATCHING THE STATIC AND HOMOGENEOUS VACUUM SOLUTIONS

In this section we show that larger spaces may be pro-

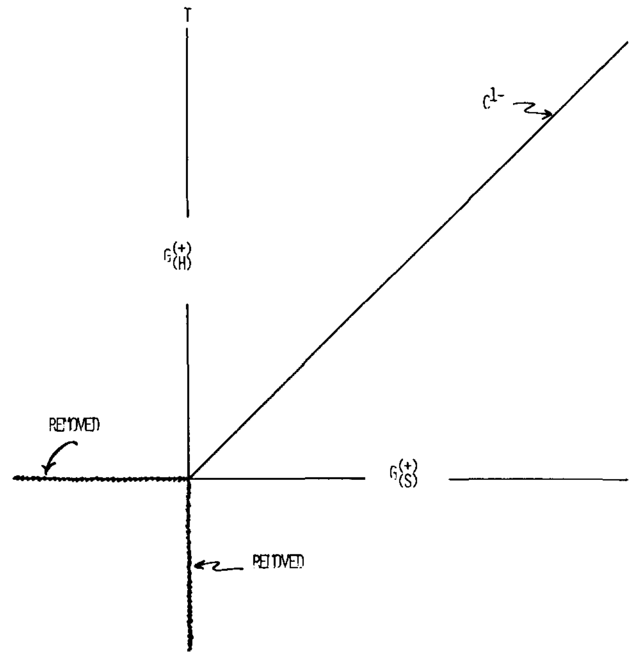

FIG. 2. The space–time $(M^{(+)}_{(u)}, g^{(+)}_{(u)})$. $g^{(+)}_{(u)}$ is Lipschitz continuous on $z = t$. The space–time cannot be extended into the third quadrant of the (z,t) plane.

duced by matching the $g^{(\pm)}_{(s,h)}$ along a hypersurface. We require that the first fundamental form of the metric be continuous across the hypersurface. We use the same coordinates on each side so that continuity of the first fundamental form reduces to the metric being continuous across the hypersurface.

The space

$$M^{(+)}_{(i)} = M^{(+)}_{(h)} \cap M^{(+)}_{(s)}$$

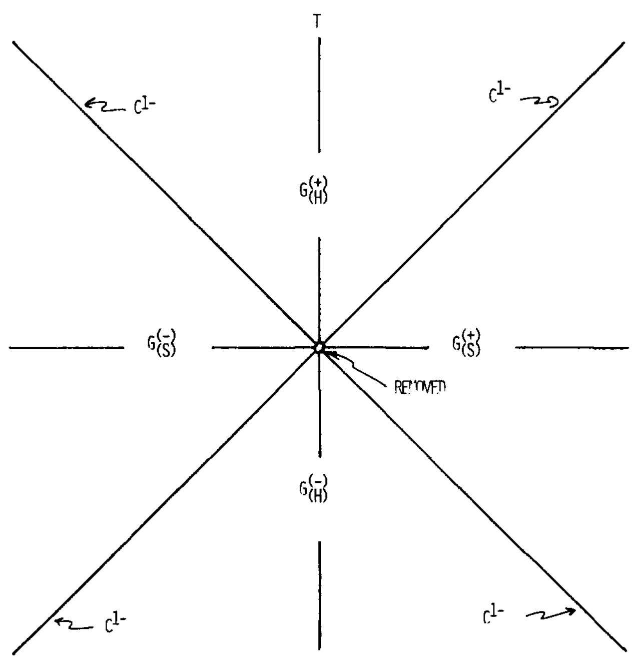

FIG. 3. The space–time (M,g). g is Lipschitz continuous on $z = \pm t$. The space-time cannot be extended onto the two-dimensional surface $z = 0 = t$.

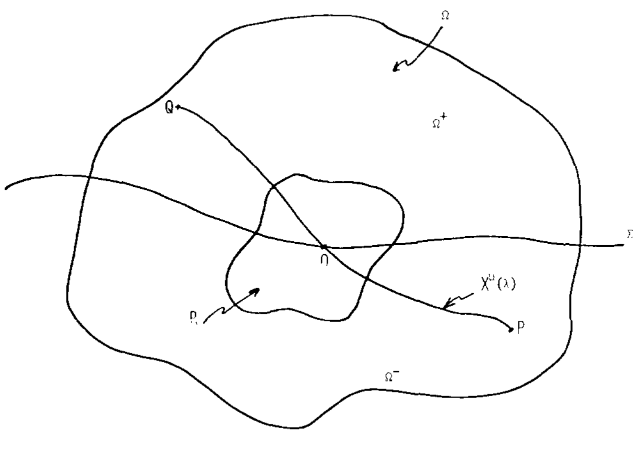

FIG. 4. A region Ω of space-time is divided into two parts $\Omega^{\pm}$ by Σ. In Ω^{+}, $\phi > 0$ and in Ω^{-}, $\phi < 0$. The curve $x^{\mu}(\lambda)$ from $P\in\Omega^{+}$ to $q\in\Omega^{-}$ crosses Σ at the point 0. R is a neighborhood of 0 and is contained in Ω. Σ divides R into two parts $R^{\pm}$.

has a global C^{1-}, nondegenerate, and nonsingular metric $g^{(+)}_{(l)}$ given by

$$g^{(+)}_{(l)} = \begin{cases} g^{(+)}_{(s)}, & t \geqslant z, \\ g^{(+)}_{(h)}, & t \leqslant z. \end{cases} \tag{11}$$

The space-time $(M^{(+)}_{(l)}, g^{(+)}_{(l)})$ is inextendable and may be depicted schematically as shown in Fig. 1.

The space

$$M^{(+)}_{(u)} = M^{(+)}_{(h)} \cup M^{(+)}_{(s)}$$

has a global C^{1-}, nondegenerate, and nonsingular metric $g^{(+)}_{(u)}$ given by

$$g^{(+)}_{(u)} = \begin{cases} g^{(+)}_{(h)}, & t \geqslant z, \\ g^{(+)}_{(s)}, & t \leqslant z. \end{cases} \tag{12}$$

The space–time $(M^{(+)}_{(u)}, g^{(+)}_{(u)})$ is inextendable and may be depicted schematically as shown in Fig. 2.

A more interesting space–time with $-\infty < z, t < \infty$ $(z \neq 0 \neq t)$ is (M,g) shown schematically in Fig. 3. g is globally C^{1-} nondegenerate and nonsingular on M. (M,g) is also inextendable. (M,g) may be considered as an extension of $(M^{(+)}_{(u)}, g^{(+)}_{(u)})$ in the sense that we cut out the region of $(M^{(+)}_{(u)}, g^{(+)}_{(u)})$ below the line $z = -t$ and fill it in as shown. But this does not extend $(M^{(+)}_{(u)}, g^{(+)}_{(u)})$ at its singularities.

IV. THE NATURE OF THE HYPERSURFACE AT THE MATCH

The surfaces $z = t$ in $(M^{(+)}_{(i,u)}, g^{(+)}_{(i,u)})$ and the surfaces $z = \pm t$ in (M,g) are null hypersurfaces across which the metric tensor is continuous but has finite jumps in its first and second derivatives. Hence the curvature tensor may be distribution valued. In the following discussion we analyze the nature of the surfaces using Taub's[2] presentation of space-times with distribution-valued curvature tensors.

Taub shows that the new part of the Ricci tensor on the surface is given by two quantities $H_{\mu\nu}\delta(\psi)$ and $J_{\mu\nu}\theta(\psi) \times [1-\theta(\psi)]$, where $\psi = 0$ is the hypersurface, δ is the Dirac delta function, and θ is the step function. With $H = H^{\mu}_{\mu}$ and $J = J^{\mu}_{\mu}$, we call $H'_{\mu\nu} = H_{\mu\nu} - \frac{1}{2}g_{\mu\nu}H$ the energy–momentum tensor of the surface and $J'_{\mu\nu} = J_{\mu\nu} - \frac{1}{2}g_{\mu\nu}J$ the noncontributing energy–momentum tensor of the surface. The Bianchi identities need not be satisfied on the surface, although they will be satisfied on either side and for any integral across the surface.

We begin our discussion with the space–time $(M^{(+)}_{(l)}, g^{(+)}_{(l)})$. Consider the null coordinates

$$\phi = z - t, \quad \omega = z + t, \tag{13}$$

in which the metric takes the form (with $\Omega_0^2 = 1$)

$$ds^2 = |\tfrac{1}{2}(\omega + \phi)|(dx^2 + dy^2) + 2[2(\omega+\phi)]^{-1/2} d\omega\, d\phi, \quad \phi \leqslant 0,$$

$$ds^2 = |\tfrac{1}{2}(\omega - \phi)|(dx^2 + dy^2) + 2[2(\omega-\phi)]^{-1/2} d\omega\, d\phi, \quad \phi \geqslant 0. \tag{14}$$

Labeling the coordinates according to $(x, y, \omega, \phi) \longleftrightarrow (x^1, x^2, x^3, x^4)$ and following the algorithm outlined in the Appendix, we conclude that

$$H_{\mu\nu} = (-2/\omega)\delta_{4\mu}\delta_{4\nu}, \quad J_{\mu\nu}(2/\omega)^2\delta_{4\mu}\delta_{4\nu}. \tag{15}$$

We also find that

$$B^{\rho\sigma\mu\nu}{}_{;\nu} = -\theta(1-\theta)\tfrac{1}{2}(\omega)^{-3}\delta^{\sigma 3}(\delta^{1\mu}\delta^{2\rho} + \delta^{2\mu}\delta^{1\rho}), \tag{16}$$

so the Bianchi identities are not satisfied on $\phi = 0$; however, the contracted Bianchi identities are satisfied on $\phi = 0$.

Now consider $(M^{(+)}_{(u)}, g^{(+)}_{(u)})$. The only difference here will be a change in sign of jump quantities. We now find that

$$H_{\mu\nu} = (2/\omega)\delta_{4\mu}\delta_{4\nu}, \quad J_{\mu\nu} = (2/\omega)^2\delta_{4\mu}\delta_{4\nu}, \tag{17}$$

and an equation similar to (16).

It is also a simple matter to find $H_{\mu\nu}$ and $J_{\mu\nu}$ for (M,g). The first quadrant of the (z,t) plane is equivalent to that part of $(M^{(+)}_{(u)}, g^{(+)}_{(u)})$. The second quadrant is equivalent to the first quadrant via $z \rightarrow -z$ $(\omega \rightarrow \phi, \phi \rightarrow -\omega)$. The third quadrant is equivalent to the first quadrant via $z \rightarrow -z$, $t \rightarrow -t$ $(\omega \rightarrow -\omega, \phi \rightarrow -\phi)$ and the fourth quadrant is equivalent to the first quadrant via $t \rightarrow -t$ $(\omega \rightarrow \phi, \phi \rightarrow \omega)$. So we have

$$H_{\mu\nu} = \begin{cases} (2/\omega)\delta_{4\mu}\delta_{4\nu}, & \phi = 0 \quad (z, t > 0), \\ (-2/\phi)\delta_{3\mu}\delta_{3\nu}, & \phi = 0 \quad (z < 0, t > 0), \\ (-2/\omega)\delta_{4\mu}\delta_{4\nu}, & \omega = 0 \quad (z, t < 0), \\ (2/\phi)\delta_{3\mu}\delta_{3\nu}, & \omega = 0 \quad (z > 0, t < 0), \end{cases} \tag{18}$$

$$J_{\mu\nu} = \begin{cases} (2/\omega)^2\delta_{4\mu}\delta_{4\nu}, & \phi = 0 \quad (z, t > 0), \\ (2/\phi)^2\delta_{3\mu}\delta_{3\nu}, & \omega = 0 \quad (z < 0, t > 0), \\ (2/\omega)^2\delta_{4\mu}\delta_{4\nu}, & \phi = 0 \quad (z, t < 0), \\ (2/\phi)^2\delta_{3\mu}\delta_{3\nu}, & \omega = 0 \quad (z > 0, t < 0) \end{cases} \tag{19}$$

and equations similar to (16).

Note that $H'_{\mu\nu}$ and $J'_{\mu\nu}$ have the algebraic properties of null pressureless fluids, i.e., $H'_{\mu\nu} = \rho_1 l_\mu l_\nu$ and $J'_{\mu\nu} = \rho_2 l_\mu l_\nu$, where $l_\mu = \delta_{4\mu}$ is a null-geodesic vector field. Also, $(M^{(+)}_{(l)}, g^{(+)}_{(l)})$ has negative energy density, whereas $(M^{(+)}_{(l)}, g^{(+)}_{(l)})$ has positive energy density as far as $H'_{\mu\nu}$ is concerned. $J'_{\mu\nu}$ does not contribute to any integrated energy.

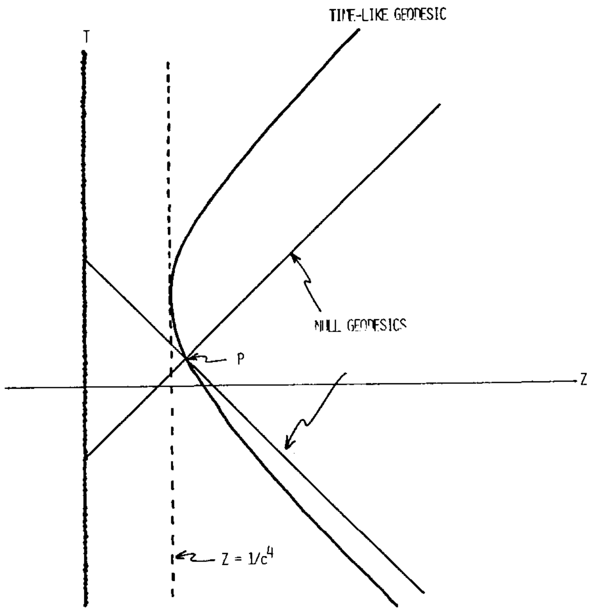

FIG. 5. Through the point P in the space–time $(M^{(+)}_{(s)}, g^{(+)}_{(s)})$ we have sketched the light cone and a typical timelike geodesic in the case $a = 0 = b$. The complete timelike geodesic bounces off of the hypersurface $z = (1/c)^4$.

V. GEODESICS

A. Geodesics in space–times with distribution-valued curvature tensors

In Fig. 4 we show a curve $x^\mu(\lambda)$ from $P \in \Omega^-$ to $q \in \Omega^+$ and crossing Σ at the point 0. A particle (allowing tachyons) will travel from P to q in such a way that

$$\delta I = \delta \left[\int_{\lambda(p)}^{\lambda(q)} \left(g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} \right) d\lambda \right] = 0, \tag{20}$$

where the variation δ is taken in the usual way. The equations of motion are then found to be

$$\left(\frac{d^2 x^\mu}{dv^2} + \Gamma^\mu_{\nu\rho} \frac{dx^\nu}{dv} \frac{dx^\rho}{dv} \right)^\pm = 0, \quad \left[\frac{dx^\mu}{dv} \right] = 0, \tag{21}$$

where v is an affine parameter. Hence we have the geodesic equations in $\Omega^\pm$ and the condition that the vector tanget to $x^\mu(v)$ be continuous across Σ.

Since $g_{\mu\nu}$ is C^3, $\Gamma^\mu_{\nu\rho}$ will be C^2. One can apply the existence and uniqueness theorem for the geodesic equations to prove the following: If $\Gamma^\mu_{\nu\rho}$ has a C^2 extension into $R^\pm$, every geodesic starting somewhere in $R^\pm$ and reaching 0 on Σ with tangent vector $(dx^\mu/dv)|_{v(0)}$ has a unique extension into $R^\mp$ satisfying (21), relative to the C^2 extension of $\Gamma^{\pm\mu}_{\nu\rho}$ into $R^\mp$.

B. Geodesics of the vacuum solutions

Horsky[5] and others (see references in Horsky) have investigated some of the properties of the geodesics of the static solution. These also fall within the class of Bianchi I spaces. The static metric given by Eq. (9) has first integrals

$$\frac{dx}{dv} = a/z, \quad \frac{dy}{dv} = b/z, \quad \frac{dt}{dv} = c(z)^{1/2},$$

$$\left(\frac{dz}{dv}\right)^2 = c^2 z - (a^2 + b^2)(z)^{-1/2} - d(z)^{1/2}, \tag{22}$$

with v an affine parameter and a, b, c, and d constants of integration. Now v may be chosen as proper length in spacelike geodesics and as proper time for timelike geodesics since $(ds)^2/(dv)^2 = -d$. If we consider geodesics for which $a = b = 0$, then null geodesics are given by $z = \pm t + \text{const.}$, while timelike geodesics are given by $dz/dt = \pm(1 - c^{-2} z^{-1/2})^{1/2}$.

These timelike geodesics are complete across $z = c^{-4}$ if we change from the (+) root to the (−) root. Fig. 5 shows a typical timelike and null geodesic through an arbitrary point P.

The homogeneous geodesics are found by replacing $z^{1/2}$ by $-t^{1/2}$ in all first-integral equations. Timelike and null geodesics are shown in Fig. 6.

C. Geodesics of the distribution-valued solutions

Once again we consider the case of geodesic motion constrained to the z axis. We know what these geodesics look like for the static and homogeneous vacuum solutions. All we have to do to display typical geodesics of the distribution-valued solutions is to continue that tangent vector continuously across $z = \pm t$. In Figs. 7, 8, and 9 we show the light cone through an arbitrary point P and the typical timelike geodesic through P for the space-times $(M^{(+)}_{(i)}, g^{(+)}_{(i)})$, $(M^{(+)}_{(n)}, g^{(+)}_{(n)})$, and (M, g).

VI. DISCUSSION AND CONCLUSION

By patching together the two vacuum solutions with plane symmetry we have constructed plane-symmetric space–times that contain null hypersurfaces across which the metric tensor is Lipschitz continuous and has finite jumps in its second derivatives. The energy–momentum tensor on the null hypersurfaces may then be calculated as pre-

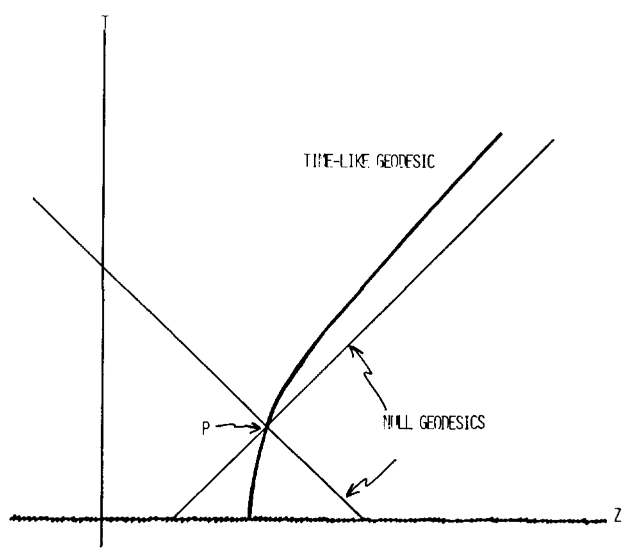

FIG. 6. Through the point P in the space–time $(M^{(+)}_{(h)}, g^{(+)}_{(h)})$ we have sketched the light cone and a typical timelike geodesic in the case $a = 0 = b$.

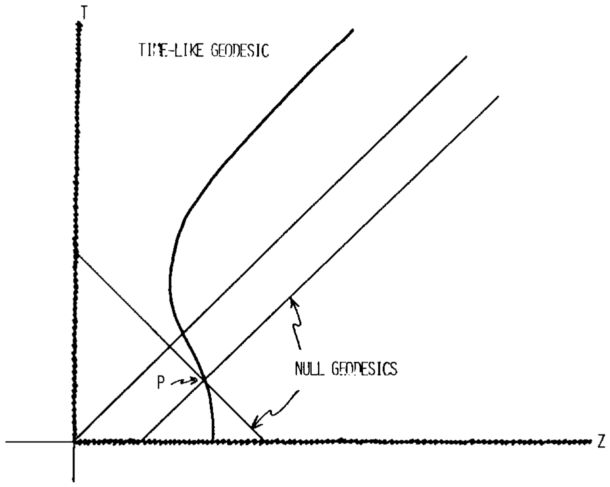

FIG. 7. Through the point P in the space–time $(M^{(+)}_{(l)}, g^{(+)}_{(l)})$ we have sketched the light cone and a typical timelike geodesic in the case $a = 0 = b$. The timelike geodesic experiences a jump in its acceleration as it crosses the null hypersurface $z = t$, which is a thin null pressureless fluid.

scribed by Taub.[2] In general, the curvature tensor will be of the form $A + B\delta(\psi) + C\theta(\psi)[1 - \theta(\psi)]$. The term proportional to $\theta(\psi)[1 - \theta(\psi)]$, where $\psi = 0$ is the surface, will not contribute to any integrated energy and momentum, however, the term proportional to $\delta(\psi)$ will contribute to integrated energy and momentum and is therefore of physical significance. We found that the energy-momentum tensor associated with $\delta(\psi)$ has the algebraic properties of a null pressureless fluid. For a null pressureless fluid $T_{\mu\nu} = \rho l_\mu l_\nu$,

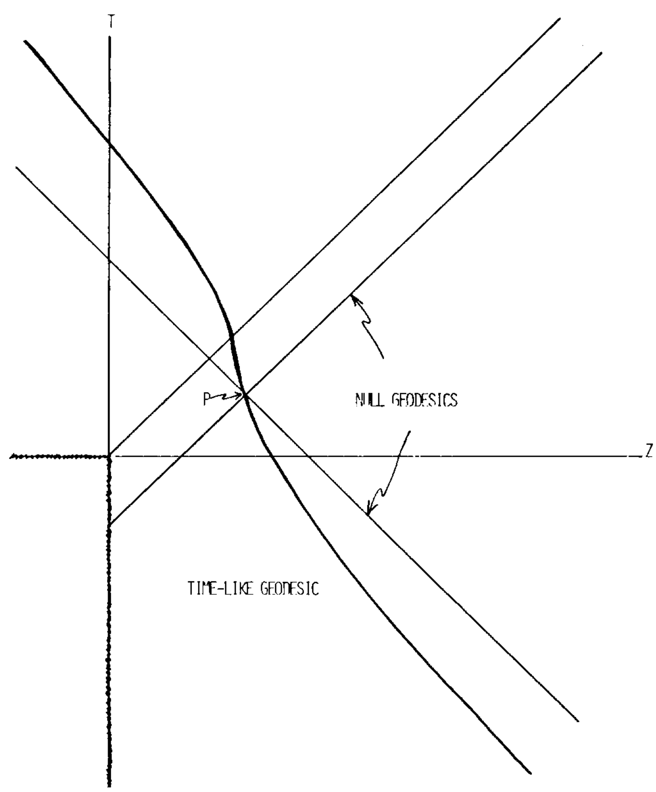

FIG. 8. Through the point P in the space time $(M^{(+)}_{(u)}, g^{(+)}_{(u)})$ we have sketched the light cone and a typical timelike geodesic in the case $a = 0 = b$. The timelike geodesic experiences a jump in its acceleration as it crosses the null hypersurface $z = t$, which is a thin null pressureless fluid.

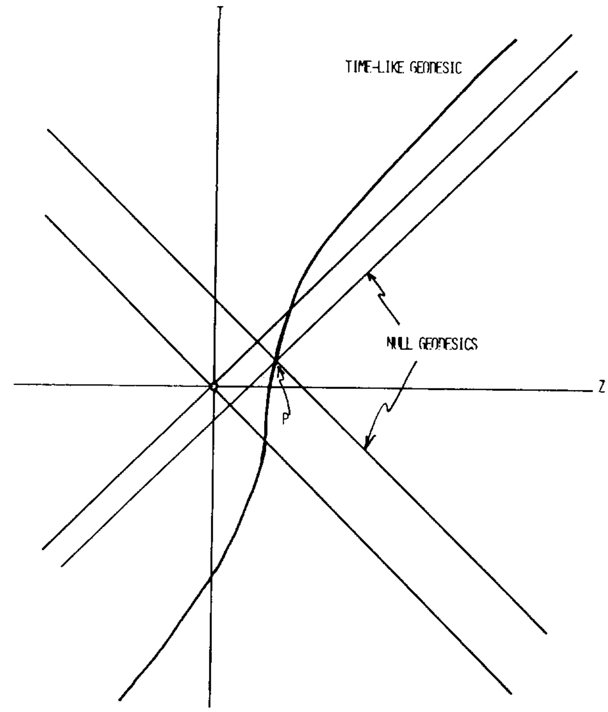

FIG. 9. Through the point P in the space–time (M,g) we have sketched the light cone and a typical timelike geodesic in the case $a = 0 = b$. The timelike geodesic experienced a jump in its acceleration in the past as it crossed the null hypersurface $z = -t$, and will experience a jump in its acceleration as it crosses the null hypersurface $z = t$.

where l_μ is a null-geodesic vector field. We look for $l_\mu = \delta_{4\mu}$ and $\rho = A(\omega)\delta(\phi) = -(2/\omega)\delta(\phi)$. The field equation is $\rho;_\mu l^\mu + \rho l^\mu;_\mu = 0$.

This equation implies that $\rho,_3 + \rho/\omega = 0$ so that $dA/d\omega = -A/\omega$, or $\rho = (C/\omega)\delta(\phi)$. We satisfy our hypothesis with $C = -2$.

Geodesic particles crossing the shock experience a jump in their acceleration and follow a different path than they would have taken had the shock not been there.

From a mathematical point of view the arbitrary patching together of space-times is a well posed problem; however, it seems to us to be rather unnatural from a physical point of view.

In order to see that the space-times discussed above may actually arise in a more natural way, start with the plane-symmetric metric.[6]

$$ds^2 = e^{(A+B)/2}(dx^2 + dy^2) + e^{2C}(dz^2 - e^{(B-A)}dt^2),$$
$$A = A(z),\ B = B(t),\ C = C(z,t)\,, \tag{23}$$

and solve $R_{\mu\nu} = 0$. One finds that the resulting equations cannot be integrated on degenerate hypersurfaces across which the metric tensor is Lipschitz continuous. One such solution is

$$ds^2 = e^{(z+t)/2}(dx^2 + dy^2) + |e^t - e^z|\,e^{-(z+t)/4}(e^z\,dz^2 - e^t\,dt^2). \tag{24}$$

Equation (23) arose naturally from integrating $R_{\mu\nu} = 0$ subject to $z \neq t$. In fact, Eq. (24) does not satisfy $R_{\mu\nu} = 0$ on $z = t$.

If in Eq. (24) we allow z and t to take on all values

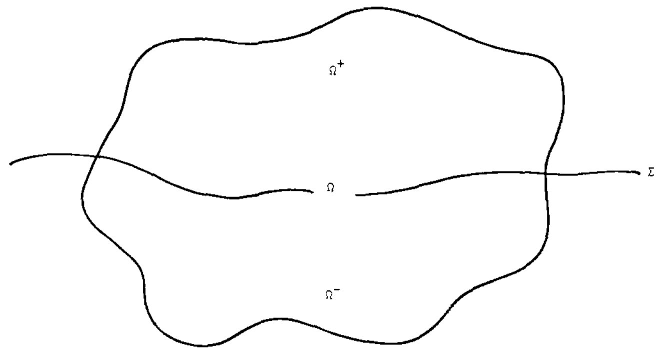

FIG. 10. A region of space–time is divided by Σ into two parts $\Omega^{\pm}$. In Ω^+, $\phi > 0$ and in Ω^-, $\phi < 0$.

$-\infty < z, t < \infty$, we can show that the resulting space-time is equivalent to $(M_{(i)}^{(+)}, g_{(i)}^{(+)})$. The C^∞ coordinate transformation $(z,t) \to (\bar{z},\bar{t})$ composed from

$$\alpha = 2e^{z/2}, \quad \beta = 2e^{t/2},$$

$$u = \alpha + \beta, \quad v = \alpha - \beta,$$

$$\bar{u} = u^2/2, \quad \bar{v} = \begin{cases} v^2/2, & v \geqslant 0, \\ -v^2/2, & v \leqslant 0, \end{cases}$$

$$\bar{u} + \bar{v} = \bar{z}, \quad \bar{u} - \bar{v} = \bar{t},$$

transforms Eq. (24) to

$$ds^2 = (\bar{z}/8)(dx^2 + dy^2) + \left[\tfrac{1}{16}(\bar{t}/8)^{-1/2}\right](d\bar{z}^2 - d\bar{t}^2), \quad \bar{t} \geqslant \bar{z},$$
$$ds^2 = (\bar{t}/8)(dx^2 + dy^2) + \left[\tfrac{1}{16}(\bar{z}/8)^{-1/2}\right](d\bar{z}^2 - d\bar{t}^2), \quad \bar{t} \leqslant \bar{z}, \tag{25}$$

which is clearly equivalent to $g_{(i)}^{(+)}$ (with an appropriate choice of Ω_0^2) and also sends the R^4 manifold (x, y, z, t) to $M_{(i)}^{(+)}$.

The null hypersurface $z = t$ showed up explicitly in the equations $R_{\mu\nu} = 0$ for (23) as a hypersurface on which the equations could not be integrated. This seems to us to be more natural from a physical point of view than the arbitrary patching together of two solutions as described in the preceeding sections.

Finally, we would like to point out that one can add the physical requirement that the energy density be positive definite. Since $(M_{(i)}^{(+)}, g_{(i)}^{(+)})$ has negative energy density, it may be regarded as unphysical although one might interpret this to mean that the shock approaches the origin from infinity and converts the homogeneous portion of the space-time into the static portion. $(M_{(i)}^{(+)}, g_{(u)}^{(+)})$ has positive energy density. The singularity on $t = 0$ for all $z < 0$ and the singularity on $z = 0$ for all $t < 0$ seem to give rise to the shock on $z = t$. In this case we say that the shock is emitted from the origin and changes the static portion of the space–time into the homogeneous portion. (M, g) has positive energy density. The singularity on the two-dimensional surface $z = 0 = t$ seems to arise from two shocks traveling in opposite directions and crossing on the two-dimensional surface $z = 0 = t$; i.e., the shocks traveling toward the origin from the infinite past intersect at $z = 0 = t$, create a singularity there, and reemit shocks that escape to the infinite future.

APPENDIX: SPACE–TIMES WITH DISTRIBUTION-VALUED CURVATURE TENSORS (Taub[2])

The surfaces considered here are orientable hypersurfaces in the space-time across which the metric tensor is continuous but has finite jumps in its first and second derivatives.

Let Σ be such a hypersurface and let Σ be described locally by the equation

$$\psi(x) = 0$$

and have normal vector

$$l^\mu = \frac{\partial \psi}{\partial x^\mu} = \psi_{,\mu}.$$

In Fig. 10 we show a region Ω of the space–time divided by Σ into two parts $\Omega^+(\Omega^-)$ where $\psi > 0$ (< 0). Assume that the metric tensor is at least C^3 in $\Omega^{\pm}$ so that the Bianchi identities are satisfied in $\Omega^{\pm}$. Consider also functions f that are continuous in $\Omega^{\pm}$ but that may have a discontinuity across Σ denoted by

$$[f] = f_+ - f_-,$$

where f_+ (f_-) is the limit of the function f in Ω^+ (Ω^-) as a point in Ω^+ (Ω^-) approaches Σ. The discontinuities in the first and second derivatives of the metric tensor are then given by

$$[g_{\mu\nu,\sigma}] = l_\sigma b_{\mu\nu},$$
$$[(g_{\mu\nu,\sigma\tau})] = l_{\sigma,\tau} b_{\mu\nu} + l_\sigma b_{\mu\nu,\tau} + l_\tau b_{\mu\nu,\sigma} + l_\sigma l_\tau \hat{b}_{\mu\nu}. \tag{26}$$

If we restrict our attention at coordinate transformations that are C^3 or better, the jumps on the left-hand sides of (26) will transform like tensors since the jumps in the nontensorial parts will be zero. Hence $b_{\mu\nu}$ and $\hat{b}_{\mu\nu}$ are tensors.

One can now calculate the following equations:

$$b = g^{\mu\nu} b_{\mu\nu} = b^\sigma_\sigma,$$

$$b'_{\mu\nu} = b_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} b,$$

$$2[\Gamma^\alpha_{\beta\gamma}] = l_\beta b^\alpha_\gamma + l_\gamma b^\alpha_\beta - l^\alpha b_{\beta\gamma},$$

$$\bar{\Gamma}^\rho_{\alpha\mu} = \tfrac{1}{2}(\Gamma^\rho_{+\,\alpha\mu} + \Gamma^\rho_{-\,\alpha\mu}),$$

$$2A^\rho_{\sigma\mu} = b^\rho_{\sigma;\mu} + b^\rho_{\mu;\sigma} - b_{\sigma\mu}{}^{;\rho} + l_\mu \hat{b}^\rho_\sigma + l_\sigma \hat{b}^\rho_\mu - l^\rho \hat{b}_{\sigma}\mu,$$

where the semicolon denotes covariant differentiation with respect to $\bar{\Gamma}^\alpha_{\beta\gamma}$.

The distribution-valued curvature tensor may now be calculated by making use of the function

$$\theta(\psi) = \begin{cases} 1, & \psi > 0 \\ \tfrac{1}{2}, & \psi = 0, \\ 0, & \psi < 0, \end{cases}$$

whose derivative is

$$\theta_{,\mu} = \frac{\partial \psi}{\partial x^\mu} \frac{\partial \theta}{\partial \psi} = l_\mu \delta(\psi),$$

where $\delta(\psi)$ is the Dirac delta function defined in the usual way,

$$\int_\Omega F \delta \psi \, d^4 x = \int_{\delta\Omega^\cdot} F \, dV = \int_{\delta\Omega^\cdot} F \, d_3 V,$$

where dV is the invariant volume element induced on Σ. Distributions are defined by equations such as

$$f^D = \theta f^+ + (1-\theta) f^- ,$$

where $a(\pm)$ restricts f to $\Omega^{\pm}$. For vector and tensor distributions, covariant differentiation is defined by equations analogous to

$$(T^\mu)^D{}_{;\nu} = (T^\mu)^D{}_{,\nu} + (\Gamma^\mu_{\nu\rho})^D (T^\rho)^D .$$

One then finds that

$$(T^\rho)^D{}_{;\mu\nu}\big|_+ - (T^\rho)^D{}_{;\mu\nu}\big|_- = -(T^\sigma)^D Q^\rho{}_{\sigma\mu\nu} ,$$

where

$$Q^\rho{}_{\sigma\mu\nu} = \delta(\psi) H^\rho{}_{\sigma\mu\nu} + (R^\rho{}_{\sigma\mu\nu})^D - \theta(1-\theta) J^\rho{}_{\sigma\mu\nu} ,$$

$$2H^\rho{}_{\sigma\mu\nu} = b^\rho_\nu l_\sigma l_\mu - b^\rho_\mu l_\sigma l_\nu - b_{\sigma\nu} l^\rho l_\mu - b_{\sigma\mu} l^\rho l_\nu ,$$

$$J^\rho{}_{\sigma\mu\nu} = [\Gamma^\tau{}_{\sigma\nu}][\Gamma^\rho{}_{\tau\mu}] - [\Gamma^\tau{}_{\sigma\mu}][\Gamma^\rho{}_{\tau\nu}] .$$

Consider

$$B^{\rho\sigma\mu\nu} = \tfrac{1}{2}(-g)^{-1/2} g^{\sigma\tau} Q^\rho{}_{\gamma\alpha\beta} \epsilon^{\mu\nu\alpha\beta} .$$

One then finds that

$$B^{\rho\sigma\mu\nu}{}_{;\nu} = -(-g)^{-1/2} g^{\sigma\tau} \theta(1-\theta) \epsilon^{\mu\alpha\beta\nu} l_\alpha \{A^\gamma{}_{\tau\beta}[\Gamma^\rho{}_{\gamma\nu}] - A^\rho{}_{\gamma\beta}[\Gamma^\gamma{}_{\sigma\nu}]\} . \qquad (27)$$

Hence the Bianchi identities are satisfied on Σ if and only if

$$\epsilon^{\mu\alpha\beta\nu} l_\alpha \{A^\gamma{}_{\sigma\beta}[\Gamma^\rho{}_{\gamma\nu}] - A^\rho{}_{\gamma\beta}[\Gamma^\gamma{}_{\sigma\nu}]\} = 0 . \qquad (28)$$

Of course it would be pleasing if the Bianchi identities were satisfied on Σ, but because of Eq. (27) we need not require Eq. (28) since from the point of view of integration

$$\int_\Omega B^{\rho\sigma\mu\nu}{}_{;\nu}\, d^4x = 0;$$

i.e., the Bianchi identities are satisfied under the integral sign.

ACKNOWLEDGMENTS

We thank the referee for pointing out the null pressureless fluid interpretation.

[1]A.H. Taub, Ann. Math. (N.Y.) **53**, 472 (1951).
[2]A.H. Taub, J. Math. Phys. **21**, 1423 (1980).
[3]G.T. Carlson, Jr. and J.L. Safko, J. Math. Phys. **19**, 1617 (1978).
[4]T.M. Davis and J.R. Ray, J. Math. Phys. **17**, 1049 (1976).
[5]J. Horsky, P. Lorenc, and J. Novotny, Phys. Lett. A **63**, No. 2, 79 (1977).
[6]G.T. Carlson, Jr., Master's Thesis (unpublished), "*Plane-Symmetric Vacuum Fields Which Admit Time-Like Reflection Invariance in Harmonic Coordinates*," University of South Carolina Library (Spring 1977).

Einstein–Cartan theory in the spin coefficient formalism [a)]

R. Gambini [a)]
Departamento de Física, Universidad Simón Bolívar Apartado Postal 80654, Caracas 108, Venezuela
L. Herrera [a)]
Departmento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Caracas, Venezuela

(Received 13 June 1979; accepted for publication 1 February 1980)

The field equations of the Einstein–Cartan theory are written down using the spin coefficient formalism developed by Newman and Penrose for the Einstein theory. The irreducible spinor decomposition of the Riemann tensor in a U^4 space is obtained.

I. INTRODUCTION

The Newman–Penrose formalism[1] has been very fruitful, not only in the study of gravitational radiation but in finding exact solutions for the Einstein, Einstein–Maxwell, and Einstein–Weyl field equations.

Our purpose in this note is to present the field equations of the Einstein–Cartan[2,3] theory following closely the program outlined by Newman and Penrose for the canonical Einstein theory.

The motivation to undertake such a work is twofold:

(a) The peculiar structure of the equations allows us to deeper understand the role of the torsion as a dynamical variable as well as a geometric quantity.

(b) We hoped that solutions for the Einstein–Cartan–Maxwell,[4] or Einstein–Cartan–Weyl field equations would be easier to find in the context of the present formalism.

In the second part of the paper we give a decomposition for the Riemann tensor in its irreducible parts, which we believe is a new result, for the case of nonsymmetric affine connection.

In the third part the equations are obtained by a procedure already used by Papapetrou[5] for the symmetric connection case.

Finally, in Appendix B, we give explicitly all the field equations.

II. RIEMANN TENSOR DECOMPOSITION

As is well known, a U^4 space is a fourth-dimensional differential manifold with a nonsymmetric, metric, connection. That is,

$$\nabla_\mu g_{\nu\alpha} = \partial_\mu g_{\nu\alpha} - \Gamma^{\beta}_{\mu\nu} g_{\beta\alpha} - \Gamma^{\beta}_{\mu\alpha} g_{\nu\beta} = 0, \tag{2.1}$$

$$\Gamma^{\beta}_{\mu\nu} = g^{\beta\alpha} \Delta^{\gamma\delta\rho}_{\mu\nu\alpha} (\tfrac{1}{2}\partial_\gamma g_{\delta\rho} - g_{\rho\epsilon} S^{\epsilon}_{\gamma\delta}), \tag{2.2}$$

where

$S^{\epsilon}_{\gamma\delta} = \frac{1}{2}(\Gamma^{\epsilon}_{\gamma\delta} - \Gamma^{\epsilon}_{\delta\gamma})$ is the torsion tensor and $\Delta^{\gamma\delta\rho}_{\mu\nu\alpha}$ is the permutation tensor defined by

$$\Delta^{\gamma\delta\rho}_{\mu\nu\alpha} = \delta^\gamma_\mu \delta^\delta_\nu \delta^\rho_\alpha + \delta^\gamma_\nu \delta^\delta_\alpha \delta^\rho_\mu - \delta^\gamma_\alpha \delta^\delta_\mu \delta^\rho_\nu. \tag{2.3}$$

Also, for the Riemann curvature tensor one has the following identities[6]:

$$R_{(\nu\mu)\lambda\chi} = R_{\nu\mu(\lambda\chi)} = 0, \tag{2.4a}$$

$$R^{\chi}_{\ [\nu\mu\lambda]} = 2\nabla_{[\nu} S^{\chi}_{\mu\lambda]} - 4S^{\rho}_{[\nu\mu} S^{\chi}_{\lambda]\rho}, \tag{2.4b}$$

$$R_{\lambda\chi\nu\mu} - R_{\nu\mu\lambda\chi} = -\tfrac{3}{2}[R_{[\chi\lambda\nu]\mu} + R_{[\chi\mu\nu]\lambda} + R_{[\mu\lambda\nu]\chi} + R_{[\lambda\chi\mu]\nu}]. \tag{2.4c}$$

It will be convenient to introduce the spinor associated with the Riemann tensor:

$$R_{A\dot{W}B\dot{X}C\dot{Y}D\dot{Z}} = R_{\nu\mu\lambda\chi} \sigma^{\nu}_{A\dot{W}} \sigma^{\mu}_{B\dot{X}} \sigma^{\lambda}_{C\dot{Y}} \sigma^{\chi}_{D\dot{Z}}, \tag{2.5}$$

where the capital letters denote the spinor indices and the σs are the usual connecting quantities.[1]

Taking (2.4) into account, one can write:

$$R_{A\dot{W}B\dot{X}C\dot{Y}D\dot{Z}} = B_{ABCD}\epsilon_{\dot{W}\dot{X}}\epsilon_{\dot{Y}\dot{Z}} + \overline{B}_{\dot{W}\dot{X}\dot{Y}\dot{Z}}\epsilon_{AB}\epsilon_{CD} + C_{AB\dot{Y}\dot{Z}}\epsilon_{CD}\epsilon_{\dot{W}\dot{X}} + \overline{C}_{\dot{W}\dot{X}CD}\epsilon_{AB}\epsilon_{\dot{Y}\dot{Z}}, \tag{2.6}$$

where bars denote complex conjugate,

$$[\epsilon_{AB}] = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

and

$$B_{ABCD} = \tfrac{1}{4} R_{A\dot{P}B}{}^{\dot{P}}{}_{C\dot{Q}D}{}^{\dot{Q}} = B_{(AB)(CD)}, \tag{2.7a}$$

$$C_{AB\dot{Y}\dot{Z}} = \tfrac{1}{4} R_{A\dot{P}B}{}^{\dot{P}}{}_{G\dot{Y}}{}^{G}{}_{\dot{Z}} = C_{(AB)(\dot{Y}\dot{Z})}, \tag{2.7b}$$

$$\overline{C}_{AB\dot{Y}\dot{Z}} = \tfrac{1}{4} R_{G\dot{Y}}{}^{G}{}_{\dot{Z}A\dot{P}B}{}^{\dot{P}} = \overline{C}_{(AB)(\dot{Y}\dot{Z})}. \tag{2.7c}$$

Let us now introduce the following decomposition of the 4-spinors B and C:

$$B_{ABCD} = S_{ABCD} + A_{ABCD}, \quad C_{AB\dot{Y}\dot{Z}} = T_{AB\dot{Y}\dot{Z}} + D_{AB\dot{Y}\dot{Z}}, \tag{2.8}$$

with

$$S_{ABCD} = \tfrac{1}{2}[B_{ABCD} + B_{CDAB}] = S_{CDAB}, \tag{2.9a}$$

$$A_{ABCD} = \tfrac{1}{2}[B_{ABCD} - B_{CDAB}] = -A_{CDAB}, \tag{2.9b}$$

$$T_{AB\dot{Y}\dot{Z}} = \tfrac{1}{2}[C_{AB\dot{Y}\dot{Z}} + \overline{C}_{AB\dot{Y}\dot{Z}}] = \overline{T}_{AB\dot{Y}\dot{Z}}, \tag{2.9c}$$

$$D_{AB\dot{Y}\dot{Z}} = \tfrac{1}{2}[C_{AB\dot{Y}\dot{Z}} - \overline{C}_{AB\dot{Y}\dot{Z}}] = -D_{AB\dot{Y}\dot{Z}}. \tag{2.9d}$$

The 4-spinors S and A can be further reduced. In fact S may be written in terms of its totally symmetric part:

$$S_{ABCD} = S_{(ABCD)} + \tfrac{1}{6}(\epsilon_{BC}\epsilon_{AD} + \epsilon_{BD}\epsilon_{AC})S, \tag{2.10}$$

with

$$S = S^{GH}_{GH},$$

and A can be expressed as

$$A_{ABCD} = \tfrac{1}{2}(\epsilon_{BC}A_{AD} + \epsilon_{AD}A_{CB}), \tag{2.11}$$

where

$$A_{AB} = A_{AG}{}^{G}{}_{B} = A_{BA}.$$

From all the expressions above, one obtains

$$R_{A\dot{W}B\dot{X}C\dot{Y}D\dot{Z}} = S_{(ABCD)}\epsilon_{\dot{W}\dot{X}}\epsilon_{\dot{Y}\dot{Z}} + S_{(\dot{W}\dot{X}\dot{Y}\dot{Z})}\epsilon_{AB}\epsilon_{CD} + \tfrac{1}{6}(\epsilon_{BC}\epsilon_{AD} + \epsilon_{BD}\epsilon_{AC})\epsilon_{\dot{W}\dot{X}}\epsilon_{\dot{Y}\dot{Z}}S$$

a) This research was supported in part by the Consejo Nacional de Investigaciones Científicas y Tecnológicas de Venezuela.

 0022-2488/80/061449-06$1.00

$$+\tfrac{1}{6}(\epsilon_{\dot X\dot Y}\epsilon_{\dot W\dot Z}+\epsilon_{\dot X\dot Z}\epsilon_{\dot W\dot Y})\epsilon_{AB}\epsilon_{CD}\bar S$$
$$+T_{AB\dot Y\dot Z}\epsilon_{CD}\epsilon_{\dot W\dot X}+T_{CD\dot W\dot X}\epsilon_{AB}\epsilon_{\dot Y\dot Z}$$
$$+D_{AB\dot Y\dot Z}\epsilon_{CD}\epsilon_{\dot W\dot X}-D_{CD\dot W\dot X}\epsilon_{AB}\epsilon_{\dot Y\dot Z}$$
$$+\tfrac{1}{2}(\epsilon_{BC}A_{AD}+\epsilon_{AD}A_{CB})\epsilon_{\dot W\dot X}\epsilon_{\dot Y\dot Z}$$
$$+\tfrac{1}{2}(\epsilon_{\dot X\dot Y}A_{\dot W\dot Z}+\epsilon_{\dot W\dot Z}A_{\dot Y\dot X})\epsilon_{AB}\epsilon_{CD}. \qquad (2.12)$$

It can be easily seen that the total number of real components for $R_{A\dot WB\dot XC\dot YD\dot Z}$ is 36, as should be. In fact one has 5 complex (10 real) components for $S_{(ABCD)}$, 1 complex (2 real) for S, 3 complex plus 3 real (9 real) for $T_{AB\dot Y\dot Z}$, 3 complex plus 3 pure imaginary (9 real) for D, and 3 complex (6 real) for A_{AB}.

Let us now see which are the tensor equivalents to each irreducible spinor component.

We start with

$$C_{A\dot WB\dot XC\dot YD\dot Z}=S_{(ABCD)}\epsilon_{\dot W\dot X}\epsilon_{\dot Y\dot Z}+S_{(\dot W\dot X\dot Y\dot Z)}\epsilon_{AB}\epsilon_{CD}.$$

As is well known,[1] its tensor equivalent $C_{\mu\nu\lambda\chi}$ has the symmetries of the Weyl tensor; thus

$$C_{\mu\nu\lambda\chi}=-C_{\nu\mu\lambda\chi}=-C_{\mu\nu\chi\lambda}=C_{\lambda\chi\mu\nu},$$
$$C^{\chi}_{\ \mu\chi\lambda}=C_{[\mu\nu\lambda]\chi}=0.$$

As far as we know, the role of this tensor in the projective and conformal structure of space–time has not been studied when torsion is present.

Next, for the spinor $T_{AC\dot W\dot Y}$ one has

$$(R_{(\mu\nu)}-\tfrac{1}{4}g_{\mu\nu}R)\sigma^{\mu}_{A\dot W}\sigma^{\nu}_{C\dot Y}=2T_{AC\dot W\dot Y},$$

where

$$R=-2(S+\bar S)=-4\,\mathrm{Re}S,$$

and for the spinor A_{AC}

$$R_{[\mu\nu]}\sigma^{\mu}_{A\dot W}\sigma^{\nu}_{C\dot Y}=\epsilon_{\dot Y\dot W}A_{AC}+\epsilon_{CA}A_{\dot W\dot Y}.$$

Finally the remaining spinor components are related with the pseudotensor

$$D_{\mu\rho}=\tfrac{1}{2}\epsilon^{\alpha\beta\gamma}_{\mu}R_{\alpha\beta\gamma\rho}=3\epsilon^{\alpha\beta\gamma}_{\mu}R_{[\alpha\beta\gamma]\rho}.$$

The second identity (2.4b) allows us to write $D_{\mu\rho}$ only in terms of the torsion tensor and therefore $D_{\mu\rho}$ vanish identically when the connection is symmetric.

The trace of $D_{\mu\nu}$ is given by

$$D=D^{\rho}_{\ \rho}=2i(S-\bar S)$$

and the tracefree symmetric part verifies

$$(D_{(\mu\rho)}-\tfrac{1}{4}g_{\mu\rho}D)\sigma^{\mu}_{E\dot V}\sigma^{\rho}_{D\dot Z}=-2iD_{DE\dot V\dot Z}.$$

Finally the antisymmetric part of $D_{\mu\nu}$ is related with the Ricci tensor by

$$R_{[\mu\rho]}=-\tilde D_{\mu\rho}=-\tfrac{1}{2}\epsilon^{\alpha\beta}_{\mu\rho}D_{\alpha\beta}.$$

Thus, one is led to the following expression for the Riemann tensor:

$$R_{\lambda\chi\nu\mu}=C_{\lambda\chi\nu\mu}-\tfrac{1}{2}(g_{\lambda\nu}R_{(\chi\mu)}-g_{\lambda\mu}R_{(\chi\nu)}+g_{\chi\mu}R_{(\lambda\nu)}-g_{\chi\nu}R_{(\lambda\mu)})-\tfrac{1}{6}R\,(g_{\lambda\mu}g_{\chi\nu}-g_{\lambda\nu}g_{\chi\mu})+\tfrac{1}{12}\epsilon_{\lambda\chi\nu\mu}D-\tfrac{1}{4}[\epsilon_{\lambda\nu\chi}{}^{\alpha}D_{\alpha\mu}+\epsilon_{\nu\chi\mu}{}^{\alpha}D_{\alpha\lambda}+\epsilon_{\chi\mu\lambda}{}^{\alpha}D_{\alpha\nu}+\epsilon_{\mu\lambda\nu}{}^{\alpha}D_{\alpha\chi}].$$

III. FIELD EQUATIONS

As in the usual Einstein theory, the spin coefficient formalism consist of three sets of equations for the three dynamical variables: (a) The components of the Riemann tensor decomposed in its irreducible parts, (b) The spin coefficients, (c) The field of tetrads from which the metric tensor is built.

Furthermore, we must consider the Einstein–Cartan Field Equations

$$G^{\mu\nu}\equiv R^{\mu\nu}-\tfrac{1}{2}g^{\mu\nu}R=k\Sigma^{\mu\nu},$$
$$T^{\mu\nu\rho}=g^{\mu\alpha}g^{\nu\beta}S_{\alpha\beta}{}^{\rho}+2g^{\rho[\mu}S^{\nu]\lambda}{}_{\lambda}=k\tau^{\mu\nu\rho}, \qquad (3.1)$$

where $\Sigma^{\mu\nu}$ is the asymmetric canonical energy momentum tensor and $\tau^{\mu\nu\rho}$ the spin angular momentum tensor.

Recently, a formalism which allows the interaction of electromagnetism with torsion was proposed[4]; in such a case the Einstein–Cartan field equations do not hold, but the spin coefficient formalism is still valid if Eqs. (3.1) are substituted by Eqs. 40–45 of Ref. 4.

In general the formalism we present here can be applied in any theory including torsion, using the corrresponding field equations.

The first group of equations (metric equations) are obtained from the relation

$$\Gamma_{abc\dot d}=\tfrac{1}{2}\sigma^{\mu\dot p}_{a}\partial_{c\dot d}\sigma_{\mu b\dot p}$$

(italic lower-case indices are diadics indices and take the value 0,1).

This equation leads, without difficulty, to

$$Z_{a\dot bc\dot d}{}^{\mu}\equiv\partial_{a\dot b}\sigma^{\mu}_{c\dot d}-\partial_{c\dot d}\sigma^{\mu}_{a\dot b}-\sigma^{\mu}_{p\dot d}\Gamma^{p}{}_{ca\dot b}-\sigma^{\mu}_{c\dot x}\bar\Gamma^{\dot x}{}_{\dot d\dot ba}+\sigma^{\mu}_{p\dot b}\Gamma^{p}{}_{ac\dot d}+\sigma^{\mu}_{a\dot x}\Gamma_{\dot b}{}^{\dot x}{}_{\dot dc}+S_{a\dot bc\dot d}{}^{\mu}=0, \qquad (3.2)$$

where

$$S_{a\dot bc\dot d}{}^{\mu}=\sigma^{\nu}_{a\dot b}\sigma^{\alpha}_{c\dot d}S^{\ \ \mu}_{\nu\alpha}.$$

Equations (3.2) are a first-order differential system for the σ's. The second group of equations will be obtained from the integrability conditions of this sytem. With this purpose we need to calculate

$$\partial_{f\dot e}\Gamma_{abc\dot d}-\partial_{c\dot d}\Gamma_{abf\dot e}.$$

One obtains

$$Y_{f\dot eabc\dot d}\equiv\partial_{f\dot e}\Gamma_{abc\dot d}-\partial_{c\dot d}\Gamma_{abf\dot e}-\Gamma_{rbc\dot d}\Gamma^{r}{}_{af\dot e}+\Gamma_{rbf\dot e}\Gamma^{r}{}_{ac\dot d}-\Gamma_{ab}{}^{r}{}_{\dot e}\Gamma_{rfc\dot d}+\Gamma_{ab}{}^{r}{}_{\dot d}\Gamma_{rcf\dot e}-\Gamma_{abc\dot x}\bar\Gamma^{\dot x}{}_{\dot d\dot ef}+\Gamma_{abf\dot x}\bar\Gamma^{\dot x}{}_{\dot e\dot dc}-\tfrac{1}{2}R_{f\dot ec\dot da}{}^{\dot p}{}_{b\dot p}+S_{f\dot ec\dot d}{}^{p}\Gamma_{abp}=0. \qquad (3.3)$$

Finally, to obtain the last group (the Bianchi identities) one must consider the integrability conditions of (3.3). To do that we have to calculate the expression

$$\epsilon^{l\dot sf\dot ed\dot br\dot x}\partial_{r\dot x}Y_{f\dot eacd\dot b},$$

where

$$\epsilon^{l\dot sf\dot ed\dot br\dot x}=i(\epsilon^{ld}\epsilon^{fr}\epsilon^{\dot s\dot x}\epsilon^{\dot e\dot b}-\epsilon^{lr}\epsilon^{fd}\epsilon^{\dot s\dot b}\epsilon^{\dot e\dot x}).$$

The last group of equations as well as the other two groups are given explicitly in Appendix B. For the Bianchi identities an algebraic computation system (REDUCE) has been used.[7]

ACKNOWLEDGMENT

We would like to thank Regina Llopis de Trias for her suggestions and assistance to the use of the algebraic computation system.

APPENDIX A

The following notation for the independent components of the Riemman tensor has been used. The quantities ψ_0, ψ_1, etc., Φ_{00}, etc, and Λ, respectively, related to the components of the Weyl tensor, Symmetric Ricci tensor, and scalar curvature are defined as usual.[1]

We have introduced new quantities related with the pseudotensor $D_{\mu\rho}$ as follows:

$$\phi_{01} = D_{00\dot{0}\dot{1}} = \tfrac{1}{2}iD_{(13)} = -\phi_{10},$$
$$\phi_{10} = D_{01\dot{0}\dot{0}} = \tfrac{1}{2}iD_{(14)} = -\phi_{01},$$
$$\phi_{12} = D_{01\dot{1}\dot{1}} = \tfrac{1}{2}iD_{(32)} = -\phi_{21},$$
$$\phi_{21} = D_{11\dot{0}\dot{1}} = \tfrac{1}{2}iD_{(42)} = -\phi_{12},$$
$$\phi_{02} = D_{00\dot{1}\dot{1}} = \tfrac{1}{2}iD_{33} = -\phi_{20},$$
$$\phi_{20} = D_{11\dot{0}\dot{0}} = \tfrac{1}{2}iD_{44} = -\phi_{02},$$
$$\phi_{00} = D_{00\dot{0}\dot{0}} = \tfrac{1}{2}iD_{11} = -\phi_{00},$$
$$\phi_{11} = D_{01\dot{0}\dot{1}} = \tfrac{1}{4}i(D_{(12)} + D_{(34)}) = -\phi_{11},$$
$$\phi_{22} = D_{11\dot{1}\dot{1}} = \tfrac{1}{2}iD_{22} = -\phi_{22},$$
$$\Sigma = (S - \bar{S})/12i = -D/24,$$
$$A_0 = A_{00} = iD_{[13]},$$
$$A_1 = A_{01} = -\tfrac{1}{2}i(D_{[12]} - D_{[34]}),$$
$$A_2 = A_{11} = iD_{[24]}.$$

APPENDIX B

We shall use the same notation as in Ref. 1 for the coordinates ($x^1 = u$, $x^2 = r$, $x^i{:}i = 3,4$), the differential operators $(\Delta, D, \delta, \bar\delta)$, the spin coefficients, and the metric functions (the tetrad). Of course the spin coefficients are now defined in terms of nonsymmetric connections.

The first group contains the following equations:

$$Z_{0\dot{0}0\dot{1}}{}^{(i)} \equiv D\xi^i - \epsilon\xi^i + \kappa X^i + \bar\xi^i\bar\epsilon - \sigma\bar\xi^i - \xi^i\bar\rho + S^{(i)}{}_{0\dot{0}0\dot{1}} = 0,$$
$$Z_{0\dot{0}0\dot{1}}{}^{(2)} \equiv D\omega - \omega\epsilon + U\kappa + \bar\Pi + \omega\bar\epsilon - \bar\omega\sigma - \omega\bar\rho + \beta + \bar\alpha + S^{(2)}{}_{0\dot{0}0\dot{1}} = 0,$$
$$Z_{0\dot{0}1\dot{1}}{}^{(i)} \equiv DX^i - \xi^i\Pi + X^i\epsilon - \bar\xi^i\bar\Pi + X^i\bar\epsilon - \bar\xi^i\tau - \xi^i\bar\tau + S^{(i)}{}_{0\dot{0}1\dot{1}} = 0,$$
$$Z_{0\dot{0}1\dot{1}}{}^{(2)} \equiv DU - \omega\Pi + U\epsilon - \bar\omega\bar\Pi + U\bar\epsilon + \gamma - \bar\omega\tau + \bar\gamma - \omega\bar\tau + S^{(2)}{}_{0\dot{0}1\dot{1}},$$
$$Z_{0\dot{1}1\dot{0}}{}^{(i)} \equiv \delta\bar\xi^i - \bar\delta\xi^i + (\beta - \bar\alpha)\bar\xi^i - (\bar\beta - \alpha)\xi^i + X^i(\bar\rho - \rho) + S^{(i)}{}_{0\dot{1}1\dot{0}} = 0,$$
$$Z_{0\dot{1}1\dot{0}}{}^{(2)} \equiv \delta\bar\omega - \bar\delta\omega - \mu + \bar\mu + (\beta - \bar\alpha)\omega - (\bar\beta - \alpha)\omega + U(\bar\rho - \rho) + S^{(2)}{}_{0\dot{1}1\dot{0}} = 0,$$
$$Z_{0\dot{1}1\dot{1}}{}^{(i)} \equiv \delta X^i - \Delta\xi^i + \xi^i(\gamma - \bar\gamma - \mu) - \overline{\lambda\xi^i} + X^i(\beta + \bar\alpha - \tau) + S^{(i)}{}_{0\dot{1}1\dot{1}} = 0,$$
$$Z_{0\dot{1}1\dot{1}}{}^{(2)} \equiv \delta U - \Delta\omega + \omega(\gamma - \bar\gamma - \mu) - \bar\lambda\omega + \bar\nu + U(\beta + \alpha - \tau) + S^{(2)}{}_{0\dot{1}1\dot{1}} = 0,$$
$$Z_{0\dot{0}0\dot{1}}{}^{(1)} \equiv \kappa + S^{(1)}{}_{0\dot{0}0\dot{1}} = 0,$$
$$Z_{0\dot{0}1\dot{1}}{}^{(1)} \equiv (\epsilon + \bar\epsilon) + S^{(1)}{}_{0\dot{0}1\dot{1}} = 0,$$
$$Z_{0\dot{1}1\dot{0}}{}^{(1)} \equiv (\bar\rho - \rho) + S^{(1)}{}_{0\dot{1}1\dot{0}} = 0,$$
$$Z_{0\dot{1}1\dot{1}}{}^{(1)} \equiv (\beta + \bar\alpha - \tau) + S^{(1)}{}_{0\dot{1}1\dot{1}} = 0.$$

For the second group of equations one has

$$Y_{0\dot{1}0\dot{1}1\dot{0}} \equiv \delta\alpha - \bar\delta\beta - (\mu\rho - \lambda\sigma) - \alpha\bar\alpha - \beta\bar\beta + 2\alpha\beta - \gamma(\rho - \bar\rho) - \epsilon(\mu - \bar\mu) + \Psi_2 - \Lambda - \Phi_{11} + \phi_{11} - i\Sigma + 2S_{0\dot{1}1\dot{0}}{}^{\rho}\Gamma_{00\rho} = 0,$$
$$Y_{0\dot{1}1\dot{1}1\dot{0}} \equiv \delta\lambda - \bar\delta\mu + (\rho - \bar\rho)\nu - (\mu - \bar\mu)\Pi - \mu(\alpha + \bar\beta) - \lambda(\bar\alpha - 3\beta) + \Psi_3 - \Phi_{21} + \phi_{21} - A_2/2 + 2S_{0\dot{1}1\dot{0}}{}^{\rho}\Gamma_{11\rho} = 0,$$
$$Y_{1\dot{1}1\dot{1}1\dot{0}} \equiv \Delta\lambda - \bar\delta\nu + (\mu + \bar\mu)\lambda + (3\gamma - \bar\gamma)\lambda - (3\alpha + \bar\beta + \Pi - \bar\tau)\nu + \Psi_4 + 2S_{1\dot{1}1\dot{0}}{}^{\rho}\Gamma_{11\rho} = 0,$$
$$Y_{0\dot{1}1\dot{1}1\dot{1}} \equiv \delta\nu - \Delta\mu - (\mu^2 + \lambda\bar\lambda) - (\gamma + \bar\gamma)\mu + \bar\nu\Pi - (\tau - 3\beta - \bar\alpha)\nu - \Phi_{22} + \phi_{22} + 2S_{0\dot{1}1\dot{1}}{}^{\rho}\Gamma_{11\rho} = 0,$$
$$Y_{0\dot{1}0\dot{1}1\dot{1}} \equiv \delta\gamma - \Delta\beta - (\tau - \bar\alpha - \beta)\gamma - \mu\tau + \sigma\nu + \epsilon\bar\nu + \beta(\gamma - \bar\gamma - \mu) - \alpha\bar\lambda - \Phi_{12} + \phi_{12} + 2S_{0\dot{1}1\dot{1}}{}^{\rho}\Gamma_{01\rho} = 0,$$
$$Y_{0\dot{0}0\dot{0}1\dot{0}} \equiv D\rho - \bar\delta\kappa - \rho^2 - \sigma\bar\sigma - \rho(\epsilon + \bar\epsilon) + \bar\kappa\tau + \kappa(3\alpha + \bar\beta - \Pi) - \Phi_{00} + \phi_{00} + 2S_{0\dot{0}1\dot{0}}{}^{\rho}\Gamma_{00\rho} = 0,$$
$$Y_{0\dot{0}0\dot{0}0\dot{1}} \equiv D\sigma - \delta\kappa - \sigma(\rho + \bar\rho) - \sigma(3\epsilon - \bar\epsilon)\sigma + (\tau - \bar\Pi + \bar\alpha + 3\beta)\kappa - \Psi_0 + 2S_{0\dot{0}0\dot{1}}{}^{\rho}\Gamma_{00\rho} = 0,$$
$$Y_{0\dot{0}0\dot{0}1\dot{1}} \equiv D\tau - \Delta\kappa - \rho(\tau + \bar\Pi) - (\bar\tau + \Pi)\sigma - (\epsilon - \bar\epsilon)\tau + (3\gamma + \bar\gamma)\kappa - \Psi_1 - \Phi_{01} + 2S_{0\dot{0}1\dot{1}}{}^{\rho}\Gamma_{00\rho} + \phi_{01} + \tfrac{1}{2}A_0 = 0,$$
$$Y_{0\dot{0}0\dot{1}1\dot{0}} \equiv D\alpha - \bar\delta\epsilon - (\rho + \bar\epsilon - 2\epsilon)\alpha - \beta\bar\sigma + \bar\beta\epsilon + \kappa\lambda + \bar\kappa\gamma - (\epsilon + \rho)\Pi - \Phi_{10} + 2S_{0\dot{0}1\dot{0}}{}^{\rho}\Gamma_{01\rho} + \phi_{10} = 0,$$
$$Y_{0\dot{0}0\dot{1}0\dot{1}} \equiv D\beta - \delta\epsilon - (\alpha + \Pi)\sigma - (\bar\rho - \bar\epsilon)\beta + (\mu + \gamma)\kappa + (\bar\alpha - \bar\Pi)\epsilon - \Psi_1 + 2S_{0\dot{0}0\dot{1}}{}^{\rho}\Gamma_{01\rho} - A_0/2 = 0,$$
$$Y_{0\dot{0}0\dot{1}1\dot{1}} \equiv D\gamma - \Delta\epsilon - (\tau + \bar\Pi)\alpha - (\bar\tau + \Pi)\beta + (\epsilon + \bar\epsilon)\gamma + (\gamma + \bar\gamma)\epsilon - \tau\Pi + \nu k - \Psi_2 + \Lambda + i\Sigma - \Phi_{11} + \phi_{11} + 2S_{0\dot{0}1\dot{1}}{}^{\rho}\Gamma_{01\rho} = 0,$$

$$Y_{0\dot{0}110\dot{1}} \equiv D\mu - \delta\Pi - (\bar\rho\mu + \sigma\lambda) - \Pi\bar\Pi + (\epsilon + \bar\epsilon)\mu + \Pi(\bar\alpha - \beta) + \nu\kappa - \Psi_2 - 2\Lambda + 2S_{0\dot{0}0\dot{1}}{}^{\rho}\Gamma_{11\rho} - 2i\Sigma - \Lambda_1 = 0,$$

$$Y_{0\dot{0}111\dot{1}} \equiv D\nu - \Delta\Pi - (\Pi + \bar\tau)\mu - (\bar\Pi + \tau)\lambda - (\gamma - \bar\gamma)\Pi + (3\epsilon + \bar\epsilon)\nu - \Psi_3 - \Phi_{21} + \phi_{21} - \Lambda_2/2 + 2S_{0\dot{0}1\dot{1}}{}^{\rho}\Gamma_{11\rho} = 0,$$

$$Y_{0\dot{1}001\dot{0}} \equiv \delta\rho - \bar\delta\sigma - \rho(\bar\alpha + \beta) + \sigma(3\alpha - \bar\beta) - (\rho - \bar\rho)\tau + (\mu - \bar\mu)\kappa + \Psi_1 - \Phi_{01} + 2S_{0\dot{1}1\dot{0}}{}^{\rho}\Gamma_{00\rho} + \phi_{01} - \Lambda_0/2 = 0,$$

$$Y_{0\dot{1}001\dot{1}} \equiv \delta\tau - \Delta\sigma - (\mu\sigma + \bar\lambda\rho) - (\tau + \beta - \bar\alpha)\tau + (3\gamma - \bar\gamma)\sigma + \kappa\bar\nu - \Phi_{02} + \phi_{02} + 2S_{0\dot{1}1\dot{1}}{}^{\rho}\Gamma_{00\rho} = 0,$$

$$Y_{1\dot{1}001\dot{0}} \equiv \Delta\rho - \bar\delta\tau + (\rho\bar\mu + \sigma\lambda) - (\bar\beta - \alpha - \bar\tau)\tau - (\gamma + \bar\gamma)\rho - \nu\kappa + \Psi_2 - \Lambda_1 + 2i\Sigma + 2\Lambda + 2S_{1\dot{1}1\dot{0}}{}^{\rho}\Gamma_{00\rho} = 0,$$

$$Y_{1\dot{1}011\dot{0}} \equiv \Delta\alpha - \bar\delta\gamma - (\rho + \epsilon)\nu + (\tau + \beta)\lambda - (\bar\gamma - \mu)\alpha - (\bar\beta - \bar\tau)\gamma + \Psi_3 - \Lambda_2/2 + 2S_{1\dot{1}1\dot{0}}{}^{\rho}\Gamma_{01\rho} = 0.$$

Finally the Bianchi Identities are

$$\begin{aligned}
&-2D\bar\Psi_3 - 2D\Phi_{12} + 2D\phi_{12} + D\bar\Lambda_2 - 2\Delta\Psi_1 - 2\Delta\Phi_{01} - 2\Delta\phi_{01} - \Delta\Lambda_0 + 2\delta\Psi_2 + 2\delta\bar\Psi_2 + 4\delta\Phi_{11} - 4\delta\Lambda + 2\Psi_0\nu \\
&-4\Psi_1 S^{3}_{23} - 4\Psi_1\mu + 4\Psi_1\gamma + 4\Psi_1 S_{12}{}^{1} - 4\bar\Psi_1\bar\lambda - 4\bar\Psi_1 S^{4}_{23} - 4\Psi_2 S^{2}_{23} + 4\Psi_2 S^{4}_{12} - 4\Psi_2 S^{1}_{13} - 6\Psi_2\tau \\
&-4\bar\Psi_2 S^{2}_{23} + 6\bar\Psi_2\bar\Pi - 4\bar\Psi_2 S^{4}_{12} - 4\bar\Psi_2 S^{1}_{13} + 4\Psi_3\sigma - 4\Psi_3 S^{4}_{13} - 4\bar\Psi_3 S^{2}_{12} - 4\bar\Psi_3\bar\epsilon - 4\bar\Psi_3 S^{3}_{13} \\
&+4\bar\Psi_3\bar\rho - 2\bar\Psi_4\bar\kappa + 2\Phi_{00}\bar\nu - 4\Phi_{01}S^{3}_{23} - 4\Phi_{01}\mu + 4\Phi_{01}\gamma + 4\Phi_{01}S^{1}_{12} - 4\Phi_{10}\bar\lambda - 4\Phi_{10}S^{4}_{23} + 2\Phi_{02}\Pi \\
&-2\Phi_{02}\bar\tau - 8\Phi_{11}S^{2}_{23} + 4\Phi_{11}\bar\Pi - 8\Phi_{11}S^{1}_{13} - 4\Phi_{11}\tau - 4\Phi_{12}S^{2}_{12} - 4\Phi_{12}\bar\epsilon - 4\Phi_{12}S^{3}_{13} + 4\Phi_{12}\bar\rho \\
&+4\Phi_{21}\sigma - 4\Phi_{21}S^{4}_{13} - 2\Phi_{22}\kappa + 2\phi_{00}\bar\nu - 4\phi_{01}S^{3}_{23} + 4\phi_{01}\gamma + 4\phi_{01}S_{12}{}^{1} + 4\phi_{10}S^{4}_{23} - 2\phi_{01}\Pi \\
&-2\phi_{02}\bar\tau - 4\phi_{11}\bar\Pi + 8\phi_{11}S^{4}_{12} - 4\phi_{11}\tau + 4\phi_{12}S^{2}_{12} + 4\phi_{12}\bar\epsilon + 4\phi_{12}S^{3}_{13} - 4\phi_{21}S^{4}_{13} \\
&+2\phi_{22}\kappa - 2\Lambda_0 S^{3}_{23} + 2\Lambda_0\gamma + 2\Lambda_0 S^{1}_{12} - 2\bar\Lambda_0 S^{4}_{23} - 2\Lambda_1\tau - 2\bar\Lambda_1\bar\Pi + 2\Lambda_2 S^{4}_{13} + 2\bar\Lambda_2 S^{2}_{12} \\
&+2\bar\Lambda_2\bar\epsilon + 2\bar\Lambda_2 S^{3}_{13} + 8\Lambda S^{2}_{23} + 8\Lambda S^{1}_{13} - 8\Sigma i S^{4}_{12} = 0,
\end{aligned}$$

$$\begin{aligned}
&-2D\Phi_{02} + 2D\phi_{02} - 2\Delta\Psi_0 + 2\delta\Psi_1 + 2\delta\Phi_{01} - 2\delta\phi_{01} - \delta\Lambda_0 - 4\Psi_0 S^{3}_{23} - 2\Psi_0\mu + 8\Psi_0\gamma + 4\Psi_0 S^{1}_{12} - 4\Psi_1\beta \\
&-4\Psi_1 S^{2}_{23} + 4\Psi_1 S^{4}_{12} - 4\Psi_1 S^{1}_{13} - 8\Psi_1\tau + 6\Psi_2\sigma - 4\Psi_2 S^{4}_{13} - 2\Phi_{00}\bar\lambda - 4\Phi_{00}S^{4}_{23} \\
&-4\Phi_{01}\beta - 4\Phi_{01}S^{2}_{23} + 4\Phi_{01}\bar\Pi - 4\Phi_{01}S^{4}_{12} - 4\Phi_{01}S^{1}_{13} - 4\Phi_{02}S^{2}_{12} + 4\Phi_{02}\epsilon - 4\Phi_{02}\bar\epsilon \\
&-4\Phi_{02}S^{3}_{13} + 2\Phi_{02}\bar\rho + 4\Phi_{11}\sigma - 4\Phi_{12}\kappa + 2\phi_{00}\bar\lambda + 4\phi_{00}S^{4}_{23} + 4\phi_{01}\beta + 4\phi_{01}S^{2}_{23} \\
&-4\phi_{01}\bar\Pi + 4\phi_{01}S^{4}_{12} + 4\phi_{01}S^{1}_{13} + 4\phi_{02}S^{2}_{12} - 4\phi_{02}\epsilon + 4\phi_{02}\bar\epsilon + 4\phi_{02}S^{3}_{13} - 2\phi_{02}\bar\rho \\
&-4\phi_{11}\sigma + 4\phi_{12}\kappa + 2\Lambda_0\beta + 2\Lambda_0 S^{2}_{23} - 2\Lambda_0 S^{4}_{12} + 2\Lambda_0 S^{1}_{13} - 2\Lambda_1\sigma + 4\Lambda_1 S^{4}_{13} \\
&-8\Lambda S^{4}_{13} - 8\Sigma i S^{4}_{13} = 0,
\end{aligned}$$

$$\begin{aligned}
&-2D\bar\Psi_2 + 2D\bar\Lambda_1 - 4D\Lambda + 4D\Sigma i - 2\Delta\Phi_{00} - 2\Delta\phi_{00} + 2\delta\bar\Psi_1 + 2\delta\Phi_{10} + 2\delta\phi_{10} - \delta\bar\Lambda_0 - 2\bar\Psi_0\bar\lambda \\
&-4\bar\Psi_0 S^{4}_{23} - 4\bar\Psi_1 S^{2}_{23} + 4\bar\Psi_1\bar\Pi - 4\bar\Psi_1\bar\alpha - 4\bar\Psi_1 S^{4}_{12} - 4\bar\Psi_1 S^{1}_{13} - 4\bar\Psi_2 S^{2}_{12} - 4\bar\Psi_2 S^{3}_{13} + 6\bar\Psi_2\bar\rho - 4\bar\Psi_3\bar\kappa \\
&-4\Phi_{00}S^{3}_{23} - 2\Phi_{00}\mu + 4\Phi_{00}\gamma + 4\Phi_{00}\bar\gamma + 4\Phi_{00}S^{1}_{12} - 4\Phi_{01}\bar\tau - 4\Phi_{10}S^{2}_{23} - 4\Phi_{10}\bar\alpha + 4\Phi_{10}S^{4}_{12} - 4\Phi_{10}S^{1}_{13} - 4\Phi_{10}\tau \\
&+2\Phi_{20}\sigma - 4\Phi_{20}S^{4}_{13} + 4\Phi_{11}\bar\rho - 4\phi_{00}S^{3}_{23} - 2\phi_{00}\mu + 4\phi_{00}\gamma + 4\phi_{00}\bar\gamma + 4\phi_{00}S^{1}_{12} \\
&-4\phi_{01}\bar\tau - 4\phi_{10}S^{2}_{23} - 4\phi_{10}\bar\alpha + 4\phi_{10}S^{4}_{12} - 4\phi_{10}S^{1}_{13} - 4\phi_{10}\tau + 2\phi_{20}\sigma \\
&-4\phi_{20}S^{4}_{13} + 4\phi_{11}\bar\rho + 2\bar\Lambda_0 S^{2}_{23} - 2\bar\Lambda_0\bar\Pi + 2\bar\Lambda_0\bar\alpha + 2\bar\Lambda_0 S^{4}_{12} + 2\bar\Lambda_0 S^{1}_{13} \\
&+4\bar\Lambda_1 S^{2}_{12} + 4\bar\Lambda_1 S^{3}_{13} - 2\bar\Lambda_1\bar\rho + 2\bar\Lambda_2\kappa - 8\Lambda S^{2}_{12} - 8\Lambda S^{3}_{13} + 8\Sigma S^{2}_{12}i + 8\Sigma S^{3}_{13}i = 0,
\end{aligned}$$

$$\begin{aligned}
&2D\bar\Psi_4 + 2\Delta\Phi_{02} + 2\Delta\phi_{02} - 2\delta\bar\Psi_3 - 2\delta\Phi_{12} - 2\delta\phi_{12} - \delta\bar\Lambda_2 + 6\bar\Psi_2\bar\lambda + 4\bar\Psi_2 S^{4}_{23} + 4\bar\Psi_3 S^{2}_{23} - 8\bar\Psi_3\bar\Pi - 4\bar\Psi_3\bar\alpha \\
&+4\bar\Psi_3 S^{4}_{12} + 4\bar\Psi_3 S^{1}_{13} + 4\bar\Psi_4 S^{2}_{12} + 8\bar\Psi_4\bar\epsilon + 4\bar\Psi_4 S^{3}_{12} - 2\bar\Psi_4\bar\rho - 4\Phi_{01}\bar\nu + 4\Phi_{02}S^{3}_{23} \\
&+2\Phi_{02}\mu - 4\Phi_{02}\gamma + 4\Phi_{02}\bar\gamma - 4\Phi_{02}S^{1}_{12} + 4\Phi_{11}\bar\lambda + 4\Phi_{12}S^{2}_{23} - 4\Phi_{12}\bar\alpha - 4\Phi_{12}S^{4}_{12} + 4\Phi_{12}S^{1}_{13} \\
&+4\Phi_{12}\tau - 2\Phi_{22}\sigma + 4\Phi_{22}S^{4}_{13} - 4\phi_{01}\bar\nu + 4\phi_{02}S^{3}_{23} + 2\phi_{02}\mu - 4\phi_{02}\gamma + 4\phi_{02}\bar\gamma \\
&-4\phi_{02}S^{1}_{12} + 4\phi_{11}\bar\lambda + 4\phi_{12}S^{2}_{23} - 4\phi_{12}\bar\alpha - 4\phi_{12}S^{4}_{12} + 4\phi_{12}S^{1}_{13} + 4\phi_{12}\tau - 2\phi_{22}\sigma \\
&+4\phi_{22}S^{4}_{13} + 2\bar\Lambda_1\bar\lambda + 4\bar\Lambda_1 S^{4}_{23} + 4\bar\Lambda_2\bar\alpha + 2\bar\Lambda_2 S^{4}_{12} + 2\bar\Lambda_2 S^{1}_{13} + \beta\Lambda S^{4}_{23} - \beta\Sigma i S^{4}_{23} = 0,
\end{aligned}$$

$$\begin{aligned}
&2D\Phi_{22} - 2D\phi_{22} + 2\Delta\Psi_2 + 2\Delta\Lambda_1 + 4\Delta\Lambda + 4\Delta\Sigma i - 2\delta\Psi_3 - 2\delta\Phi_{21} + 2\delta\phi_{21} - \delta\Lambda - 4\Psi_1\bar\nu + 4\Psi_2 S^{3}_{23} + 6\Psi_2\mu \\
&-4\Psi_2 S^{1}_{12} - 4\Psi_3\beta + 4\Psi_3 S^{2}_{23} - 4\Psi_3 S^{4}_{12} + 4\Psi_3 S^{1}_{13} + 4\Psi_3\tau - 2\Psi_4\sigma + 4\Psi_4 S^{4}_{13} \\
&+2\Phi_{20}\bar\lambda + 4\Phi_{20}S^{4}_{23} + 4\Phi_{11}\mu - 4\Phi_{12}\Pi - 4\Phi_{21}\beta + 4\Phi_{21}S^{2}_{23} - 4\Phi_{21}\bar\Pi + 4\Phi_{21}S^{4}_{12} + 4\Phi_{21}S^{1}_{13} \\
&+4\Phi_{22}S^{2}_{12} + 4\Phi_{22}\epsilon + 4\Phi_{22}\bar\epsilon + 4\Phi_{22}S^{3}_{13} - 2\Phi_{22}\bar\rho - 2\phi_{20}\bar\lambda - 4\phi_{20}S^{4}_{23} - 4\phi_{11}\mu + 4\phi_{12}\Pi \\
&+4\phi_{21}\beta - 4\phi_{21}S^{2}_{23} + 4\phi_{21}\bar\Pi - 4\phi_{21}S^{4}_{12} - 4\phi_{21}S^{1}_{13} - 4\phi_{22}S^{2}_{12} - 4\phi_{22}\epsilon - 4\phi_{22}\bar\epsilon
\end{aligned}$$

$$-4\phi_{22}S^{3}_{13}+2\phi_{22}\bar{\rho}-2\Lambda_{0}\nu+4\Lambda_{1}S^{3}_{23}+2\Lambda_{1}\mu-4\Lambda_{1}S^{1}_{12}-2\Lambda_{2}\beta+2\Lambda_{2}S^{2}_{23}$$
$$-2\Lambda_{2}S^{4}_{12}+2\Lambda_{2}S^{1}_{13}+2\Lambda_{2}\tau+8\Lambda S^{3}_{23}-8\Lambda S^{1}_{12}+8\Sigma iS^{3}_{23}-8\Sigma iS^{1}_{12}=0,$$

$$+2D\bar{\Psi}_{3}+2D\Phi_{12}-2D\phi_{12}+D\bar{\Lambda}_{2}+2\Delta\Psi_{1}-2\Delta\Phi_{01}-2\Delta\phi_{01}+\Delta\Lambda_{0}-2\delta\Psi_{2}+2\delta\bar{\Psi}_{2}+4\Delta\phi_{11}+4\Delta\Sigma i-2\Psi_{0}\nu$$
$$+4\Psi_{1}S^{3}_{23}+4\Psi_{1}\mu-4\Psi_{1}\gamma-4\Psi_{1}S^{1}_{12}-4\bar{\Psi}_{1}\bar{\lambda}-4\bar{\Psi}_{1}S^{4}_{23}+4\Psi_{2}S^{2}_{23}-4\Psi_{2}S^{4}_{12}+4\Psi_{2}S^{1}_{13}+6\Psi_{2}\tau-4\bar{\Psi}_{2}S^{2}_{23}$$
$$+6\bar{\Psi}_{2}\bar{\Pi}-4\bar{\Psi}_{2}S^{4}_{12}-4\bar{\Psi}_{2}S^{1}_{13}-4\Psi_{3}\sigma+4\Psi_{3}S^{4}_{13}-4\bar{\Psi}_{3}S^{2}_{12}-4\bar{\Psi}_{3}\bar{\epsilon}-4\bar{\Psi}_{3}S^{3}_{13}+4\bar{\Psi}_{3}\bar{\rho}-2\bar{\Psi}_{4}\bar{\kappa}+2\Phi_{00}\bar{\nu}-4\Phi_{01}S^{3}_{23}$$
$$+4\Phi_{01}\gamma+4\Phi_{01}S^{1}_{12}+4\Phi_{10}S^{4}_{23}-2\Phi_{02}\Pi-2\Phi_{02}\bar{\tau}-4\Phi_{11}\bar{\Pi}+8\Phi_{11}S^{4}_{12}-4\Phi_{11}\tau+4\Phi_{12}S^{2}_{12}+4\Phi_{12}\bar{\epsilon}$$
$$+4\Phi_{12}S^{3}_{13}-4\Phi_{21}S^{4}_{13}+2\Phi_{22}\kappa+2\phi_{00}\bar{\nu}-4\phi_{01}S^{3}_{23}-4\phi_{01}\mu+4\phi_{01}\gamma+4\phi_{01}S^{1}_{12}-4\phi_{10}\bar{\lambda}-4\phi_{10}S^{4}_{23}$$
$$+2\phi_{02}\Pi-2\phi_{02}\bar{\tau}-8\phi_{11}S^{2}_{23}+4\phi_{11}\bar{\Pi}-8\phi_{11}S^{1}_{13}-4\phi_{11}\tau-4\phi_{12}S^{2}_{12}-4\phi_{12}\bar{\epsilon}-4\phi_{12}S^{3}_{13}+4\phi_{12}\bar{\rho}+4\phi_{21}\sigma$$
$$-4\phi_{21}S^{4}_{13}-2\phi_{22}\kappa+2\Lambda_{0}S^{3}_{23}-2\Lambda_{0}\gamma-2\Lambda_{0}S^{1}_{12}-2\bar{\Lambda}_{0}S^{4}_{23}+2\Lambda_{1}\tau-2\bar{\Lambda}_{1}\bar{\Pi}-2\Lambda_{2}S^{4}_{13}+2\bar{\Lambda}_{2}S^{2}_{12}+2\bar{\Lambda}_{2}\bar{\epsilon}$$
$$+2\bar{\Lambda}_{2}S^{3}_{13}+8\Lambda S^{4}_{12}-8\Sigma S^{2}_{23}i-8\Sigma iS^{1}_{13}=0,$$

$$-2D\Psi_{2}+2D\bar{\Psi}_{2}-4D\phi_{11}+4D\Sigma i-2\delta\bar{\Psi}_{1}-2\delta\Phi_{10}+2\delta\phi_{10}-\delta\bar{\Lambda}_{0}+2\bar{\delta}\Psi_{1}+2\bar{\delta}\Phi_{01}+2\bar{\delta}\phi_{01}+\delta\Lambda_{0}-2\Psi_{0}\lambda$$
$$+2\bar{\Psi}_{0}\bar{\lambda}-4\Psi_{1}\alpha-4\Psi_{1}S^{3}_{34}-4\Psi_{1}S^{1}_{14}+4\Psi_{1}\Pi-4\bar{\Psi}_{1}\bar{\Pi}-4\bar{\Psi}_{1}S^{4}_{34}+4\bar{\Psi}_{1}\bar{\alpha}+4\bar{\Psi}_{1}S^{1}_{13}-4\Psi_{2}S^{4}_{14}-4\Psi_{2}S^{3}_{13}+6\Psi_{2}\rho$$
$$-4\Psi_{2}S^{2}_{34}+4\bar{\Psi}_{2}S^{4}_{14}+4\Psi_{2}S^{3}_{13}-6\bar{\Psi}_{2}\bar{\rho}-4\bar{\Psi}_{2}S^{2}_{34}-4\Psi_{3}S^{2}_{13}-4\Psi_{3}\kappa+4\bar{\Psi}_{3}\bar{\kappa}+4\bar{\Psi}_{3}S^{2}_{14}+2\Phi_{00}\mu-2\Phi_{00}\bar{\mu}-4\Phi_{01}\alpha$$
$$-4\Phi_{01}S^{3}_{34}-4\Phi_{01}S^{1}_{14}-4\Phi_{10}S^{4}_{34}+4\Phi_{10}\bar{\alpha}+4\Phi_{10}S^{1}_{13}+2\Phi_{02}\bar{\sigma}-2\Phi_{20}\sigma+4\Phi_{11}\rho-4\Phi_{11}\bar{\rho}-8\Phi_{11}S^{2}_{34}+4\Phi_{12}S^{2}_{14}$$
$$-4\Phi_{21}S^{2}_{13}-2\phi_{00}\mu-2\phi_{00}\bar{\mu}-4\phi_{01}\alpha-4\phi_{01}S^{3}_{34}-4\phi_{01}S^{1}_{14}+4\phi_{01}\Pi+4\phi_{10}\bar{\Pi}+4\phi_{10}S^{4}_{34}-4\phi_{10}\bar{\alpha}-4\phi_{10}S^{1}_{13}+2\phi_{02}\bar{\sigma}$$
$$+2\phi_{20}\sigma-8\phi_{11}S^{4}_{14}-8\phi_{11}S^{3}_{13}+4\phi_{11}\rho+4\phi_{11}\bar{\rho}-4\phi_{12}\bar{\kappa}-4\phi_{12}S^{2}_{14}-4\phi_{21}S^{2}_{13}-4\phi_{21}\bar{\kappa}-2\Lambda_{0}\alpha-2\Lambda_{0}S^{3}_{34}-2\Lambda_{0}S^{1}_{14}$$
$$-2\bar{\Lambda}_{0}S^{4}_{34}+2\bar{\Lambda}_{0}\bar{\alpha}+2\bar{\Lambda}_{0}S^{1}_{13}+2\Lambda_{1}\rho-2\bar{\Lambda}_{1}\bar{\rho}+2\Lambda_{2}S^{2}_{13}-2\bar{\Lambda}_{2}S^{2}_{14}+8\Lambda S^{2}_{34}+8\Sigma S^{4}_{14}i+8\Sigma S^{3}_{13}i=0,$$

$$-2D\Psi_{1}+2D\Phi_{01}-2D\phi_{01}+D\Lambda_{0}-2\delta\Phi_{00}+2\delta\phi_{00}+2\bar{\delta}\Psi_{0}-8\Psi_{0}\alpha-4\Psi_{0}S^{3}_{34}-4\Psi_{0}S^{1}_{14}+4\Psi_{1}\epsilon+2\Psi_{0}\Pi$$
$$-4\Psi_{1}S^{4}_{14}-4\Psi_{1}S^{3}_{13}+8\Psi_{1}\rho-4\Psi_{1}S^{2}_{34}-4\Psi_{2}S^{2}_{13}$$
$$-6\Psi_{2}\kappa+4\Phi_{00}\beta-2\Phi_{00}\bar{\Pi}-4\Phi_{00}S^{4}_{34}+4\Phi_{00}\bar{\alpha}+4\Phi_{00}S^{1}_{13}+4\Phi_{01}S^{4}_{14}$$
$$-4\Phi_{01}\epsilon+4\Phi_{01}S^{3}_{13}-4\Phi_{01}\bar{\rho}-4\Phi_{01}S^{2}_{34}-4\Phi_{10}\sigma+2\Phi_{02}\bar{\kappa}+4\Phi_{02}S^{2}_{14}+4\Phi_{11}\bar{\kappa}-4\phi_{00}\beta+2\phi_{00}\bar{\Pi}+4\phi_{00}S^{4}_{34}-4\phi_{00}\bar{\alpha}$$
$$-4\phi_{00}S^{1}_{13}-4\phi_{01}S^{4}_{14}+4\phi_{01}\epsilon-4\phi_{01}S^{3}_{13}+4\phi_{01}\bar{\rho}+4\phi_{01}S^{2}_{34}+4\phi_{10}\sigma-4\phi_{02}\bar{\kappa}-4\phi_{02}S^{2}_{14}+4\phi_{11}\kappa$$
$$+2\Lambda_{0}S^{4}_{14}-2\Lambda_{0}\epsilon+2\Lambda_{0}S^{3}_{13}+2\Lambda_{0}S^{2}_{34}+4\Lambda_{1}S^{2}_{13}+2\Lambda_{1}\kappa-8\Lambda S^{2}_{13}-8\Sigma S^{2}_{13}i=0,$$

$$-2D\bar{\Psi}_{3}+2D\Phi_{12}+2D\phi_{12}-D\bar{\Lambda}_{2}+2\delta\bar{\Psi}_{2}+2\delta\bar{\Lambda}_{1}+4\delta\Lambda-4\delta\Sigma i-2\bar{\delta}\Phi_{02}-2\bar{\delta}\phi_{02}-4\bar{\Psi}_{1}\bar{\lambda}+6\bar{\Psi}_{2}\bar{\Pi}+4\bar{\Psi}_{2}S^{4}_{34}$$
$$-4\bar{\Psi}_{2}S^{1}_{13}-4\bar{\Psi}_{3}S^{4}_{14}-4\bar{\Psi}_{3}\bar{\epsilon}-4\bar{\Psi}_{3}S^{3}_{13}+4\bar{\Psi}_{3}\bar{\rho}+4\bar{\Psi}_{3}S^{2}_{34}-2\bar{\Psi}_{4}\bar{\kappa}-4\bar{\Psi}_{4}S^{2}_{14}+4\Phi_{01}\bar{\mu}+4\Phi_{02}\alpha+4\Phi_{02}S^{3}_{34}+4\Phi_{02}S^{1}_{14}$$
$$-2\Phi_{02}\Pi-4\Phi_{02}\bar{\beta}-4\Phi_{11}\bar{\Pi}+4\Phi_{12}S^{4}_{14}+4\Phi_{12}\epsilon+4\Phi_{12}S^{3}_{13}-4\Phi_{12}\rho+4\Phi_{12}S^{2}_{34}+4\Phi_{22}S^{2}_{13}+2\Phi_{22}\kappa$$
$$+4\phi_{01}\bar{\mu}+4\phi_{02}\alpha+4\phi_{02}S^{3}_{34}+4\phi_{02}S^{1}_{14}-2\phi_{02}\Pi-4\phi_{02}\bar{\beta}-4\phi_{11}\bar{\Pi}+4\phi_{12}S^{4}_{14}+4\phi_{12}\bar{\epsilon}+4\phi_{12}S^{3}_{13}-4\phi_{12}\rho$$
$$+4\phi_{12}S^{2}_{34}+4\phi_{22}S^{2}_{13}+2\phi_{22}\kappa-4\bar{\Lambda}_{0}\bar{\lambda}+2\bar{\Lambda}_{1}\bar{\Pi}+4\bar{\Lambda}_{1}S^{4}_{34}-4\bar{\Lambda}_{1}S^{1}_{13}-2\bar{\Lambda}_{2}S^{4}_{14}-2\bar{\Lambda}_{2}\bar{\epsilon}$$
$$-2\bar{\Lambda}_{2}S^{3}_{13}+2\bar{\Lambda}_{2}\bar{\rho}+2\bar{\Lambda}_{2}S^{2}_{34}+8\Lambda S^{4}_{34}+8\Lambda S^{1}_{13}-8\Sigma S^{4}_{34}i+8\Sigma iS^{1}_{13}=0,$$

$$2D\Psi_{2}+2D\bar{\Psi}_{2}-4D\Phi_{11}-4D\Lambda-2\delta\bar{\Psi}_{1}+2\delta\Phi_{10}-2\delta\phi_{10}-\delta\bar{\Lambda}_{0}-2\bar{\delta}\Psi_{1}+2\bar{\delta}\Phi_{01}+2\bar{\delta}\phi_{01}-\delta\Lambda_{0}+2\Psi_{0}\lambda+2\bar{\Psi}_{0}\bar{\lambda}$$
$$+4\Psi_{1}\alpha+4\Psi_{1}S^{3}_{34}+4\Psi_{1}S^{1}_{14}-4\Psi_{1}\Pi-4\bar{\Psi}_{1}\bar{\Pi}-4\bar{\Psi}_{1}S^{4}_{34}+4\bar{\Psi}_{1}\bar{\alpha}+4\bar{\Psi}_{1}S^{1}_{13}+4\Psi_{2}S^{4}_{14}+4\Psi_{2}S^{3}_{13}-6\Psi_{2}\rho+4\Psi_{2}S^{2}_{34}$$
$$+4\bar{\Psi}_{2}S^{4}_{14}+4\bar{\Psi}_{2}S^{3}_{13}-6\bar{\Psi}_{2}\bar{\rho}-4\bar{\Psi}_{2}S^{2}_{34}+4\Psi_{3}S^{2}_{13}+4\Psi_{3}\kappa+4\bar{\Psi}_{3}\bar{\kappa}+4\bar{\Psi}_{3}S^{2}_{14}-2\Phi_{00}\mu-2\Phi_{00}\bar{\mu}-4\Phi_{01}\alpha-4\Phi_{01}S^{3}_{34}$$
$$-4\Phi_{01}S^{1}_{14}+4\Phi_{01}\Pi+4\Phi_{10}\bar{\Pi}+4\Phi_{10}S^{4}_{34}-4\Phi_{10}\bar{\alpha}-4\Phi_{10}S^{1}_{13}+2\Phi_{02}\bar{\sigma}+2\Phi_{20}\sigma-8\Phi_{11}S^{4}_{14}-8\Phi_{11}S^{3}_{13}+4\Phi_{11}\rho$$
$$+4\Phi_{11}\bar{\rho}-4\Phi_{12}\bar{\kappa}-4\Phi_{12}S^{2}_{14}-4\Phi_{21}S^{2}_{13}-4\Phi_{21}\kappa+2\phi_{00}\mu-2\phi_{00}\bar{\mu}-4\phi_{01}\alpha-4\phi_{01}S^{3}_{34}-4\phi_{01}S^{1}_{14}-4\phi_{10}S^{4}_{34}+4\phi_{10}\bar{\alpha}$$
$$+4\phi_{10}S^{1}_{13}+2\phi_{02}\bar{\sigma}-2\phi_{20}\sigma+4\phi_{11}\rho-4\phi_{11}\bar{\rho}-8\phi_{11}S^{2}_{34}+4\phi_{12}S^{2}_{14}-4\phi_{21}S^{2}_{13}+2\Lambda_{0}\alpha+2\Lambda_{0}S^{3}_{34}+2\Lambda_{0}S^{1}_{14}$$
$$-2\bar{\Lambda}_{0}S^{4}_{34}+2\bar{\Lambda}_{0}\bar{\alpha}+2\bar{\Lambda}_{0}S^{1}_{13}-2\Lambda_{1}\rho-2\bar{\Lambda}_{1}\bar{\rho}-2\Lambda_{2}S^{2}_{13}-2\bar{\Lambda}_{2}S^{2}_{14}-8\Lambda S^{4}_{14}-8\Lambda S^{3}_{13}-8\Sigma iS^{2}_{34}=0,$$

$$-2\Delta\Psi_{2}+2\Delta\bar{\Psi}_{2}-4\Delta\phi_{11}+4\Delta\Sigma i+2\delta\Psi_{3}+2\delta\Phi_{21}+2\delta\phi_{21}-\delta\Lambda_{2}-2\bar{\delta}\bar{\Psi}_{3}-2\bar{\delta}\Phi_{12}+2\bar{\delta}\phi_{12}+\bar{\delta}\bar{\Lambda}_{2}$$
$$+4\Psi_{1}\nu-4\Psi_{1}S^{1}_{24}-4\bar{\Psi}_{1}\bar{\nu}+4\bar{\Psi}_{1}S^{1}_{23}-4\Psi_{2}S^{4}_{24}+4\Psi_{2}S^{1}_{34}-4\Psi_{2}S^{3}_{23}-6\Psi_{2}\mu+4\bar{\Psi}_{2}S^{4}_{24}+4\bar{\Psi}_{2}S^{1}_{34}-4\Psi_{2}S^{3}_{23}$$
$$-6\Psi_{2}\mu+4\bar{\Psi}_{2}S^{4}_{24}+4\Psi_{2}S^{1}_{34}+4\bar{\Psi}_{2}S^{3}_{23}+6\Psi_{2}\bar{\mu}+4\Psi_{3}\beta-4\Psi_{3}S^{2}_{23}+4\Psi_{3}S^{4}_{34}-4\Psi_{3}\tau+4\bar{\Psi}_{3}S^{3}_{34}+4\bar{\Psi}_{3}S^{2}_{24}$$
$$-4\bar{\Psi}_{3}\bar{\beta}+4\bar{\Psi}_{3}\bar{\tau}+2\Psi_{4}\sigma-2\bar{\Psi}_{4}\bar{\sigma}-4\Phi_{01}S^{1}_{24}+4\Phi_{10}S^{1}_{23}+2\Phi_{02}\lambda-2\Phi_{20}\bar{\lambda}+8\Phi_{11}S^{1}_{34}-4\Phi_{11}\mu+4\Phi_{11}\bar{\mu}$$
$$+4\Phi_{12}S^{3}_{34}+4\Phi_{12}S^{2}_{24}-4\Phi_{12}\bar{\beta}+4\Phi_{21}\beta-4\Phi_{21}S^{2}_{23}+4\Phi_{21}S^{4}_{34}-2\Phi_{22}\rho+2\Phi_{22}\bar{\rho}+4\phi_{01}\nu-4\phi_{01}S^{1}_{24}+4\phi_{10}\bar{\nu}$$

$$-4\phi_{10}S^{1}_{23} - 2\phi_{02}\lambda - 2\phi_{20}\bar{\lambda} - 8\phi_{11}S^{4}_{24} - 8\phi_{11}S^{3}_{23} - 4\phi_{11}\mu - 4\phi_{11}\bar{\mu} - 4\phi_{12}S^{3}_{34} - 4\phi_{12}S^{2}_{24} + 4\phi_{12}\bar{\beta} - 4\phi_{12}\bar{\tau} + 4\phi_{21}\beta$$
$$-4\phi_{21}S^{2}_{23} + 4\phi_{21}S^{4}_{34} - 4\phi_{21}\tau + 2\phi_{22}\rho + 2\phi_{22}\bar{\rho} - 2\Lambda_{0}S^{1}_{24} + 2\bar{\Lambda}_{0}S^{1}_{23} + 2\Lambda_{1}\mu - 2\bar{\Lambda}_{1}\bar{\mu} - 2\Lambda_{2}\beta + 2\Lambda_{2}S^{2}_{23}$$
$$-2\Lambda_{2}S^{4}_{34} - 2\bar{\Lambda}_{2}S^{3}_{34} - 2\bar{\Lambda}_{2}S^{2}_{24} + 2\bar{\Lambda}_{2}\bar{\beta} - 8\Lambda S^{1}_{34} + 8\Sigma S^{4}_{24}i + 8\Sigma i S^{3}_{23} = 0,$$

$$-2\Delta\Psi_{1} + 2\Delta\Phi_{01} - 2\Delta\phi_{01} + \Delta\Lambda_{0} + 2\delta\Psi_{2} - 2\delta\Lambda_{1} + 4\delta\Lambda + 4\delta\Sigma i - 2\bar{\delta}\Phi_{02} + 2\bar{\delta}\phi_{02} + 2\Psi_{0}\nu$$
$$-4\Psi_{0}S^{1}_{24} - 4\Psi_{1}S^{4}_{24} + 4\Psi_{1}S^{1}_{34} - 4\Psi_{1}S^{3}_{23} - 4\Psi_{1}\mu + 4\Psi_{1}\gamma - 4\Psi_{2}S^{2}_{23} + 4\Psi_{2}S^{4}_{34} - 6\Psi_{2}\tau + 4\Psi_{3}\sigma - 2\Phi_{00}\bar{\nu}$$
$$+4\Phi_{00}S^{1}_{23} + 4\Phi_{01}S^{4}_{24} + 4\Phi_{01}S^{1}_{34} + 4\Phi_{01}S^{3}_{23} + 4\Phi_{01}\bar{\mu} - 4\Phi_{01}\gamma + 4\Phi_{02}\alpha + 4\Phi_{02}S^{3}_{34} + 4\Phi_{02}S^{2}_{24}$$
$$-4\Phi_{02}\bar{\beta} + 2\Phi_{02}\bar{\tau} + 4\Phi_{11}\tau - 4\Phi_{12}\rho + 2\phi_{00}\bar{\nu} - 4\phi_{00}S^{1}_{23} - 4\phi_{01}S^{4}_{24} - 4\phi_{01}S^{1}_{34} - 4\phi_{01}S^{3}_{23} - 4\phi_{01}\bar{\mu} + 4\phi_{01}\gamma$$
$$-4\phi_{02}\alpha - 4\phi_{02}S^{3}_{34} - 4\phi_{02}S^{2}_{24} + 4\phi_{02}\bar{\beta} - 2\phi_{02}\bar{\tau} - 4\phi_{11}\tau + 4\phi_{12}\rho + 2\Lambda_{0}S^{4}_{24} - 2\Lambda_{0}S^{1}_{34} + 2\Lambda_{0}S^{3}_{23} + 2\Lambda_{0}\mu - 2\Lambda_{0}\gamma$$
$$+4\Lambda_{1}S^{2}_{23} - 4\Lambda_{1}S^{4}_{34} + 2\Lambda_{1}\tau - 2\Lambda_{2}\sigma - 8\Lambda S^{2}_{23} + 8\Lambda S^{4}_{34} - 8\Sigma S^{2}_{23}i + 8\Sigma S^{4}_{34}i = 0,$$

$$-2\Delta\bar{\Psi}_{3} + 2\Delta\Phi_{12} + 2\Delta\phi_{12} - \Delta\bar{\Lambda}_{2} - 2\delta\Phi_{22} - 2\delta\phi_{22} + 2\bar{\delta}\bar{\Psi}_{4} + 6\bar{\Psi}_{2}\bar{\nu} - 4\bar{\Psi}_{2}S^{1}_{23} - 4\bar{\Psi}_{3}S^{4}_{24} - 4\bar{\Psi}_{3}S^{1}_{34}$$
$$-4\bar{\Psi}_{3}S^{3}_{23} - 8\bar{\Psi}_{3}\bar{\mu} - 4\bar{\Psi}_{3}\bar{\gamma} - 4\Psi_{4}S^{3}_{34} - 4\bar{\Psi}_{4}S^{2}_{24} + 8\bar{\Psi}_{4}\bar{\beta} - 2\bar{\Psi}_{4}\bar{\tau} - 2\Phi_{02}\nu + 4\Phi_{02}S^{1}_{24} - 4\Phi_{11}\bar{\nu} + 4\Phi_{12}S^{4}_{24}$$
$$-4\Phi_{12}S^{1}_{34} + 4\Phi_{12}S^{3}_{23} + 4\Phi_{12}\mu + 4\Phi_{12}\bar{\gamma} + 4\Phi_{21}\bar{\lambda} - 4\Phi_{22}\beta + 4\Phi_{22}S^{3}_{23} - 4\Phi_{22}S^{4}_{34} - 4\Phi_{22}\alpha + 2\Phi_{22}\tau - 2\phi_{02}\nu$$
$$+4\phi_{02}S^{1}_{24} - 4\phi_{11}\bar{\nu} + 4\phi_{12}S^{4}_{24} - 4\phi_{12}S^{1}_{34} + 4\phi_{12}S^{3}_{23} + 4\phi_{12}\mu + 4\phi_{12}\bar{\gamma} + 4\phi_{21}\bar{\lambda} - 4\phi_{22}\beta + 4\phi_{22}S^{2}_{23} - 4\phi_{22}S^{4}_{34}$$
$$-4\phi_{22}\bar{\alpha} + 2\phi_{22}\tau + 2\bar{\Lambda}_{1}\bar{\nu} - 4\bar{\Lambda}_{1}S^{1}_{23} - 2\bar{\Lambda}_{2}S^{4}_{24} - 2\bar{\Lambda}_{2}S^{1}_{34} - 2\bar{\Lambda}_{2}S^{3}_{23} - 2\bar{\Lambda}_{2}\bar{\gamma} - 8\Lambda S^{1}_{23} + 8\Sigma i S^{1}_{23} = 0,$$

$$2\Delta\Psi_{2} + 2\Delta\bar{\Psi}_{2} - 4\Delta\Phi_{11} - 4\Delta\Lambda - 2\delta\Psi_{3} + 2\delta\Phi_{21} + 2\delta\phi_{21} + \delta\Lambda_{2} - 2\bar{\delta}\bar{\Psi}_{3} + 2\bar{\delta}\Phi_{12} - 2\bar{\delta}\phi_{12} + \bar{\delta}\bar{\Lambda}_{2} - 4\Psi_{1}\nu$$
$$+4\Psi_{1}S^{1}_{24} - 4\bar{\Psi}_{1}\bar{\nu} + 4\bar{\Psi}_{1}S^{1}_{23} + 4\Psi_{2}S^{4}_{24} - 4\Psi_{2}S^{1}_{34} + 4\Psi_{2}S^{3}_{23} + 6\Psi_{2}\mu + 4\bar{\Psi}_{2}S^{4}_{24} + 4\bar{\Psi}_{2}S^{1}_{34} + 4\bar{\Psi}_{2}S^{3}_{23} + 6\bar{\Psi}_{2}\bar{\mu} - 4\Psi_{3}\beta$$
$$+4\Psi_{3}S^{2}_{23} - 4\Psi_{3}S^{4}_{34} + 4\Psi_{3}\tau + 4\bar{\Psi}_{3}S^{3}_{34} + 4\bar{\Psi}_{3}S^{2}_{24} - 4\bar{\Psi}_{3}\bar{\beta} + 4\bar{\Psi}_{3}\bar{\tau} - 2\Psi_{4}\sigma - 2\bar{\Psi}_{4}\bar{\sigma} + 4\Phi_{01}\nu - 4\Phi_{01}S^{1}_{24} + 4\Phi_{10}\bar{\nu}$$
$$-4\Phi_{10}S^{1}_{23} - 2\Phi_{02}\lambda - 2\Phi_{20}\bar{\lambda} - 8\Phi_{11}S^{4}_{24} - 8\Phi_{11}S^{3}_{23} - 4\Phi_{11}\mu - 4\Phi_{11}\bar{\mu} - 4\Phi_{12}S^{3}_{34} - 4\Phi_{12}S^{2}_{24} + 4\Phi_{12}\bar{\beta} - 4\Phi_{12}\bar{\tau}$$
$$+4\Phi_{21}\beta - 4\Phi_{21}S^{2}_{23} + 4\Phi_{21}S^{4}_{34} - 4\Phi_{21}\tau + 2\Phi_{22}\rho + 2\Phi_{22}\bar{\rho} - 4\phi_{01}S^{1}_{24} + 4\phi_{10}S^{1}_{23} + 2\phi_{02}\lambda - 2\phi_{20}\bar{\lambda} + 8\phi_{11}S^{1}_{34} - 4\phi_{11}\mu$$
$$+4\phi_{11}\bar{\mu} + 4\phi_{12}S^{3}_{34} + 4\phi_{12}S^{2}_{24} - 4\phi_{12}\bar{\beta} + 4\phi_{21}\beta - 4\phi_{21}S^{2}_{23} + 4\phi_{21}S^{4}_{34} - 2\phi_{22}\rho + 2\phi_{22}\bar{\rho} + 2\Lambda_{0}S^{1}_{24} + 2\bar{\Lambda}_{0}S^{1}_{23} - 2\Lambda_{1}\mu$$
$$-2\bar{\Lambda}_{1}\bar{\mu} + 2\Lambda_{2}\beta - 2\Lambda_{2}S^{2}_{23} + 2\Lambda_{2}S^{4}_{34} - 2\bar{\Lambda}_{2}S^{3}_{34} - 2\bar{\Lambda}_{2}S^{2}_{24} - 2\bar{\Lambda}_{2}\bar{\beta} - 8\Lambda S^{4}_{24} - 8\Lambda S^{3}_{23} + 8\Sigma S^{1}_{34}i = 0.$$

[1] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[2] F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, Rev. Mod. Phys. **48**, 393 (1976).
[3] A. Trautman, Ann. N.Y. Acad. Sci. **262**, 241 (1975).
[4] S. Hojman, M. Rosembaum, and M. P. Ryan, Phys. Rev. D **17**, 3141 (1978).
[5] A. Papapetrou, Ann. Inst. Henri Poincaré **13**, 271 (1970).
[6] J. A. Schouten, *Ricci-Calculus*, 2nd ed. (Springer, Berlin, 1954).

[7] After submitting this paper for publication we received a preprint by S. Jogia and J. B. Griffiths (University of Technology, Loughborough) with a discussion of the same problem.

Some properties of static general relativistic stellar models [a)]

Lee Lindblom
Physics Department, University of California, Santa Barbara, California 93106

(Received 30 November 1979; accepted for publication 29 February 1980)

A number of properties of static general relativistic stellar models are presented which appear to be relevant to the ongoing search for a proof that all such models must have spherical symmetry. It is shown that any such model, having conformally flat spatial sections, must have spherical symmetry. A general procedure is described which allows one to construct the type of "divergence equals positive quantity" identities for static stellar models, which were used to prove that static black holes must have spherical symmetry. This procedure is used to produce a large new class of identities for the exterior vacuum regions of static stellar models and identities are constructed for the interior regions of uniform density models. These identities are used to prove that static uniform density stellar models must have spherical symmetry.

I. REVIEW OF STATIC STELLAR MODELS

In this paper we study the solutions of Einstein's field equations with matter corresponding to a perfect fluid which is in static equilibrium. It has been shown previously[1-6] that such a stellar model is a spacetime with a metric which can be represented by the line element

$$ds^2 = -V^2 dt^2 + g_{ab} dx^a dx^b . \tag{1}$$

The components of g_{ab} and the function V are independent of the coordinate t, and the tensor g_{ab} represents the positive definite metric on each t = constant submanifold, each of which has the same topology as R^3. Einstein's equations for this system can be written in the form

$$\nabla_a \nabla^a V = 4\pi V(\rho + 3p) , \tag{2}$$

$$R_{ab} = V^{-1} \nabla_a \nabla_b V + 4\pi(\rho - p) g_{ab} . \tag{3}$$

The tensor R_{ab} represents the three-dimensional Ricci tensor of the metric g_{ab}, and ∇_a represents the three-dimensional covariant derivative compatible with g_{ab}. The functions ρ and p represent the mass density and pressure of the fluid, respectively. These are related by an equation of state; i.e., a given relationship of the form $\rho = \rho(p)$. The contracted Bianchi identities for the three-dimensional curvature, and Eq. (3) imply the equivalent of Euler's equation:

$$\nabla_a p = -V^{-1}(\rho + p) \nabla_a V . \tag{4}$$

In the discussion that follows, it will be helpful to define a number of additional quantities:

$$W = \nabla^a V \nabla_a V , \tag{5}$$

$$n^a = W^{-1/2} \nabla^a V , \tag{6}$$

$$\beta^{ab} = g^{ab} - n^a n^b , \tag{7}$$

$$H_{ab} = \beta_a{}^c \beta_b{}^d \nabla_c n_d , \tag{8}$$

$$\psi_{ab} = H_{ab} - \tfrac{1}{2} \beta_{ab} H . \tag{9}$$

These quantities describe the geometry of the V = const. 2-surfaces. The unit vector field n^a is orthogonal to these surfaces; β_{ab} is the intrinsic metric; H_{ab} is the extrinsic curvature tensor and ψ_{ab} represents the trace-free part of H_{ab}.

The quantities which define the geometry of the V = const. 2-surfaces are related to one another by splitting Einstein's equations into pieces tangent and orthogonal to these surfaces in the standard way. In the discussion that follows, we make use of two of the equations obtained by this splitting of Eqs. (2) and (3):

$$W^{-1} \nabla^a V \nabla_a W = -2W^{1/2} H + 8\pi V(\rho + 3p) , \tag{10}$$

$$\begin{aligned} W^{-1} \nabla^a V \nabla_a H &= -\tfrac{1}{2} W^{-1/2} H^2 + V^{-1} H - \beta^{ab} \nabla_a (\beta_b{}^c \nabla_c W^{-1/2}) \\ &\quad - W^{-1/2} \psi_{ab} \psi^{ab} - 8\pi W^{-1/2} (\rho + p) . \end{aligned} \tag{11}$$

The derivation of these equations can be found in the literature.[1-3,5]

It has long been suspected that no nonspherical, asymptotically flat solutions exist to Eqs. (2) and (3). This belief is motivated by an analogous theorem for Newtonian stellar models[7] and a similar result for the vacuum ($\rho = p = 0$) black hole solutions of Eqs. (2) and (3).[8-10] It has also been shown that stationary (nonstatic) general relativistic stellar models (made of dissipative fluids) must be axisymmetric.[11] Little progress has been made on the problem of static relativistic stellar models however. It has been shown that if

$$\beta^{ab} \nabla_b W = 0 \tag{12}$$

then the model must be spherical.[3] It has also been shown that no "almost" spherical static stellar models exist.[3,4]

For the remainder of this paper, we discuss some of the properties of these stellar models which appear to be relevant to the ongoing search for a proof that spherical symmetry is necessary. In Sec. II it is shown that if the 3-geometry described by g_{ab} is conformally flat, then the stellar model must be spherical. In Sec. III we describe a procedure which allows one to construct, for stellar models, the type of identities which were used[8-10] to prove that static black holes must have spherical symmetry. In Sec. IV we use the procedure described in Sec. III to construct identities applicable in the vacuum exterior regions of any static stellar model, and in the interior regions of models with uniform density. And finally in Sec. V, we use these identities to show that, in the special case of uniform density models, spherical symmetry is necessary.

[a)]Supported by the National Science Foundation.

II. CONFORMAL FLATNESS AND SPHERICAL SYMMETRY

The conformal properties of a three-dimensional manifold are not described by the Weyl tensor (which vanishes identically) but by a certain third-rank tensor field,[12] defined by

$$R_{abc} = \nabla_c R_{ab} - \nabla_b R_{ac} + \tfrac{1}{4}(g_{ac}\nabla_b R - g_{ab}\nabla_c R\,). \quad (13)$$

This expression can be evaluated in terms of the functions V, ρ, etc. by using Eq. (3) for R_{ab}. Making this substitution, and using the quantities defined in Egs. (5)–(9), it is straightforward to verfy the following;

$$R_{abc}R^{\,abc} = 8V^{-4}W^2\{\psi_{ab}\psi^{ab} + \tfrac{1}{8}W^{-2}\beta^{ab}\nabla_a W\nabla_b W\}\,. \quad (14)$$

It is interesting to note that the matter variables do not appear explicitly in this expression; this is precisely the same expression which was found to hold in a vacuum spacetime.[9] Expression (14), however, is true for any geometry satisfying Eqs. (2) and (3).

If the geometry g_{ab} were conformally flat, then the tensor R_{abc} must vanish.[12] From Eq. (14) it would follow that ψ_{ab} and $\beta^{ab}\nabla_b W$ must also vanish in this case. Consequently, it would follow from the standard arguments[3] that the stellar model would necessarily be spherical. Thus we have established the following lemma.

Lemma: If the spatial geometry g_{ab} of a static general relativistic stellar model [i.e., *a solution of Eqs.* (2) *and* (3)] *is conformally flat, then the stellar model necessarily has spherical symmetry.*

Another expression for the square of the conformal tensor, which will be useful in the following section, is the following:

$$\begin{aligned}\tfrac{1}{4}V^4W^{-1}R_{abc}R^{\,abc} &= \nabla_a\nabla^a W - V^{-1}\nabla^a V\nabla_a W - \tfrac{3}{4}W^{-1}\nabla^a W\nabla_a W\\ &\quad + 8\pi W(\rho+p) + 4\pi V W^{-1}(\rho+3p)\nabla^a V\nabla_a W\\ &\quad - 16\pi^2V^2(\rho+3p)^2 - 8\pi V\nabla^a V\nabla_a\rho\,. \end{aligned} \quad (15)$$

This expression is derived using essentially the same procedure as that described to derive Eq. (14); and we note that this expression agrees in the vacuum limit with an analogous expression derived previously.[9]

III. CONSTRUCTION OF DIVERGENCE IDENTITIES

The proof that static black holes must have spherical symmetry depends on constructing an identity which has the form of a divergence equaling a positive definite quantity (which vanishes if and only if the spacetime is spherical). One positive definite quantity, which might be suitable for such an identity, has been identified in the last section: $R_{abc}R^{\,abc}$. The existence of another suitable quantity is implied by Eq. (12). If the spacetime were spherical, then the vector $\nabla_a W$ must be proportional to $\nabla_a V$; we call the proportionality factor F. The function $F = F(V,W)$ can be determined explicitly (as shall be discussed in detail in Sec. IV) once the equation of state of the fluid in the stellar model is specified. Taking F as a known function, it follows that the quantity:

$$[\nabla_a W - F\nabla_a V\,][\nabla^a W - F\nabla^a V\,] \quad (16)$$

is positive and vanishes if and only if the model is spherical.

Having identified two suitable positive definite quantities, we are lead to ask when an identity of the following form can be found:

$$\begin{aligned}\nabla_a\{K_1(V,W)\nabla^a V + K_2(V,W)\nabla^a W\}\\ = \tfrac{1}{4}Q_1(V,W)V^4W^{-1}R_{abc}R^{\,abc}\\ + Q_2(V,W)W^{-1}|\nabla_a W - F\nabla_a V\,|^2, \end{aligned} \quad (17)$$

where K_1, K_2, Q_1 and Q_2 are functions of V and W with $Q_1 > 0$ and $Q_2 > 0$. (We note that every identity used to prove the spherical symmetry of black hole spacetimes has been of this form.[8-10,13]) On examining each side of Eq. (17) (using Eq. (15) to evaluate the first term on the right) we find, in addition to functions of V and W, terms linear in the three functions: $\nabla^a\nabla_a W$, $\nabla^a V\nabla_a W$, and $\nabla^a W\nabla_a W$. If we require that the coefficients of these functions on one side of the equation equal the corresponding coefficients on the other side, the following four constraints on the functions K_1, K_2, Q_1, and Q_2 are implied:

$$K_2 = Q_1\,, \quad (18)$$

$$Q_2 = W\frac{\partial Q_1}{\partial W} + \tfrac{3}{4}Q_1\,, \quad (19)$$

$$\begin{aligned}\frac{\partial Q_1}{\partial V} &= -2F\frac{\partial Q_1}{\partial W} - \frac{\partial K_1}{\partial W}\\ &\quad - \left[V^{-1} - 4\pi V W^{-1}(\rho+3p) + \frac{3}{2}W^{-1}F\right]Q_1, \end{aligned} \quad (20)$$

$$\begin{aligned}\frac{\partial K_1}{\partial V} &= F^2\frac{\partial Q_1}{\partial W} - 4\pi K_1 V W^{-1}(\rho+3p)\\ &\quad - \Big[16\pi^2V^2W^{-1}(\rho+3p)^2\\ &\quad - \tfrac{3}{4}W^{-1}F^2 + 8\pi V\frac{d}{dV}(\rho+p)\Big]Q_1. \end{aligned} \quad (21)$$

Thus, if these partial differential equations can be solved for K_1, K_2, Q_1, and Q_2 (with $Q_1 > 0$ and $Q_2 > 0$), then an identity in the form of Eq. (17) will exist. Note that V and W play the role of independent variables in these equations. Each of the coefficients in these equations is a known function of V and W: F was assumed to be a known function, while ρ and p are explicit functions of V determined by integrating Eq. (4).

A large number of solutions clearly exist to Eqs. (18)–(21). Equations (20) and (21) form a linear system of equations for Q_1 and K_1. One can imagine solving these equations as a Cauchy initial value problem. On an initial surface, say $V = V_i$, we arbitrarily specify the functions $Q_1(V_i,W)$ and $K_1(V_i,W)$. Equations (20) and (21) allow us to compute the normal derivatives of these functions; consequently the equations can be integrated to find Q_1 and K_1 (at least for V sufficiently close to V_i). Thus a large number of solutions to these equations exist, each of which corresponds to an identity of the form in Eq. (17). In order to be useful as tools for proving the spherical symmetry of stars, we must limit the choice of functions to those for which $Q_1 > 0$ and $Q_2 > 0$. At present, it is not known whether or not there exist solutions with positive Q_1 and Q_2 in general. We see in the next section that in some special cases, however, positive solutions do exist.

IV. IDENTITIES FOR VACUUM AND UNIFORM DENSITY SPACES

In this section we use the procedure described above to derive identities for the special cases of vacuum spacetimes (the exterior regions of stars) and for the interior regions of uniform density stars. To do this we must explicitly integrate Eqs. (20) and (21) for these cases. Before these equations can be explicitly integrated, we need to discuss how the function F can be determined for a given equation of state. For the spherical solutions of Eqs. (2) and (3) the functions W and H (the trace of the 2-dimensional extrinsic curvature) depend only on V; let $W_0(V)$ and $H_0(V)$ denote those functions. Equations (10) and (11) imply that W_0 and H_0 satisfy the differential equations

$$\frac{dW_0}{dV} = -2W_0^{1/2}H_0 + 8\pi V(\rho+3p), \tag{22}$$

$$\frac{dH_0}{dV} = -\tfrac{1}{2}W_0^{-1/2}H_0^2 + V^{-1}H_0 - 8\pi W_0^{-1/2}(\rho+p). \tag{23}$$

The function H_0 can be eliminated from these equations, to obtain a single equation for W_0:

$$\frac{d^2W_0}{dV^2} = V^{-1}\frac{dW_0}{dV} + \tfrac{3}{4}W_0^{-1}\left(\frac{dW_0}{dV}\right)^2 - 8\pi VW_0^{-1}\frac{dW_0}{dV}(\rho+3p) + 16\pi^2W_0^{-1}V^2(\rho+3p)^2 + 8\pi V\frac{d}{dV}(\rho+p). \tag{24}$$

Given an equation of state, $\rho=\rho(p)$, Eq. (4) can be integrated to determine the functions $\rho(V)$ and $p(V)$. Using these functions, Eq. (24) can be integrated to determine $W_0(V)$: the function to which W would be equal if the solution were spherical. Given this function, W_0, it is easy now to find the function F. In fact, F can be chosen in an infinite number of ways. One obvious choice is $F=dW_0/dV$, but $F=dW_0/dV+(W_0-W)^n$ or $F=W^nW_0^{-n}\,dW_0/dV$ would do just as well. Thus for each equation of state which we specify there exist an infinite number of different choices of the function F; and for each F there exist an infinite number of identities in the form of Eq. (17).

Let us now explicitly utilize the procedure, which is outlined above, to obtain identities that are relevant to the study of stellar models. We begin with the simplest case: identities which describe the vacuum exterior regions of any stellar model. The first step is to solve Eq. (24) for W_0:

$$W_0 = \frac{1}{16M^2}(1-V^2)^4, \tag{25}$$

and thus

$$\frac{dW_0}{dV} = -\frac{V}{2M^2}(1-V^2)^3. \tag{26}$$

In these expressions, the constant M represents the asymptotically defined mass of the star. These solutions now allow us to choose the function F in any number of different ways. We select two different choices, each of which allows us to explicitly solve Eqs. (20) and (21) in a straightforward manner.

For our first choice we let

$$F = \frac{W}{W_0}\frac{dW_0}{dV} = -8VW(1-V^2)^{-1}. \tag{27}$$

This expression is substituted in Eqs. (20) and (21), and the equations are integrated. The general solution of these equations is given by:

$$Q_1 = V^{-1}(1-V^2)^{-2}\{A(x) - 8M^2(1-V^2)^{-1}B'(x)\}, \tag{28}$$

$$K_1 = B(x) + 8VW(1-V^2)^{-1}Q_1, \tag{29}$$

where $x=W/W_0$; A and B are arbitrary functions, and $B'(x)=dB/dx$. It is clearly possible to choose A and B in such a way that $Q_1>0$ and $Q_2>0$; so that these identities are potentially useful in our search for a proof of spherical symmetry. This class of identities contains, as special cases, every identity that has been constructed to prove the spherical symmetry of black holes. For example, Robinson's identities[10] are given by

$$A(x) = -c, \tag{30}$$

$$B(x) = -(c+d)x/8M^2, \tag{31}$$

where c and d are arbitrary constants. The two identities originally discovered by Israel[8] (or in this notation by Müller zum Hagen, *et al.*[9]) are obtained by setting

$$A(x) = 0, \tag{32}$$

$$B(x) = -4M^{-1/2}x^{1/4}, \tag{33}$$

for one identity, and

$$A(x) = -2M^{3/2}x^{-3/4}, \tag{34}$$

$$B(x) = -M^{-1/2}x^{1/4}, \tag{35}$$

for the other. We see that the technique described here yields a considerable degree of generalization over previously known identities.

Another choice of the function F, for the vacuum case, is

$$F = \left(\frac{W}{W_0}\right)^{3/4}\frac{dW_0}{dV} = -4M^{-1/2}VW^{3/4}. \tag{36}$$

This choice of F also allows Eqs. (20) and (21) to be integrated in general in a straightforward manner. The general solutions for Q_1 and K_1 are:

$$Q_1 = V^{-1}W^{3/4}C(y) - \tfrac{1}{20}M^{3/2}V^{-1}W^{-1/2}D'(y), \tag{37}$$

$$K_1 = D(y) + 4M^{-1/2}VW^{3/4}Q_1, \tag{38}$$

where $y = W^{1/4} - W_0^{1/4}$; $C(y)$ and D(y) are arbitrary functions of y. Clearly it is possible to select the functions C and D so that $Q_1>0$ and $Q_2>0$. Consequently this represents another large class of divergence identities which may be useful in the study of the spherical symmetry of static stars.

Let us move on now to a consideration of the non-vacuum interior regions of static stellar models. The first problem one encounters is solving Eq. (24) for an arbitrary equation of state. I have not determined how this can be accomplished in general, yet. For the special case of uniform density stars, however, the solution can be found. We begin by integrating Eq. (4) for this case, to find that

$$p = \rho V^{-1}(V_s - V), \tag{39}$$

where ρ is the constant density of the star, and V_s is the value of the potential V at the suface of the star. Using this expression for p, it is easy to verify that

$$\frac{dW_0}{dV} = \tfrac{8}{3}\pi V(\rho + 3p) = \tfrac{8}{3}\pi\rho(3V_s - 2V) \quad (40)$$

is the first integral of Eq. (24). We also find that with the choice

$$F = \frac{dW_0}{dV} = \tfrac{8}{3}\pi V(\rho + 3p), \quad (41)$$

Eqs. (20) and (21) can be integrated in a straightforward fashion, with the result:

$$Q_1 = V^{-1}E(W - W_0), \quad (42)$$

$$K_1 = -\tfrac{8}{3}\pi(\rho + 3p)E(W - W_0), \quad (43)$$

where E is an arbitrary function of $W - W_0$. A more general integral of these equations exists, which involves an additional arbitrary function of $W - W_0$. While it is straightforward to obtain the more general solution, it is rather lengthy and complicated and it will not be needed in the proof that uniform density stars mush have spherical symmetry.

V. UNIFORM DENSITY STARS MUST BE SHPERICAL

In this final section we will show how the particular identities derived in the last section can be used to prove that spherical symmetry is necessary in the special case of uniform density stellar models. This discussion is a somewhat more detailed version of the proof given in Ref. 14. Before proceeding directly to the proof it is necessary to discuss in more detail the smoothness assumptions and boundary conditions for the solutions of Eqs. (2) and (3) which are appropriate for stellar models. We assume that V and g_{ab} are C^3 except at the boundary ($V = V_s$) between the interior and exterior of the star. This assumption guarantees that suitable coordinates exist so that V and g_{ab} are analytical functions.[15,16] At the suface of the star, the differentiability is reduced however. The exact differentiability can be inferred by requiring that Eqs. (2) and (3) [and consequently Eqs. (10) and (11)] are satisfied even at the surface of the star. Equation (11) implies that the extrinsic curvature H must be continuous at the surface[17]; while Eq. (10) shows that $\nabla_a W$ will have a discontinuity in the direction of the normal to the surface if the density function has a discontinuity there. The magnitude of this discontinuity is given by

$$\lim_{V\to V_s^+} W^{-1}\nabla^a V\nabla_a W - \lim_{V\to V_s^-} W^{-1}\nabla^a V\nabla_a W = -8\pi V_s \lim_{p\to 0+}\rho. \quad (44)$$

In the case of uniform density stars the density function must have a discontinuity at the surface, while other equations of state may not have this discontinuity. To make use of the formalism derived above, we must also take care to properly match the function W_0 across the surface of the star. To this end we will choose the mass constant of Eq. (25) and the constant obtained from integrating Eq. (40) so that W_0 is continuous while its first derivative satisfies the discontinuity equation:

$$\lim_{V\to V_s^+} W^{-1}\nabla^a V\nabla_a W_0 - \lim_{V\to V_s^-} W^{-1}\nabla^a V\nabla_a W_0 = -8\pi V_s \lim_{p\to 0+}\rho. \quad (45)$$

This leads to the following function W_0:

$$W_0 = \tfrac{2}{3}\pi\rho(1 - V^2)^4(1 - V_s^2)^{-3} \quad \text{for } V > V_s, \quad (46)$$

$$W_0 = \tfrac{8}{3}\pi\rho V(3V_s - V) + \tfrac{2}{3}\pi\rho(1 - 9V_s^2) \quad \text{for } V < V_s. \quad (47)$$

Note that by choosing W_0 in this way, the gradient $\nabla_a(W - W_0)$ is continuous even at the surface of the star.

We are now prepared to proceed with the proof of the following theorem:

Theorem: A static asymptotically flat general relativistic stellar model, which is made of uniform (positive) density fluid, is necessarily spherically symmetric.

The goal of the first step in the proof is to use the identities derived in Sec. IV to establish that the function $W - W_0$ must attain its maximum value on the surface of the star. Integrate Eq. (17) over the exterior region of the star using Eqs. (18), (19), (28), and (29) with $B = 0$. The divergence on the left-hand side is converted to a boundary integral at the surface of the star and at infinity. The surface integral at infinity vanishes if A is bounded. Therefore the following relationship is true:

$$-\int_{V=V_s^+} ({}^2g)^{1/2}V^{-1}W^{-1/2}(1 - V^2)^{-2}A(W/W_0)\nabla^a V \times[\nabla_a W - F\nabla_a V]d^2x = I, \quad (48)$$

where

$$I = \int ({}^3g)^{1/2}\{\tfrac{1}{4}Q_1 V^4 W^{-1}R_{abc}R^{abc} + Q_2 W^{-1}|\nabla_a W - F\nabla_a V|^2\}d^3x. \quad (49)$$

Let us choose the function $A(U)$ so that it vanishes for $U < U_0$ and smoothly increases to positive values for $U > U_0$. In this case Q_1 and Q_2 are nonnegative functions. If the maxiumum value that W/W_0 assumed in the exterior of the star were larger than the maximum value which it assumed on the surface of the star, one could choose the constant U_0 to lie somewhere between these values. In this case the boundary integral on the left of Eq. (48) would vanish. Since the volume integral on the right would vanish in this situation only if the star were spherical (see Sec. II), we conclude that W/W_0 attains its maximum value (relative to the exterior region) on the surface of the star, or that the star is spherical. Thus W/W_0 attains its maximum value (relative to the exterior region) on the surface of the star, since this also occurs in the spherical case. Since W_0 also attains its maximum on the surface of the star, it follows that $W - W_0$ also attains its maximum value (relative to the exterior region) on the surface of the star. Next integrate Eq. (17) using Eqs. (18), (19), (42), and (43) over the interior of the star. By appropriately choosing the function E, in an argument analogous to that described above for the exterior, it is straightforward to show that $W - W_0$ attains its maximum value (relative to the interior region) on the surface of the star. Thus the absolute maximum value of $W - W_0$ occurs somewhere on the surface of the star. Also, since the gradient $\nabla_a(W - W_0)$ was shown to be continuous, it must vanish at this maximum point.

The next step in the proof is to show that the function $W - W_0$ is in fact constant in the interior of the star. This is accomlished using a maximum principle for elliptic differential equations.[18] In the interior of the star, Eqs. (17), (18), (19), (42), and (43) with $E = 1$ imply that

$$\nabla^a[V^{-1}\nabla_a(W - W_0)] \geqslant 0. \tag{50}$$

The maximum principle for this type of differential equation states (roughly, see Ref. 18 for a precise statement) that if $W - W_0$ satisfies Eq. (50) and has a maximum at a boundary point and if the outward normal derivative of $W - W_0$ is not positive at this maximum point, then $W - W_0$ must be constant. Since the gradient of $W - W_0$ vanishes at the maximum point, $W - W_0$ must be constant. From this it follows (again using Eqs. (17)–(19), (42), and (43) with $E = 1$) that $R_{abc} = 0$ in the interior of the star.

The final step is to show that W/W_0 is constant in the exterior of the star. Chose $A = 1$ and $B = 0$ in Eqs. (17)–(19), (28), and (29) to find that

$$\nabla_a[V^{-1}(1 - V^2)^{-2}W_0\nabla^a(W/W_0)] \geqslant 0 \tag{51}$$

in the exterior region. We know that W/W_0 attains its maximum (relative to the exterior region) on the boundary of the star. To employ the maximum principle, we must compute the outward directed (that is out of the exterior region) normal derivative of W/W_0. We find

$$\begin{aligned} d(W/W_0)/dn &= -\lim_{V\to V_s^+} W^{-1/2}\nabla^a V\,\nabla_a(W/W_0) \\ &= \lim_{V\to V_s^+} W^{1/2}W_0^{-2}(W - W_0)dW_0/dV. \end{aligned} \tag{52}$$

At the maximum point $W - W_0 \geqslant 0$ since this quantity vanishes at infinity. Therefore $d(W/W_0)/dn \leqslant 0$ at the maximum point since $dW_0/dV < 0$ there [see Eq. (46)]. The maximum principle therefore guarantees that W/W_0 is constant in the exterior of the star. It follows from Eqs. (17)–(19), (42), and (43) that $R_{abc} = 0$ in the exterior of the star also. Consequently (see Sec. II) the star must be spherical.

The argument given above has implicitly assumed that only a single star was present. The argument can be easily generalized to eliminate the possibility of mutiple static uniform density stars. Even if multiple static stars existed, the argument using Eqs. (48) and (49) would still imply that the maximum of W/W_0 would occur on the surface of one of the stars. If one chooses this maximal star to supply the parameters ρ and V_s for Eqs. (46) and (47), the argument given above will go through exactly as before, with the conclusion that the spacetime is spherical, and consequently only one star is present.

ACKNOWLEDGMENTS

I would like especially to thank S. Chandrasekhar for his enthusiastic support of this work. I also thank D. Brill, J. Hartle, G. Horowitz and B. Mashhoon for helpful suggestions.

[1] A. Avez, R. Acad. Sci. Paris **257**, 3316 (1963).
[2] A. Avez, Ann. Inst. Henri Poincaré **A 1**, 291 (1964).
[3] H.P. Künzle, Commun. Math. Phys. **20**, 85 (1971).
[4] H.P. Künzle and J.R. Savage, preprint (1979).
[5] L. Lindblom, *Fundamental Properties of Equilibrium Stellar Models*, (Ph.D. Dissertation, University of Maryland, College Park, Maryland, 1978), pp. 90–115.
[6] L. Lindblom and D. Brill, in *Essays in General Relativity: A Festschrift for Abraham Taub* ed. by F. Tipler (Academic Press, New York, 1980).
[7] L. Lindblom, J. Math. Phys. **18**, 2352 (1977).
[8] W. Israel, Phys. Rev. **164**, 1776 (1967).
[9] H. Müller zum Hagen, D.C. Robinson, and H.J. Seifert, Gen. Rel. Grav. **4**, 53 (1973).
[10] D.C. Robinson, Gen. Rel. Grav. **8**, 695 (1977).
[11] L. Lindblom, Astrophys.J. **208**, 873 (1976).
[12] L.P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, New Jersey, 1925), pp. 89–92.
[13] H. Müller zum Hagen, D.C. Robinson, and H.J. Seifert, Gen. Rel. Grav. **5**, 61 (1974).
[14] L. Lindblom (to be published) (1980).
[15] H. Müller zum Hagen, Proc. Cambridge Philos. Soc. **67**, 415 (1970).
[16] L. Lindblom, J. Math. Phys. **17**, 1810 (1976).
[17] W. Israel, Nuovo Cimento B **44**, 1 (1966); B **48**, 463 (1967).
[18] L. Bers, F. John, and M. Schecter, *Lectures in Applied Mathematics: Partial Differential Equations* Vol. III (Interscience, New York, 1964), p. 150.

A functional integral formalism for quantum spin systems

Dalcio Kisling Dacol
Department of Physics, University of California, Berkeley, California 94720

(Received 20 March 1979; accepted for publication 15 June 1979)

A functional integral algorithm for the partition function and for the temperature Green's function generating functional of spin 1/2 systems is developed. Starting from the representation of the spin operators in terms of fermion creation and destruction operators we write the above mentioned objects as functional integrals over an infinite dimensional Grassmann algebra. We show how to eliminate the integration over the Grassmann algebra in favor of an integration over function space. The use of the formalism is illustrated with two simple examples.

1. INTRODUCTION

A convenient starting point in the modern approaches to critical phenomena, especially when using renormalization group techniques, is to write the partition function as an integral over the order parameter space. In the case of classical spin systems, like the n-vector model, this procedure is straightforward and is known by the name of Gaussian transformation.[1] For quantum spin systems the Gaussian transformation cannot be applied due to the noncommutativity of the spin operators. In this paper we deal with the noncommutativity problem by writing the spin operators in terms of fermion creation and destruction operators, thus transforming the spin system problem into a many-fermion problem.[2] Once the spin Hamiltonian is written as a fermion Hamiltonian we just use well-known expressions for the partition function and the Green's function generating functional of a many-fermion system as functional integrals over infinite dimensional Grassmann algebras.[3] Then it turns out to be possible to eliminate the integration over the Grassmann algebra by using a functional generalization of the Gaussian transformation. In this way we obtain expressions for the partition function and the temperature Green's function generating functional as functional integrals over function space. Finally, as an illustration, we apply this formalism to two simple systems.

2. FERMION REPRESENTATION OF THE SPIN OPERATORS

Consider a lattice with N points. With each lattice point $\mathbf{R}_i$ we associate a spin $\frac{1}{2}$ operator $\mathbf{S}_i$. Spin operators associated with distinct lattice points commute with each other. Following Ref. 2 we write $[S_j^+ = S_j^x + iS_j^y, S_j^- = (S_j^+)^*]$

$$S_i^+ = (b_i + b_i^*)a_i, \quad S_i^- = a_i^*(b_i + b_i^*), \quad S_i^z = \tfrac{1}{2} - a_i^* a_i. \tag{2.1}$$

The a's and b's are fermion destruction operators and they obey the following anticommutation relations:

$$[a_i, a_j^*]_+ = [b_i, b_j^*]_+ = \delta_{ij},$$
$$[a_i, a_j]_+ = [b_i, b_j]_+ = [a_i, b_j]_+ = [a_i, b_j^*]_+ = 0, \tag{2.2}$$
$$\delta_{ij} = 1 \quad \text{if } \mathbf{R}_i = \mathbf{R}_j, \quad \delta_{ij} = 0 \quad \text{if } \mathbf{R}_i \neq \mathbf{R}_j.$$

Using the relations (2.2) one can easily show that the operators defined in (2.1) do satisfy the angular momentum commutation relations and that spin operators associated with different lattice points commute.

Let $H[\mathbf{S}]$ be a Hamiltonian for the spin system written in terms of the operators $\mathbf{S}_i$ and let $H[a,b]$ denote the same Hamiltonian after replacing the spin operators by their expressions in terms of the fermion operators. Let us indicate by Tr_s the trace operation in spin space and by Tr_f the trace operation in the space spanned by the fermion operators. Then it is straightforward to show that for the partition function we obtain

$$Z = \mathrm{Tr}_s[e^{-\beta H[\mathbf{S}]}] = 2^{-N}(\mathrm{Tr}_f[e^{-\beta H[a,b]}]). \tag{2.3}$$

The first equality above is the definition of the partition function and the factor 2^{-N} in the right-hand side is a consequence of the fact that the fermion space has dimension 2^{2N} (we have two types of fermions per lattice site) and the spin space has dimension 2^N.

In the following sections we will consider the Hamiltonian ($J_{ii} = K_{ii} = 0$)

$$H[\mathbf{S}] = -\sum_{ij}[J_{ij}(S_i^x S_j^x + S_i^y S_j^y) + K_{ij}(S_i^z S_j^z)] - \sum_i [\mathbf{h}\cdot\mathbf{S}_i]. \tag{2.4}$$

Using the representation (2.1) and normal ordering the fermion operators we get

$$\begin{aligned} H[a,b] = -N\gamma &+ \sum_i [\alpha a_i^* a_i - h^+ a_i^*(b_i + b_i^*) \\ &- h^-(b_i + b_i^*)a_i] \\ &+ \sum_{ij} [K_{ij} a_i^* a_j^* a_i a_j - J_{ij} a_i^*(b_i b_j \\ &+ b_i^* b_j^* + b_i^* b_j - b_j^* b_i)a_j]. \end{aligned} \tag{2.5}$$

In order to get the above expression we assumed

$$\begin{aligned} J_{ij} &= (1/N)\sum_{\mathbf{k}} \{ J(\mathbf{k}) \exp[i\mathbf{k}\cdot(\mathbf{R}_i - \mathbf{R}_j)]\}, \\ K_{ij} &= (1/N)\sum_{\mathbf{k}} \{ K(\mathbf{k}) \exp[i\mathbf{k}\cdot(\mathbf{R}_i - \mathbf{R}_j)]\}, \end{aligned} \tag{2.6}$$

where the $\mathbf{k}$ summation is performed over the first Brillouin zone. The constants in (2.5) are

$$\begin{aligned} \gamma &= \tfrac{1}{4}[2h_z + K(0)], \\ \alpha &= h_z + K(0), \\ h^{\pm} &= (h_x \pm ih_y)/2. \end{aligned} \tag{2.7}$$

3. FUNCTIONAL INTEGRAL REPRESENTATION FOR Z

For a Hamiltonian defined as a polynomial in fermion creation and destruction operators it has been shown[3] that the partition function can be expressed as a functional integral over an infinite dimensional Grassmann algebra. With the Hamiltonian $H[a,b]$ [Eq. (2.5)], we get

$$Z=2^{-N}\left[\lim_{M\to\infty}\prod_{n=1}^{M}\prod_{i}\int da_i^n\, d\bar{a}_i^n\, db_i^n\, d\bar{b}_i^n(e^{-S})\right], \quad (3.1)$$

$$S=\sum_{n=1}^{M}\left(\sum_i\{\bar{a}_i^n(a_i^{n+1}-a_i^n)+\bar{b}_i^n \times(b_i^{n+1}-b_i^n)+(\beta/M)H[a^n,b^n]\}\right). \quad (3.2)$$

In Eq. (3.2) by $H[a^n,b^n]$ we mean the functional obtained by replacing in (2.5) the fermion operators a_i, a_i^*, b_i, b_i^*, by their Grassmann algebra representatives a_i^n, $\bar{a}_i^n$, b_i^n, $\bar{b}_i^n$. The a^n's and b^n's obey antiperiodical boundary conditions with respect to the index n, i.e., $a_i^{M+1}=-a_i^1$, etc. For details about Grassmann algebras and the definition of integration over them we refer the reader to Berezin's book.[4] The operations implied in (3.1) and (3.2) can be written in the following symbolic form:

$$Z=2^{-N}\left[\prod_i\int Da_i\, D\bar{a}_i\, Db_i\, D\bar{b}_i(e^{-S})\right], \quad (3.3)$$

$$S=\int_0^\beta\left[\sum_i\left(\bar{a}_i\frac{\partial}{\partial t}a_i+\bar{b}_i\frac{\partial}{\partial t}b_i\right)+H[a(t),b(t)]\right]. \quad (3.4)$$

In (3.4), $a_i(t_n)=a_i^n$ with $t_n=(n-1)\beta/M$, and the antiperiodical boundary condition in n now means $a_i(\beta) = -a_i(0)$, etc.

Our next step is to show how to eliminate the Grassmann variables from the problem and express Z as a functional integral over function space. Consider the part of S [Eq (3.2)] that is of fourth order in the Grassmann variables: This term can be written as

$$-\frac{\beta}{M}\sum_{n=1}^{M}\left\{\sum_{\mathbf{k}}[K(\mathbf{k})\bar{\eta}_{\mathbf{k}}^n\eta_{\mathbf{k}}^n+J(\mathbf{k})\bar{\lambda}_{\mathbf{k}}^n\lambda_{\mathbf{k}}^n]\right\}, \quad (3.5)$$

$$\eta_{\mathbf{k}}^n=(1/\sqrt{N})\sum_j\{\bar{a}_j^n a_j^n\exp[-(i\mathbf{k}\cdot\mathbf{R}_j)]\},$$

$$\lambda_{\mathbf{k}}^n=(1/\sqrt{N})\sum_j\{(b_j^n+\bar{b}_j^n)a_j^n\exp[-(i\mathbf{k}\cdot\mathbf{R}_j)]\}.$$

Notice that the η's and λ's are commutative quantities since they are linear combinations of quadratic forms of Grassmann variables. By using the identity[5]

$$\int d^2z\,(\exp\{-[\pi w\bar{z}z+(\sqrt{\pi})(uz+v\bar{z})]\}) = (\exp[uv/w])/w, \quad d^2z=dx\,dy, \quad z=x+iy,$$

which is valid for any complex numbers u and v and for $\mathrm{Re}(w)>0$, we get

$$\exp\left(\frac{\beta}{M}\right)\sum_{n=1}^{M}\left\{\sum_{\mathbf{k}}[K(\mathbf{k})\bar{\eta}_{\mathbf{k}}^n\eta_{\mathbf{k}}^n+J(\mathbf{k})\bar{\lambda}_{\mathbf{k}}^n\lambda_{\mathbf{k}}^n]\right\} = \prod_{n=1}^{M}\prod_{\mathbf{k}}\left\{(M^{-2})\int d^2\varphi_{\mathbf{k}}^n\int d^2\Theta_{\mathbf{k}}^n \times\exp(-A[\varphi,\Theta,\eta,\lambda])\right\}, \quad (3.6)$$

$$A[\varphi,\Theta,\eta,\lambda]=\left(\frac{1}{M}\right)\sum_{n=1}^{M}\left(\sum_{\mathbf{k}}[\pi\bar{\varphi}_{\mathbf{k}}^n\varphi_{\mathbf{k}}^n+\pi\bar{\Theta}_{\mathbf{k}}^n\Theta_{\mathbf{k}}^n \times D(\mathbf{k})(\bar{\lambda}_{\mathbf{k}}^n\varphi_{\mathbf{k}}^n+\lambda_{\mathbf{k}}^n\bar{\varphi}_{\mathbf{k}}^n) - F(\mathbf{k})(\bar{\eta}_{\mathbf{k}}^n\Theta_{\mathbf{k}}^n+\eta_{\mathbf{k}}^n\bar{\Theta}_{\mathbf{k}}^n)]\right),$$

$$D(\mathbf{k})=[\pi\beta J(\mathbf{k})]^{1/2}, \quad F(\mathbf{k})=[\pi\beta K(\mathbf{k})]^{1/2}. \quad (3.7)$$

Since the Grassmann variables obey antiperiodical boundary conditions with respect to the index n, it follows that the η's and λ's should obey periodical boundary conditions and therefore the φ's and Θ's must also obey periodical boundary conditions. From (3.6) and (3.7) after some obvious manipulations we obtain, in symbolic form,

$$\exp\int_0^\beta dt\left\{\sum_{ij}[K_{ij}\bar{a}_i a_i\bar{a}_j a_j+J_{ij}\bar{a}_i(b_i+\bar{b}_i)(b_j+\bar{b}_j)a_j]\right\} = \left\{\prod_i\int D^2\varphi_i\int D^2\Theta_i\exp(-A[\varphi,\Theta,a,b])\right\}\Big/ \left\{\prod_i\int D^2\varphi_i\int D^2\Theta_i\exp(-A[\varphi,\Theta,0,0])\right\}, \quad (3.8)$$

$$A[\varphi,\Theta,a,b]=(1/\beta)\int_0^\beta dt\left[\pi\sum_i(\bar{\varphi}_i\varphi_i+(\bar{\Theta}_i\Theta_i) +\sum_{ij}D_{ij}(\bar{\varphi}_j(b_i+\bar{b}_i)a_i+\varphi_j\bar{a}_j(b_i+\bar{b}_i)) -\sum_{ij}F_{ij}((\bar{\Theta}_j+\Theta_j)\bar{a}_i a_i)\right], \quad (3.9)$$

$$D_{ij}=(1/N)\sum_{\mathbf{k}}D(\mathbf{k})\exp[i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_j)], \quad (3.10)$$

$$F_{ij}=(1/N)\sum_{\mathbf{k}}F(\mathbf{k})\exp[i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_j)]. \quad (3.11)$$

The functions $\varphi_i(t)$ and $\Theta_i(t)$ must satisfy periodical boundary conditions in t, i.e., $\varphi_i(0)=\varphi_i(\beta)$, $\Theta_i(0)=\Theta_i(\beta)$. After substituting (3.8) into (3.3) the dependence of the integrand on the Grassmann variables is quadratic and therefore we can perform the functional integral over these variables.[4] The final result is

$$Z=Q\exp[N(\beta\gamma+\ln(1+e^{-\beta\alpha}))], \quad (3.12)$$

$$Q=\left\{\prod_i\int D^2\varphi_i\int D\sigma_i\exp(-L\{\varphi,\sigma\})\right\}\Big/ \left\{\prod_i\int D^2\varphi_i\int D\sigma_i\exp(-L_0[\varphi,\sigma])\right\}, \quad (3.13)$$

$$L_0[\varphi,\sigma]=(\pi/\beta)\int_0^\beta dt\left[\sum_i(\bar{\varphi}_i\varphi_i+\sigma_i^2)\right], \quad (3.14)$$

$$L[\varphi,\sigma]=L_0[\varphi,\sigma]-\sum_i\{\mathrm{Tr}[\ln(1-S_i)] +\tfrac{1}{2}\mathrm{Tr}[\ln(1-R_i)]\}, \quad (3.15)$$

$$S_i(t,t')=(2/\beta)A(t-t')\sum_j[F_{ij}\sigma_j(t')], \quad (3.16)$$

$$R_i(t,t')=(2/\beta^2)\int_0^\beta ds\,B(t-s)\Big\{P_i(s,t') \times\Big[\sum_j D_{ij}\bar{\varphi}_j(s)-\beta h^-\Big] \times\Big[\sum_k D_{ik}\varphi_k(t')-\beta h^+\Big]-P_i(t',s)$$

$$\times\Big[\sum_j D_{ij}\bar{\varphi}_j(t') - \beta h^-\Big]$$
$$\times\Big[\sum_k D_{ik}\varphi_k(s) - \beta h^+\Big]\Big\}, \quad (3.17)$$

$$A(t-s) = (1/\beta)\sum_m \{A_m \exp[-if_m(t-s)]\},$$

$$B(t-s) = (1/\beta)\sum_m \{B_m \exp[-if_m(t-s)]\}, \quad (3.18)$$

$$A_m = (-if_m + \alpha)^{-1}, \quad B_m = (-if_m)^{-1},$$
$$f_m = \pi(2m+1)/\beta,$$

Σ_m means summation over all integers m. For $P_i(t,t')$ we have

$$P_i = [(1-S_i)^{-1}]A, \quad (3.19)$$

where P_i, S_i, and A are interpreted as continuous matrices with elements $P_i(t,t')$, $S_i(t,t')$, and $A(t-t')$. In (3.13) $\varphi_i(t)$ is the complex function introduced before, $\sigma_i(t)$ is a real function [the real part of the previous $\Theta_i(t)$], and the integral over the imaginary part of $\Theta_i(t)$ drops out because the a_i in (3.9) couple only to the real part of $\Theta_i(t)$. The trace of a matrix M with components $M(t,t')$ is defined here as

$$\lim_{\eta\to 0^+}\int_0^\beta dt\, M(t,t+\eta). \quad (3.20)$$

This prescription is a remnant of the normal ordering procedure[6] and its role is exactly the same played by the convergence factor one introduces whenever one encounters a loop in the Feynman graph expansion of the temperature Green's function.[7] $A(t-t')$ and $B(t-t')$ are essentially the "free" temperature Green's function for type a and type b fermions, $P_i(t,t')$ is the Green's function for type a fermions in the presence of an external field $-2\,\Sigma_j F_{ij}\,\sigma_j(t)$.

Equation (3.12) is the expression of the partition function for the Hamiltonian (2.5) as a functional integral over the order parameter space. It is straightforward to show that the magnetization is proportional to the expectation values of the fields φ and σ.

4. THE GREEN'S FUNCTION GENERATING FUNCTIONAL

By introducing sources in the right-hand side of (3.3) we can define a generating functional from which we can obtain the temperature Green's functions as functional derivatives with respect to the sources.[8] Let us define $Z[\eta,\lambda]$ by

$$Z[\eta,\lambda] = 2^{-N}\Big\{\prod_i\int Da_i\, D\bar{a}_i\, Db_i\, D\bar{b}_i \exp(-S[\eta,\lambda])\Big\}, \quad (4.1)$$

$$S[\eta,\lambda] = S + \int_0^\beta dt\Big\{\sum_i [\bar{\eta}_i(b_i+\bar{b}_i)a_i + \eta_i\bar{a}_i(b_i+\bar{b}_i) + \lambda_i(\tfrac{1}{2}-\bar{a}_i a_i)]\Big\}. \quad (4.2)$$

In (4.2) S is given by Eq. (3.4), $\eta_i(t)$ and $\lambda_i(t)$ are respectively a complex and a real function of lattice position and t and they satisfy periodical boundary conditions with respect to t, i.e., $\eta_i(0) = \eta_i(\beta)$, $\lambda_i(0) = \lambda_i(\beta)$.

The Green's functions of interest are

$$G_{ij}^{+-}(t,s) = -\langle T(S_i^+(t)S_j^-(s))\rangle,$$
$$G_{ij}^{zz}(t,s) = -\langle T(S_i^z(t)S_j^z(s))\rangle. \quad (4.3)$$

In (4.3) the symbol $T(\cdots)$ means a time ordering operation, $S_i^\alpha(t)$ means the operator S_i^α in the imaginary time Heisenberg picture,[7] and for an operator X, $\langle X\rangle = \mathrm{Tr}[X\exp(-\beta H)]/\mathrm{Tr}[\exp(-\beta H)]$. From the results of Ref. 3 concerning the Green's functions for fermion operators it follows that

$$G_{ij}^{+-}(t,s) = -\left(\frac{\delta^2 W[\eta,\lambda]}{\delta\eta_j(s)\delta\bar{\eta}_i(t)}\right)_{\eta=\lambda=0}, \quad (4.4)$$

$$G_{ij}^{zz}(t,s) = -\left(\frac{\delta^2 W[\eta,\lambda]}{\delta\lambda_j(s)\delta\lambda_i(t)}\right)_{\eta=\lambda=0}, \quad (4.5)$$

$$W[\eta,\lambda] = Z[\eta,\lambda]/Z[0,0]. \quad (4.6)$$

Other Green's functions can be obtained by taking the appropriate functional derivatives of W; for example, the magnetization is given by the first order functional derivatives.

$Z[\eta,\lambda]$ can be cast into a form similar to Eq. (3.12), the only modifications being:

(1) In the equivalent of Eq. (3.15) add a term $\frac{1}{2}\int_0^\beta dt\,[\Sigma_i\,\lambda_i(t)]$ to the right-hand side.

(2) In the equivalent of Eq. (3.16) replace $\Sigma_j\,F_{ij}\,\sigma_j(t')$ by $\Sigma_j\,F_{ij}\,\sigma_j(t') + \beta\lambda_i(t')$.

(3) In the equivalent of Eq. (3.17) replace $\Sigma_j\,D_{ij}\,\varphi_j(t')$ by $\Sigma_j\,D_{ij}\,\varphi_j(t') + \beta\eta_j(t')$.

After these modifications one can make a change of variables, $\varphi\to\varphi'$, $\sigma\to\sigma'$, where $\varphi_i'(t) = \Sigma_j\,D_{ij}\,\varphi_j(t) + \beta\eta_i(t) - \beta h^+$, $\sigma_i'(t) = \Sigma_j\,F_{ij}\,\sigma_j(t) + \beta\lambda_i(t)$. The effect of this change of variables is to remove the sources and the magnetic field from S_i and R_i [Eq. (3.16) and Eq. (3.17)] and thus to show that η and λ are sources for φ' and σ', therefore establishing the connection between the spin Green's functions and expectation values of products of (φ')'s and (σ')'s.

5. EXAMPLES OF THE USE OF THE FORMALISM

We will now, as an illustration, proceed to apply the formalism developed above to two simple cases. First we will consider a system of noninteracting spins in the presence of an external magnetic field. The partition function of this system is given by [from (2.5), (3.3), and (3.4) with $J_{ij} = K_{ij} = 0$]

$$Z = [2Q\exp(\beta h_z/2)]^N, \quad (5.1)$$

$$Q = \Big(\int Da\, D\bar{a}\, Db\, D\bar{b}\, e^{-E}\Big)\Big/\Big(\int Da\, D\bar{a}\, Db\, D\bar{b}\, e^{-E_0}\Big), \quad (5.2)$$

$$E_0 = \int_0^\beta dt\Big(\bar{a}\frac{\partial}{\partial t}a + \bar{b}\frac{\partial}{\partial t}b\Big), \quad (5.3)$$

$$E = E_0 + \int_0^\beta dt\,[h_z\bar{a}a - h^+(\bar{a}(b+\bar{b})) - h^-(b+\bar{b})a]. \quad (5.4)$$

In order to arrive at (5.1) we used the fact that for a noninteracting system the partition function is a product of N identical factors and that

$$\prod_i\int Da_i\, D\bar{a}_i\, Db_i\, D\bar{b}_i$$

$$\times\exp\left\{-\int_0^\beta dt\left[\sum_i\left(\bar{a}_i\frac{\partial}{\partial t}a_i+b_i\frac{\partial}{\partial t}b_i\right)\right]\right\}=4^N. \quad (5.5)$$

After integrating over the a's we get

$$Q=\int Db\,D\bar{b}\,q[b,\bar{b}]\exp\left[-\int_0^\beta dt\left(\bar{b}\frac{\partial}{\partial t}b\right)\right]\Big/ \int Db\,D\bar{b}\exp\left[-\int dt\left(\bar{b}\frac{\partial}{\partial t}b\right)\right], \quad (5.6)$$

$$q[b,\bar{b}]=\exp\{\mathrm{Tr}[\ln(1+h_zB)]\} +h^+h^-\int_0^\beta dt\int_0^\beta ds \times A(t-s)[b(t)+\bar{b}(t)][b(s)+\bar{b}(s)]. \quad (5.7)$$

Finally, after performing the b integration, we obtain

$$Q=\exp\{\mathrm{Tr}[\ln(1+h_zB)]+\tfrac{1}{2}\mathrm{Tr}[\ln(1+V)]\}, \quad (5.8)$$

$$V(t,s)=(1/\beta)\sum_m[(h_x^2+h_y^2)/(f_m^2+h_z^2)] \times\exp[-if_m(t-s)]. \quad (5.9)$$

Now

$$\mathrm{Tr}[\ln(1+h_zB)]=\lim_{\eta\to0^+}\left\{\sum_m[\ln(1-h_z/if_m)]e^{i\eta f_m}\right\} =-\beta h_z/2+\ln[\cosh(\beta h_z/2)], \quad (5.10)$$

$$\mathrm{Tr}[\ln(1+V)] =\lim_{\eta\to0^+}\left[\sum_m[\ln(1+(h_x^2+h_y^2))/(f_m^2+h_z^2)]e^{i\eta f_m}\right] =2\{\ln[\cosh(\beta h/2)]-\ln[\cosh(\beta h_z/2)]\}, \quad (5.11)$$

$h=(h_x^2+h_y^2+h_z^2)^{1/2}$. Thus we get $Q=\cosh(\beta h/2)e^{-\beta h_z/2}$ and so $Z=[2\cosh(\beta h/2)]^N$, which is the correct expression.[2]

As our second example we will consider the case of the Ising model with a magnetic field in the z direction, in this case $J_{ij}=0$ and $h_x=h_y=0$. Since the Ising spin model is in fact a classical spin model (n-vector model with $n=1$), its partition function has an integral representation in terms of a real field defined on the lattice—(it is not a functional integral but a multiple integral over N real variables, where N is the number lattice points).[1] Since Eqs. (3.12)–(3.17) present an alternative representation for the partition function it is interesting to see how one obtains the integral representation mentioned above starting from (3.12)–(3.17). Let $\mu^i(t)=\Sigma_j F_{ij}\,\sigma_j(t)$, and write

$$\mu^j(t)=\sum_m\mu_m^j e^{-itv_m},\quad v_m=2\pi m/\beta,$$

then

$$\mathrm{Tr}[\ln(1-S_i)]=-\left(\frac{2}{\beta}\right)\mu_0^i T_1-\sum_{n=2}^{\infty}\left(\frac{(2/\beta)^n}{n}\right) \times\left[\sum_{m_1}\cdots\sum_{m_{n-1}}(\mu_{m_n}^i\mu_{m_1}^i\cdots\mu_{m_{n-1}}^i) \times T_n(m_1,\ldots,m_{n-1})\right], \quad (5.12)$$

$$T_1=\lim_{\eta\to0^+}\left[\sum_m A_m e^{i\eta f_m}\right],\quad m_n=-\sum_{a=1}^{n-1}m_a,$$

and

$$T_n(m_1,\ldots,m_{n-1})=\lim_{\eta\to0^+}\left(\frac{1}{n!}\left\{\sum_m e^{i\eta f_m}\times[A_mA_{m-m_1}A_{m-m_1-m_2}\cdots A_{m-m_1-\cdots-m_{n-1}}\right.\right.$$
$$+\text{ similar terms involving all other permutations among the } m_a\ \ (1\leqslant a\leqslant n-1)]\Big\}\Big). \quad (5.13)$$

It is straightforward, although very tedious, to show that $T_n(m_1,\ldots,m_{n-1})=0$ unless $m_1=m_2=\cdots=m_{n-1}=0$. Therefore

$$\mathrm{Tr}[\ln(1-S_i)] =-\sum_{n=1}^{\infty}\left(\frac{(2/\beta)^n}{n}\right)(\mu_0^i)^n\lim_{\eta\to0^+}\left(\sum_m A_m^n e^{i\eta f_n}\right) =\lim_{\eta\to0^+}\left(\sum_m\{\ln[1-(2/\beta)\mu_0^iA_m]\}e^{i\eta f_m}\right) =\ln\{1+\exp[-\beta(\alpha-2\mu_0^i/\beta)]\}-\ln(1+e^{-\beta\alpha}). \quad (5.14)$$

Now $\mu_m^i=\Sigma_j F_{ij}\,\sigma_j^m$, and $\sigma_j^{-m}=\bar{\sigma}_j^m$ because σ_j is real. The change of variables from σ_j to $\mathrm{Re}\sigma_j^m$ and $\mathrm{Im}\sigma_j^m$ in (3.13) has unity Jacobian if we take $m\geqslant0$ only. Since $\mathrm{Tr}[\ln(1-S_i)]$ depends only on σ_j^0, the integrals over the σ_j^m, $m\neq0$, in the numerator are cancelled by similar integrals in the denominator and for the partition function we get (with $\sigma_j=\sigma_j^0$)

$$Z=\prod_i\int_{-\infty}^{\infty}d\sigma_i\,e^{-E(\sigma)}, \quad (5.15)$$

$$E(\sigma)=\sum_i\left(-\beta\gamma+\pi\sigma_i^2 -\ln\left\{1+\exp\left[-(\beta\alpha-2\sum_j F_{ij}\,\sigma_j)\right]\right\}\right). \quad (5.16)$$

In (5.15) Z is expressed as a multiple integral over the N real variables σ_i, $-\infty<\sigma_i<\infty$. Let us make the following change of variables $\sigma_i=F(0)/2\pi-\varphi_j$ in (5.15). We get

$$Z=\prod_i\int_{-\infty}^{\infty}d\varphi_i\exp\left(-\sum_i\left\{\pi\varphi_i^2 -\ln\left[2\cosh(\beta h_z/2+\sum_j F_{ij}\,\varphi_j)\right]\right\}\right). \quad (5.17)$$

The expression (5.17) for Z is the representation mentioned before, which can be obtained through a Gaussian transformation.[1] In order to make this point more explicit let us observe that the Hamiltonian for this system can be written as

$$H=-(1/\pi\beta)\sum_i\left(\sum_j F_{ij}\,S_j^z\right)^2-h_z\left(\sum_i S_i^z\right), \quad (5.18)$$

and the partition function can be written as

$$Z=\mathrm{Tr}[e^{-\beta H}] =\mathrm{Tr}\left\{\left[\exp(1/\pi)\sum_i\left(\sum_j F_{ij}\,S_j^z\right)^2\right]\exp(\beta h_z\sum_i S_i^z)\right\} =\prod_i\int_{-\infty}^{\infty}d\varphi_i\exp\left(-\pi\sum_i\varphi_i^2\right)\mathrm{Tr}\left[\exp\sum_i \times\left(\beta h_z+2\sum_j F_{ij}\,\varphi_j\right)S_i^z\right]. \quad (5.19)$$

Now

$$\mathrm{Tr}\Big[\exp\sum_i\Big(\beta h_z+2\sum_j F_{ij}\,\varphi_j\Big)S_i^z\Big]$$
$$=\exp\sum_i\Big\{\ln\Big[2\cosh\Big(\beta h_z/2+\sum_j F_{ij}\,\varphi_j\Big)\Big]\Big\}$$

and so we see that (5.17) is equal to (5.19). This calculation provides a consistency check on the manipulations carried out in Sec. 3.

As a final remark we would like to point out that applications of the techniques discussed in this paper are not restricted to spin models. For example, we can use this formalism to eliminate the spin variables in the functional integral treatment of problems like electrons interacting with spin $\frac{1}{2}$ magnetic impurities in solids, or a system of two level atoms in interaction with a radiation field, etc.

ACKNOWLEDGMENTS

I am grateful to Professor L.M. Falicov for a helpful suggestion. I also want to acknowledge financial support from CAPES (an agency of the Brazilian government) in the form of a graduate studies fellowship.

[1]D.J. Amit, *Field Theory, The Renormalization Group, and Critical Phenomena* (McGraw-Hill, London, 1978).
[2]D.C. Mattis, *The Theory of Magnetism* (Harper and Row, New York, 1965).
[3]D.E. Soper, Phys. Rev. D **18**, 4590 (1978); Y. Ohnuki and T. Kashina, Prog. of Theor. Phys. **60**, 548 (1978).
[4]F.A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[5]B.R. Mollow nd R.J. Glauber, Phys. Rev. **160**, 1076 (1967).
[6]H. Kleinert, "Nonlinear Field Equations in Collective Phenomena," Freie Universitat preprint, Berlin, 1977).
[7]A.L., Fetter and J.D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971).
[8]E.S. Abers and B.W. Lee, Phys. Rep. C **9** 1 (1973).

Structure of the BBGKY hierarchy near phase transition

G. V. Ramanathan [a)]
Department of Mathematics, University of Illinois, Box 4348, Chicago, IL 60680
C. Jedrzejek [b),c)]
Departmen of Energy Engineering, University of Illinois, Box 4348, Chicago, IL 60680

(Received 6 November 1979; accepted for publication 25 January 1980)

A nonperturbative method, as opposed to diagrammatic expansions, is used to study critical phenomena in a fluid with a small hard core and a weak, long-ranged attractive potential. Using the natural small parameter related to the inverse of the range of the attractive potential, spatially uniformly valid asymptotic estimates are made for the magnitudes of all correlations (which are defined as the excess from the generalized superposition approximation) in a region near phase transition in arbitrary number of dimensions. It is shown that if the dimension of the space is larger than four, the correlation hierarchy truncates at the three-body level. The pair correlation satisfies a linear equation. The solution is precisely of Ornstein-Zernike form. For dimensions smaller than four, the hierarchy is still an infinite chain, but considerably simpler than the BBGKY hierarchy. In this case, at the critical point, the correlations are shown to satisfy a scaling law which is the same as that for S^4 spin systems.

1. INTRODUCTION

In an earlier paper[1] one of us presented an asymptotic analytical technique for the study of critical phenomena in a simple fluid in three dimensions. In this work we use this technique to investigate the behavior of correlation functions near phase transition in an arbitrary number of dimensions. Some of our results are analogous to those obtained for spin systems by the use of the renormalization-group method.[2,3] The motivation for our work comes from a desire to understand critical phenomena using basic Hamiltonian mechanics as the starting point instead of an effective phenomenological Hamiltonian, such as that of Laudau–Ginsburg–Wilson, and to study the similarities and differences between fluids and isotropic-spin models.[4,5]

It is well known that the spatial dimension d of a spin system plays a crucial role in the determination of its critical behavior. For $d>4$ the mean-field theory holds and the critical exponents are classical. It can be shown that the pair correlation to its leading order satisfies a linear equation that is decoupled from the effects of higher correlations, even near the critical point.

In this work we derive a similar result for a fluid, namely, that for $d>4$ the critical exponents are classical. However, this happens in a less obvious manner than for spin systems. The BBGKY (Bogoliubov–Born–Green–Kirkwood–Yvon) hierarchy is shown to truncate at the three-body level and not at the two-body level. Still the pair-correlation to its leading order, even including the three-body effects that become significant near phase transition, satisfies a linear equation.

For $d\leqslant 4$ the BBGKY hierarchy can be simplified in the same manner as was shown for $d=3$ in Ref. 1 but one is still left with an infinite chain of equations. In this paper we show that the three-body and higher correlations given by this chain satisfy certain scaling relations at the critical point. These scaling relations turn out to precisely the same as those obtained by Wilson[2] for spin systems.

Before presenting the technical details we shall summarize the asymptotic technique and explain how the above-mentioned results are derived.

Consider a uniform system of identical molecules interacting through an isotropic pair potential consisting of a small hard core and a long-range attractive part. (This will be made precise later.) In this model the hard core plays a minor role, even in the region of phase transition. It keeps the system stable.[6] The ratio of the average interparticular distance to the range of the attractive potential is the fundamental small parameter ϵ in our analysis. (Later on we shall choose it to be similar to the Kac potential where we denote the small parameter by ϵ instead of the usual γ. We do this mainly to distinguish our nonperturbative technique from the perturbative γ expansion.[7-10] Using the infinitesimal nature of ϵ, we determine the asymptotic orders of all the correlation functions by making successive self-consistent estimates as follows. We start with the equation for the pair correlation and obtain a first estimate of its order by neglecting the three-body function. Then we go to the three-body equation, neglect the four-body function, and use the first estimate on the two-body function to obtain a first estimate for the order of the three-body function. This is inserted back in the full equation for the pair function and the resulting modification, if any, of the first estimate on the pair function is investigated. The general s-body function is estimated in a recursive manner by using the earlier estimates on all its predecessors and the first estimate on its immediate successor.

In this analysis a natural parameter arises, namely, $\mu\equiv 1-n\int\psi\, d\mathbf{x}$, where ψ is essentially the negative of the nondimensionalized attractive potential outside the hard

[a)]Supported by a grant from the National Science Foundation, NSF 79-01718.
[b)]Supported partly by a grant from the National Science Foundation, NSF GK 4-3133.
[c)]Permanent address: Department of Theoretical Physics, Institute of Physics, Jagellonian University, Krakow, Poland.

 0022-2488/80/061465-08$1.00

core and -1 inside and n is the number density. As long as this parameter is of the order of unity and positive, the first estimates remain unchanged by the inclusion of higher correlations. In fact their effects are shown to become successively more and more insignificant. This implies that to the leading order of each correlation the hierarchy can be decoupled and solved by an iterative procedure, where each iteration yields an asymptotically smaller term than the one added by the previous one. This has been shown in Ref. 1 for $d = 3$.

However as μ is reduced, the correlation functions begin to develop very long ranges. Even though their amplitudes remain small, when integrated over all space their effects become significant. This is most readily seen if one considers the Fourier transforms of the correlation functions. When μ is sufficiently small, yet positive, in a certain asymptotic (in terms of ϵ) region, for *small* wave numbers, for every s, the $s+1$-body term in the s-body equation becomes as significant as the terms involving its predecessors. In this region, for some value of μ, the compressibility will be infinite. We call this region in μ the region of phase transition.

It should be stressed that the above phenomenon occurs only for small wave numbers. Suppose that μ is in the region of phase transition and hence much smaller than unity. Even if any one of the wave numbers of the s-body function is much larger that $\epsilon\mu^{1/2}$ in magnitude, we show that the effect of the $s+1$-body function in the s-body equation is negligible, whatever the value of μ may be. Thus the hierarchy can be truncated in this case at the s-level. Only when each of the wave numbers is of order $\epsilon\mu^{1/2}$ does the $s+1$-body term make a nonnegligible contribution.

We discover the order of μ in the region of phase transition as follows. Assume that $\mu \ll 1$ and consider the equation for pair correlation for small values of the wave number. We use the estimate on the three-body function in the three-body term in the equation and see what the order of μ should be such that this term is of the same order as that of the pair-correlation function. The result is that $\mu \sim \epsilon^d$ for $d > 4$ and $\mu \sim \epsilon^{2d/6-d}$ for $d \leqslant 4$. It is here that the number of dimensions plays an important role. We show that for $d > 4$, even though the three-body term is significant, the major contribution comes from one of its wave numbers of order ϵ. For $d \leqslant 4$, the major contribution comes from wave numbers of order $\epsilon\mu^{1/2}$. Going then to the three-body equation, being only interested in the regime where one of the wave numbers is of order $\epsilon \gg \epsilon\mu^{1/2}$, we see that the four-body term is insignificant. Thus the hierarchy can be truncated at the three-body level. We solve the two equations and obtain an expression for the pair-correlation function to its leading order, which for small wave numbers is of Ornstein-Zernike form.

For $d \leqslant 4$, when $\mu \sim \epsilon^{2d/6-d}$, the major contributions come from small wave numbers and the hierarchy cannot be truncated. However, certain simplifications occur and the resulting set of equations is considerably simpler than the BBGKY hierarchy. The potential does not appear explicitly anywhere in these equations, but only its integral. At the critical point we assume that the correlation functions are homogeneous functions of the wave numbers. Under this assumption we show that the equations for the three-body and higher correlations are invariant under linear scaling of all the wave numbers if the degree of homogeneity of the s-body correlation is $\frac{1}{2}[s\,a_2 - (s-1)d]$, where a_2 is the degree of homogeneity of the pair function. This is the same formula as obtained by Wilson[2] for spin systems.

2. EQUATIONS FOR CORRELATIONS

We shall denote by $\{p\}_r^s$ a set of p particles chosen from, $r, r+1, \ldots, r+s$, by $f_{\{p\}_r^s}$ the reduced probability distribution of this set, and by $\alpha_{\{p\}_r^s}$ the corresponding correlation functions to be defined shortly. When convenient, we shall simply list the particles. For example, f_{234} and α_{14} will stand respectively for the three-body reduced distribution of particles 2,3, and 4 and the pair correlation of 1 and 4. The correlation functions are defined in terms of the distributions as corrections to generalized superposition:[11-13]

$$f_{12} \equiv f_1 f_2 (1 + \alpha_{12}),$$

$$f_{123} \equiv \frac{f_{12} f_{13} f_{23}}{f_1 f_2 f_3} (1 + \alpha_{123}),$$

$$f_{1234} \equiv \frac{f_{123} f_{124} f_{134} f_{234}}{f_{12} f_{13} f_{14} f_{23} f_{24} f_{34}} f_1 f_2 f_3 f_4 (1 + \alpha_{1234}),$$

$$f_{\{s\}_1^s} \equiv \frac{\Pi f_{\{s-1\}_1^s}}{\Pi f_{\{s-2\}_1^s}} \cdots (\Pi f_{\{1\}_1^s})^{(-1)^s} (1 + \alpha_{\{s\}_1^s}),$$

$\Pi f_{\{j\}_1^s}$ is the product over all distinct sets of j particles chosen from 1, 2,...,s. The function α_{12} is the usual pair-correlation function often denoted in the literature by h.

For correlations in a spatially homogeneous system of identical particles at thermal equilibrium, assuming the existence of the bulk limit, one can derive the following hierarchy of equations from the BBGKY hierarchy.[13] Here n will denote the number density, θ the temperature multiplied by Boltzmann's constant, and ϕ_{ij} the two-body potential between particles i and j, assumed to be spherically symmetric:

$$\frac{\partial}{\partial \mathbf{x}_1} \log(1 + \alpha_{12}) = -\frac{1}{\theta} \frac{\partial \phi_{12}}{\partial \mathbf{x}_1} - \frac{n}{\theta} \int \frac{\partial \phi_{13}}{\partial \mathbf{x}_1} (1 + \alpha_{13}) \times (\alpha_{23} + \alpha_{123} + \alpha_{23}\alpha_{123}) d\mathbf{x}_3. \qquad (2.1)$$

For $s \geqslant 3$

$$\frac{\partial}{\partial \mathbf{x}_1} \log(1 + \alpha_{\{s\}_1^s}) = -\frac{n}{\theta} \int \frac{\partial \phi_{1,s+1}}{\partial \mathbf{x}_1} (1 + \alpha_{1,s+1}) \times \sum \prod_{i=1}^{m} \alpha_{\{p_i\}_1^s, s+1}\, d\mathbf{x}_{s+1}, \qquad (2.2)$$

where the summation is over all products satisfying the following conditions:

(i) The sets $\{p_i\}_1^s$ occurring in each product are distinct.
(ii) If $p_i = 1$, the particle 1 is not a member of $\{p_i\}_1^{s+1}$.
(iii) If $\cup_{i=1}^m \{p_i\}_1^s = \{p\}_1^s$, $p = s-1$ or $p = s$.
(iv) If $p = s-1$, the particle 1 is not a member of $\{p\}_1^s$.

This means that, for example, for $s = 4$ terms such as $\alpha_{25}\alpha_{235}$ and $\alpha_{125}\alpha_{235}$ cannot occur in the sum, the former because of (iii) and the latter because of (iv).

We now consider a potential that consists of a hard core of radius δ and an attractive part of the form $-\epsilon^d v(\epsilon r)$, where ϵ is an infinitesimally small parameter and v a positive, continuous, bounded function such that

$$\int_0^\infty r^{d-1} v(r) dr < \infty .$$

We shall assume that δ is of the order ϵ^d for $d > 4$ and $\epsilon^{2d/(6-d)}$ for $d \leqslant 4$. The exponent is chosen such that the effect of the hard core will be significant in a certain asymptotic region near phase transition. This will be made clearer later in the analysis.

One obtains consistent asymptotic estimates of the magnitudes of the correlations in terms of ϵ. To begin with, a first estimate on the order of α_{12} is made from the equation obtained by neglecting α_{123}. Then in the equation for α_{123}, α_{1234} is neglected and the first estimate on α_{12} is used to find the order of α_{123}. This is put back in the equation for α_{12} and one looks for modifications, if any, of the first estimate. This procedure is continued recursively for all the correlations. The technical details of the method are presented in Ref. 1.

We shall start with a very weak corollary of the results given in Ref. 1. It is that the three-body and higher correlations are all uniformly small in their domains and the pair correlation is small outside the hard core. Without going into any mathematical detail, one could convince oneself of this by inspection of Eq. (2.1) and (2.2) and observing that outside the hard core the potential is small. It follows then that the logarithms in (2.1) and (2.2) can be expanded and the three-body and higher correlations can be neglected next to terms on the order of unity. More precisely, if we set

$$\psi_{12} = -1 + \exp(-\phi_{12}/\theta), \tag{2.3}$$

we can write,

$$\frac{\partial \alpha_{12}}{\partial \mathbf{x}_1} - n \int \frac{\partial \psi_{13}}{\partial \mathbf{x}_1} \alpha_{23}\, d\mathbf{x}_3 = \frac{\partial \psi_{12}}{\partial \mathbf{x}_1} + n \int \frac{\partial \psi_{13}}{\partial \mathbf{x}_1} \alpha_{123}\, d\mathbf{x}_3 + o\left(\frac{\partial \alpha_{12}}{\partial \mathbf{x}_1} \right), \tag{2.4}$$

$$\frac{\partial \alpha_{123}}{\partial \mathbf{x}_1} - n \int \frac{\partial \psi_{14}}{\partial \mathbf{x}_1} \alpha_{234}\, d\mathbf{x}_4 = n \int \frac{\partial \psi_{14}}{\partial \mathbf{x}_1} \alpha_{24}\alpha_{34}\, d\mathbf{x}_4 + n \int \frac{\partial \psi_{14}}{\partial \mathbf{x}_1} \alpha_{1234}\, d\mathbf{x}_4 + o\left(\frac{\partial \alpha_{123}}{\partial \mathbf{x}_1} \right), \tag{2.5}$$

$$\frac{\partial \alpha_{1234}}{\partial \mathbf{x}_1} - n \int \frac{\partial \psi_{15}}{\partial \mathbf{x}_1} \alpha_{2345}\, d\mathbf{x}_5 = n \int \frac{\partial \psi_{15}}{\partial \mathbf{x}_1} [\alpha_{25}\alpha_{35}\alpha_{45} + \alpha_{25}\alpha_{345} + \alpha_{35}\alpha_{245} + \alpha_{45}\alpha_{235}] + n \int \frac{\partial \psi_{15}}{\partial \mathbf{x}_1} \alpha_{12345}\, d\mathbf{x}_5 + o\left(\frac{\partial \alpha_{1234}}{\partial \mathbf{x}_1} \right). \tag{2.6}$$

More generally, for $s \geqslant 3$,

$$\frac{\partial \alpha_{\{s\}_1^s}}{\partial \mathbf{x}_1} - n \int \frac{\partial \psi_{1,s+1}}{\partial \mathbf{x}_1} \alpha_{\{s\}_2^{s+1}}\, d\mathbf{x}_{s+1} = n \int \frac{\partial \psi_{1,s+1}}{\partial \mathbf{x}_1} \sum \prod_{j=1}^{m} \alpha_{\{p_j\}_2^s, s+1}\, d\mathbf{x}_{s+1} + n \int \frac{\partial \psi_{1,s+1}}{\partial \mathbf{x}_1} \alpha_{\{s+1\}_1^{s+1}}\, d\mathbf{x}_{s+1} + o\left(\frac{\partial \alpha_{\{s\}_1^s}}{\partial \mathbf{x}_1} \right), \tag{2.7}$$

where the sum is over all products such that the $\{p_j\}$'s in each product partition the set $\{s-1\}_2^s$. For details the reader is referred to Refs. 1 and 13.

Let us define the function $G_{12}(\mathbf{x}_1 - \mathbf{x}_2)$ by

$$G_{12} - n \int \psi_{13} G_{23}\, d\mathbf{x}_3 = \psi_{12}. \tag{2.8}$$

Clearly (2.8) has a unique integrable solution G_{12} as long as the quantity

$$\mu \equiv 1 - n \int \psi_{12}\, d\mathbf{x}_2 \tag{2.9}$$

remains positive.[14] If $\mu \sim 1$, it can be readily shown that the three-body term in (2.4) is unimportant and, more generally, for $s \geqslant 2$,

$$\left| n \int \frac{\partial \psi_{1,s+1}}{\partial \mathbf{x}_1} \alpha_{\{s+1\}_1^{s+1}}\, d\mathbf{x}_{s+1} \right| = o\left(\left| \frac{\partial \alpha_{\{s\}_1^s}}{\partial \mathbf{x}_1} \right| \right).$$

In fact, one can solve every equation of the hierarchy to the leading order in each correlation to obtain

$$\alpha_{12} = G_{12} + o(\alpha_{12}), \tag{2.10}$$

$$\alpha_{123} = n \int G_{14}\alpha_{24}\alpha_{34}\, d\mathbf{x}_4 + o(\alpha_{123}), \tag{2.11}$$

$$\alpha_{\{s\}_1^s} = n \int G_{1,s+1} \sum \prod \alpha_{\{p_j\}_2^s, s+1}\, d\mathbf{x}_{s+1} + o(\alpha_{\{s\}_1^s}), \quad s \geqslant 3, \tag{2.12}$$

where the summation is the same as the one in (2.7). The range of G_{12} is of the order of $(\epsilon\mu^{1/2})^{-1}$. This can be seen by Fourier transforming (2.8) and writing a Taylor expansion for the numerator and denominator for small wave numbers. Thus, if μ is of order unity, all the correlations have ranges of the order of the range of the potential. The asymptotic orders of the correlations are obtained from (2.10)–(2.12) as

$$\alpha_{12} \sim \epsilon^d, \tag{2.13}$$

$$\alpha_{\{s\}_1^s} \sim \epsilon^{(s-1)d}. \tag{2.14}$$

One also has estimates on the derivatives from the hierarchy itself:

$$\left| \frac{\partial \alpha_{\{s\}_1^s}}{\partial \mathbf{x}_1} \right| \sim \epsilon^{(s-1)d+1} \tag{2.15}$$

for particle separations of the order of ϵ^{-1}.

It is easy to verify that

$$\left| \int \frac{\partial \psi_{1,s+1}}{\partial \mathbf{x}_1} \alpha_{\{s+1\}_1^{s+1}}\, d\mathbf{x}_{s+1} \right| = O(\epsilon^{sd+1}) < \left| \frac{\partial \alpha_{\{s\}_1^s}}{\partial \mathbf{x}_1} \right| \sim \epsilon^{(s-1)d+1}.$$

If $\mu \sim 1$,

$$\left| \frac{\partial \alpha_{\{s\}_1^s}}{\partial \mathbf{x}_1} - \frac{n}{\theta} \int \frac{\partial \psi_{1,s+1}}{\partial \mathbf{x}_1} \alpha_{\{s+1\}_1^{s+1}}\, d\mathbf{x}_{s+1} \right|$$

is of the same order as $|\partial \alpha_{\{s\}_1^s} / \partial \mathbf{x}_1|$ and the term containing $\alpha_{\{s+1\}_1^{s+1}}$ in the s-body equation can be neglected to the leading order in $\alpha_{\{s\}_1^s}$. Thus the hierarchy truncates at every level.

If $\mu \ll 1$, however, the range of G_{12}, and hence the range of each correlation function, becomes much larger than the potential. For particle separations much larger than the range of the potential, the left-hand side of each of the equations (2.4)–(2.7) becomes small. Therefore the term involving the higher correlation function has to be retained. The asymptotic order of μ for which the left-hand side of the equation balances the term involving the higher correlation will be derived in the next section. Here the dimensionality of the system plays a strong role. For dimensions lower than 4, at every level of the hierarchy the corresponding higher correlation becomes significant. Thus the hierarchy cannot be truncated. For dimensions higher than 4, even though the α_{123} term is significant in the equation, the term containing α_{1234} in the equation for α_{123} is insignificant. Thus the hierarchy truncates at the three level. We call this asymptotic region in μ the region of phase transition. This is the region in which the compressibility becomes infinitely large.

3. CORRELATION FUNCTIONS NEAR PHASE TRANSITION

We now suppose that $\mu \ll 1$ and determine the region in which the higher correlations become significant. In order to do this we use the recursive procedure outlined before. We first obtain an estimate on each correlation by truncating the hierarchy at every level. We then insert this estimate in the term containing $\alpha_{\{s+1\}_1^{s+1}}$ in the $\alpha_{\{s\}_1^s}$ equation and determine the asymptotic region in μ in which the higher correlation term becomes comparable to $\alpha_{\{s\}_1^s}$. We shall demonstrate that for dimensions higher than 4, for any μ, the four and higher correlations remains insignificant. For dimensions lower than four we still have an infinite hierarchy but it is considerably simpler than the BBGKY hierarchy.

Let us define Fourier transforms of the correlation functions by

$$\tilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2,...\mathbf{k}_{s-1}) = \int \alpha_{\{s\}_1^s} \exp - i \sum_{j=1}^{s-1} \mathbf{k}_j\cdot(\mathbf{x}_1 - \mathbf{x}_{j+1}) \prod_{j=1}^{s-1} d\mathbf{x}_{j+1}.$$

We define $\tilde{G}(\mathbf{k})$ and $\tilde{\psi}(\mathbf{k})$ similarly. Observing that

$$\left| \int \frac{\partial \psi_{1,s+1}}{\partial \mathbf{x}_1} \alpha_{\{s+1\}_1^{s+1}}\, d\mathbf{x}_{s+1} \right| \sim \left| \frac{\partial}{\partial \mathbf{x}_1} \int \psi_{1,s+1} \alpha_{1,s+1}\, d\mathbf{x}_{s+1} \right|,$$

we get the following estimates from (2.4)–(2.7):

$$\tilde{\alpha}(\mathbf{k}_1) \sim \tilde{G}(\mathbf{k}_1) + n^2 \tilde{G}(\mathbf{k}_1) \int \tilde{\psi}(\mathbf{k}_2) \tilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2) d\mathbf{k}_2, \tag{3.1}$$

$$\tilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2) \sim n\tilde{G}(\mathbf{k}_1+\mathbf{k}_2)\tilde{\alpha}(\mathbf{k}_1)\tilde{\alpha}(\mathbf{k}_2) + n^2\tilde{G}(\mathbf{k}_1+\mathbf{k}_2)\int \tilde{\psi}(\mathbf{k}_3)\tilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) d\mathbf{k}_3, \tag{3.2}$$

$$\tilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) \sim n\tilde{G}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)[\tilde{\alpha}(\mathbf{k}_1)\tilde{\alpha}(\mathbf{k}_2)\tilde{\alpha}(\mathbf{k}_3) + \tilde{\alpha}(\mathbf{k}_1)\tilde{\alpha}(\mathbf{k}_2,\mathbf{k}_3) + \tilde{\alpha}(\mathbf{k}_2)\tilde{\alpha}(\mathbf{k}_1,\mathbf{k}_3) + \tilde{\alpha}(\mathbf{k}_3)\tilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2)] + n^2\tilde{G}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\int \tilde{\psi}(\mathbf{k}_4)\tilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4) d\mathbf{k}_4, \tag{3.3}$$

$$\tilde{\alpha}(\mathbf{k}_1,...,\mathbf{k}_s) = n\tilde{G}(\mathbf{k}_1+\cdots+\mathbf{k}_s)\sum\prod\tilde{\alpha}(\{\mathbf{k}_j\}) + n^2\tilde{G}(\mathbf{k}_1+\cdots+\mathbf{k}_s) \times \int \tilde{\psi}(\mathbf{k}_{s+1})\tilde{\alpha}(\mathbf{k}_1,...,\mathbf{k}_{s+1}) d\mathbf{k}_{s+1}, \tag{3.4}$$

where the sets $\{\mathbf{k}_j\}$ in each product partition $\{\mathbf{k}_1,...,\mathbf{k}_s\}$. It should be pointed out that originally the correlation functions G and ψ are all nondimensional quantities. After Fourier transforming they become dimensional. They can be nondimensionalized by dividing by an appropriate power of n that is not dependent upon ϵ. Therefore, in order to avoid unwieldy notation, we shall loosely speak of the orders of these functions and other dimensional quantities. For instance, by $|\mathbf{k}| = O(\epsilon)$ we shall mean $|\mathbf{k}|n^{-d} = O(\epsilon)$.

We begin with asymptotic estimates for $\tilde{\psi}(\mathbf{k})$ and $\tilde{G}(\mathbf{k})$. Since the potential is assumed to consist of a small hard core and an attractive part of the form $-\epsilon^d v(\epsilon r)$, where $v(r)$ is a positive, continuous, bounded function, it is easily seen from (2.3) that

$$\tilde{\psi}(\mathbf{k}) \sim \epsilon^d \int v(\epsilon\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{x} = \tilde{v}(\epsilon^{-1}\mathbf{k}).$$

Since $v(\mathbf{x})$ is continuous and bounded, $\tilde{v}(\mathbf{k})$ is integrable. Therefore, for $\epsilon = O(|\mathbf{k}|)$, $\tilde{v}(\epsilon^{-1}\mathbf{k}) = O(\epsilon^d|\mathbf{k}|^{-d})$, and for $|\mathbf{k}| = O(\epsilon)$, $\tilde{v}(\epsilon^{-1}\mathbf{k}) \sim 1$. (This is a rather weak estimate but it is sufficient for our purposes.)
Thus

$$\tilde{\psi}(\mathbf{k}) = O(\epsilon^d|\mathbf{k}|^{-d}) \quad \text{for } \epsilon = O(|\mathbf{k}|) \tag{3.5}$$

and

$$\tilde{\psi}(\mathbf{k}) \sim 1 \quad \text{for } |\mathbf{k}| = O(\epsilon). \tag{3.6}$$

We can write them together as $\tilde{\psi}(\mathbf{k}) = O(\epsilon^d/|\mathbf{k}|^d + \epsilon^d)$.
Fourier transforming (2.8), we obtain

$$\tilde{G}(\mathbf{k}) = \tilde{\psi}(\mathbf{k})/[1 - n\tilde{\psi}(\mathbf{k})]. \tag{3.7}$$

Since it follows from the positivity of $v(\mathbf{x})$ that

$$|\tilde{v}(\mathbf{k})| \leqslant \tilde{v}(0),$$

$$G(\mathbf{k}) \sim \psi(\mathbf{k}) \sim \epsilon^d|\mathbf{k}|^{-d} \quad \text{for } \epsilon = O(|\mathbf{k}|). \tag{3.8}$$

For $|\mathbf{k}| = O(\epsilon)$ we expand $\tilde{\psi}(\mathbf{k})$ around $|\mathbf{k}| = 0$ and make use of the fact that the attractive part of the potential is of the form $-\epsilon^d v(\epsilon r)$ to obtain

$$G(\mathbf{k}) \sim \epsilon^2(k^2 + \epsilon^2\mu)^{-1} \quad \text{for } |\mathbf{k}| = O(\epsilon), \tag{3.9}$$

where $\mu = 1 - n\tilde{\psi}(o)$. Therefore

$$\tilde{G}(\mathbf{k}) \sim \epsilon^2 k^{-2} \quad \text{for } |\mathbf{k}| = O(\epsilon) \quad \text{and } \epsilon\mu^{1/2} = O(|\mathbf{k}|) \tag{3.10}$$

and

$$\tilde{G}(\mathbf{k}) \sim \mu^{-1} \quad \text{for } |\mathbf{k}| = O(\epsilon\mu^{1/2}). \tag{3.11}$$

The following first estimates for $\tilde{\alpha}(\mathbf{k}_1)$ and $\tilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2)$ are obtained by neglecting the higher correlation term in each of (3.1) and (3.2):

$$\tilde{\alpha}(\mathbf{k}_1) \sim \tilde{G}(\mathbf{k}_1), \tag{3.12}$$

$$\tilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2) \sim \tilde{G}(\mathbf{k}_1+\mathbf{k}_2)\tilde{\alpha}(\mathbf{k}_1)\tilde{\alpha}(\mathbf{k}_2) \sim \tilde{G}(\mathbf{k}_1+\mathbf{k}_2)\tilde{G}(\mathbf{k}_1)\tilde{G}(\mathbf{k}_2). \tag{3.13}$$

We now use (3.13) in the right-hand side of (3.1) to estimate the term that was neglected in the first estimate for $\tilde{\alpha}(\mathbf{k}_1)$. By balancing this term with the term retained, we shall find the asymptotic order of μ for which the higher correlations become significant. The term retained in the first estimate for $\widetilde{\alpha}(\mathbf{k}_1)$ is $\widetilde{G}(\mathbf{k}_1)$ and the term neglected is $n^2\widetilde{G}(\mathbf{k}_1)\int\tilde{\psi}(\mathbf{k}_2)\alpha(\mathbf{k}_1,\mathbf{k}_2)d\,\mathbf{k}_2$, which from (3.13) is of order $n^2\widetilde{G}(\mathbf{k}_1)\int\tilde{\psi}(\mathbf{k}_2)\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2)\widetilde{G}(\mathbf{k}_1)G(\mathbf{k}_2)d\,\mathbf{k}_2$. Contributions for this last integral come from various regions of $\mathbf{k}_2$, namely, (a) $(|\mathbf{k}_1|) = O(|\mathbf{k}_2|)$. (b) $|\mathbf{k}_2|\sim|\mathbf{k}_1|$ and $|\mathbf{k}_1+\mathbf{k}_2|\sim|\mathbf{k}_1|$, (c) $|\mathbf{k}_2|\sim|\mathbf{k}_1|$ and $|\mathbf{k}_1+\mathbf{k}_2| = o(|\mathbf{k}_1|)$, (d) $|\mathbf{k}_2| = o(|\mathbf{k}_1|)$. It is readily seen that in each of these regions the term neglected in the first estimate for $\widetilde{\alpha}(\mathbf{k}_1)$ is of order

$$\widetilde{G}(\mathbf{k}_1)\,|\mathbf{k}_2|^d\tilde{\psi}(\mathbf{k}_2)\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2)\widetilde{G}(\mathbf{k}_1)\widetilde{G}(\mathbf{k}_2). \tag{3.14}$$

First suppose that $|\mathbf{k}_1| = O(\epsilon\mu^{1/2})$. From (3.5)–(3.11) it is readily seen that if $\epsilon = o(\mathbf{k}_2)$, (3.14) is $o(\epsilon^d\mu^{-1}\widetilde{G}(\mathbf{k}_1))$. If $|\mathbf{k}_2| = O(\epsilon)$, it is of order

$$\epsilon^4\mu^{-1}|\mathbf{k}_2|^d(\mathbf{k}_2^2+\epsilon^2\mu)^{-2}\widetilde{G}(\mathbf{k}_1). \tag{3.15}$$

If $d>4$, the maximum contribution for this comes from the largest values of $|\mathbf{k}_2|$ in this region, namely, $|\mathbf{k}_2|\sim\epsilon$. Thus (3.14) is of order $\epsilon^d\mu^{-1}\widetilde{G}(\mathbf{k}_1)$. This becomes comparable to the term retained in (3.1), namely, $\widetilde{G}(\mathbf{k}_1)$ when

$$\mu\sim\epsilon^d \quad (d>4). \tag{3.16}$$

If $d\leqslant 4$, (3.15) has maximum value when $|\mathbf{k}_2|\sim\epsilon\mu^{1/2}$. Then (3.14) is of order $\epsilon^d\mu^{(d-6)/2}\widetilde{G}(\mathbf{k}_1)$. This is larger than $\epsilon^d\mu^{-1}$, which, in turn, is larger than the contribution from the region $\epsilon = O(|\mathbf{k}_2$Thus the three-body term becomes significant when

$$\mu\sim\epsilon^{2d/(6-d)} \quad (d\leqslant 4). \tag{3.17}$$

If μ is larger than ϵ^d $(d>4)$ or $\epsilon^{2d/(6-d)}$ $(d\leqslant 4)$, the three-body term in (3.1) remains insignificant and the hierarchy can be truncated at the two-body level.

We emphasize that the major contribution to the integral in (3.1) comes from $|\mathbf{k}_2|\sim\epsilon$ for $d>4$, from $|\mathbf{k}_2|\sim\epsilon\mu^{1/2}$ for $d<4$, and from all of $|\mathbf{k}_2| = O(\epsilon)$ for $d=4$. We shall use these facts later. Next suppose that $\epsilon\mu^{1/2} = o(|\mathbf{k}_1|)$. We consider the various regions of $|\mathbf{k}_2|$ in the same manner as the previous case and conclude that as long as μ is given by (3.16), (3.17), or larger, (3.14) is $o(\widetilde{G}(\mathbf{k}_1))$.

We now summarize the results. If $\mu\equiv 1-n\tilde{\psi}(o)$ is much larger than ϵ^d for $d>4$ and $\epsilon^{2d/(6-d)}$ for $d\leqslant 4$, the hierarchy truncates at the two-body level,

i.e.,

$$\widetilde{\alpha}(\mathbf{k}_1) = \widetilde{G}(\mathbf{k}_1) + o(\widetilde{G}(\mathbf{k}_1)). \tag{3.18}$$

When μ becomes of the order ϵ^d for $d>4$ and $\epsilon^{2d/(6-d)}$ for $d\leqslant 4$, the three-body term becomes significant for small values of $|\mathbf{k}_1|$. More precisely, when $|\mathbf{k}_1| = O(\epsilon\mu^{1/2})$, where $\mu\sim\epsilon^d$ $(d>4)$ and $\mu\sim\epsilon^{2d/(6-d)}$ $(d\leqslant 4)$, the three-body term is of order $\widetilde{G}(\mathbf{k}_1)$. For larger values of $|\mathbf{k}_1|$, i.e., if $\epsilon\mu^{1/2} = o(|\mathbf{k}_1|)$, the three-body term is $o(\widetilde{G}(\mathbf{k}_1))$ and the hierarchy truncates at the two-body level. Finally, for $d>4$ and $|\mathbf{k}_1| = O(\epsilon\mu^{1/2})$ the main contribution to the three-body term comes from $|\mathbf{k}_2|$ of order ϵ.

4. THREE-BODY FUNCTION FOR $d>4$

In the last section we showed that when μ is sufficiently small the contribution from the three-body term to $\widetilde{\alpha}(\mathbf{k}_1)$ becomes significant when $|\mathbf{k}_1| = O(\epsilon\mu^{1/2})$. We now study the next equation in the hierarchy. We shall show that for $d>4$ the hierarchy truncates at the three-body level. We shall now suppose that $\mu\sim\epsilon^d$ and $|\mathbf{k}_1|\sim\epsilon\mu^{1/2}$. This is the region where the three-body term, namely, $\widetilde{G}(\mathbf{k}_1)\int\tilde{\psi}(\mathbf{k}_2)\alpha(\mathbf{k}_1,\mathbf{k}_2)d\,\mathbf{k}_2$ becomes significant. Recall that for $d>4$ the major contribution for this integral comes from the region $|\mathbf{k}_2|\sim\epsilon$. Now consider Eqs. (3.2) and (3.3), where $|\mathbf{k}_1| = O(\epsilon\mu^{1/2})$ and $|\mathbf{k}_2|\sim\epsilon$.

We now have the following first estimates for the three- and four-body functions.

$$\widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2)\sim\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2)\widetilde{G}(\mathbf{k}_1)\widetilde{G}(\mathbf{k}_2), \tag{4.1}$$

$$\begin{aligned}\widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\sim{}&\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\widetilde{G}(\mathbf{k}_1)\widetilde{G}(\mathbf{k}_2)\widetilde{G}(\mathbf{k}_3)\\&\times[1+\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2)+\widetilde{G}(\mathbf{k}_1+\mathbf{k}_3)\\&+\widetilde{G}(\mathbf{k}_2+\mathbf{k}_3)].\end{aligned} \tag{4.2}$$

We insert estimate (4.2) in the term in (3.2) that was neglected in the first estimate to obtain

$$\begin{aligned}&\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2)\widetilde{G}(\mathbf{k}_1)\widetilde{G}(\mathbf{k}_2)\int\tilde{\psi}(\mathbf{k}_3)\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\widetilde{G}(\mathbf{k}_3)\\&\quad\times[1+\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2)+\widetilde{G}(\mathbf{k}_1+\mathbf{k}_3)+\widetilde{G}(\mathbf{k}_2+\mathbf{k}_3)]d\,\mathbf{k}_3\end{aligned} \tag{4.3}$$

and compare it with the retained term, namely, $\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2)\widetilde{G}(\mathbf{k}_1)\widetilde{G}(\mathbf{k}_2)$. As for the two-body case, the integral in (4.3) is estimated to be

$$\begin{aligned}&|\mathbf{k}_3|^d\tilde{\psi}(\mathbf{k}_3)\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\widetilde{G}(\mathbf{k}_3)[1+\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2)\\&\quad+\widetilde{G}(\mathbf{k}_1+\mathbf{k}_3)+\widetilde{G}(\mathbf{k}_2+\mathbf{k}_3)].\end{aligned} \tag{4.4}$$

Recall that $|\mathbf{k}_1| = O(\epsilon\mu^{1/2})$ and $|\mathbf{k}_2|\sim\epsilon$. Therefore, using (3.5)–(3.11), we see that if $\epsilon = o(|\mathbf{k}_3$(4.4) is $o(1)$ and so the contribution from that region is negligible. If $|\mathbf{k}_3| = O(\epsilon)$, since $|\mathbf{k}_2|\sim\epsilon$, every term containing $\mathbf{k}_2$ in the argument is order 1. Also $\tilde{\psi}(\mathbf{k}_3)\sim 1$. Then (4.4) is of the same order as

$$\begin{aligned}&|\mathbf{k}_3|^d\widetilde{G}(\mathbf{k}_3)[1+\widetilde{G}(\mathbf{k}_1+\mathbf{k}_3)]\\&\quad\sim|\mathbf{k}_3|^d\epsilon^2(\mathbf{k}_3^2+\epsilon^2\mu)^{-1}[1+\epsilon^2/(\mathbf{k}_1+\mathbf{k}_3)^2+\epsilon^2\mu].\end{aligned}$$

For $d>4$ the major contribution to this term comes from the largest values of $|\mathbf{k}_3|$ in this region, i.e., $|\mathbf{k}_3|\sim\epsilon$. Then it is of order $\epsilon^d = o(1)$. Thus we find that the term that was neglected in the first estimate for $\widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2)$ is indeed insignificant. The hierarchy can be legitimately truncated at the three-body level.

A similar set of estimates can also be made in configuration space instead of $\mathbf{k}$ space and it can be shown that for dimensions higher than 4 the hierarchy can be truncated at the three-body level. The $\mathbf{k}$-space estimates are technically a little simpler. In configuration space the corresponding results are as follows:

$$\psi(r)\sim\epsilon^d, \tag{4.5}$$

$$G(r)\sim\epsilon^2 r^{2-d}. \tag{4.6}$$

If $\epsilon^d = o(\mu)$ $(d>4)$ or $\epsilon^{2d/(6-d)}$ $(d\leqslant 4)$,

$$\alpha_{12} = G_{12} + o(G_{12}) \tag{4.7}$$

for all $|\mathbf{x}_1-\mathbf{x}_2|$.

If $\mu\sim\epsilon^d$ $(d>4)$ or $\mu\sim\epsilon^{2d/(6-d)}$ $(d\leqslant 4)$,

$$\alpha_{12} = G_{12} + o(G_{12}) \quad \text{if } |\mathbf{x}_1-\mathbf{x}_2| = o(\epsilon^{-1/2}\mu^{-1/2}). \tag{4.8}$$

When $|\mathbf{x}_1-\mathbf{x}_2|\sim\epsilon^{-1}\mu^{-1/2}$ or larger, however, the three-body term involving α_{123} becomes significant in the the two-body equation (2.4). Observe that, since the range of ψ_{13} in this integral is of order ϵ^{-1}, we need the estimate for α_{123} when $|\mathbf{x}_1 - \mathbf{x}_3| \sim \epsilon^{-1}$. By a method similar to the one used to estimate $\widetilde{\alpha}(\mathbf{k}_1, \mathbf{k}_2)$ in $\mathbf{k}$ space, it can be shown that for $d > 4$, if $|\mathbf{x}_1 - \mathbf{x}_2| \sim \epsilon^{-1}\mu^{-1/2}$ or larger and $|\mathbf{x}_1 - \mathbf{x}_3| \sim \epsilon^{-1}$, the four-body term in the three-body equation is insignificant, i.e., the hierarchy can be truncated at the three-body level. Then (2.5) can be integrated to obtain

$$\alpha_{123} = n \int G_{14}\alpha_{24}\alpha_{34}\, d\mathbf{x}_4 + o\Big(n \int G_{14}\alpha_{24}\alpha_{34}\, d\mathbf{x}_4\Big). \qquad (4.9)$$

It may be pointed out that (4.7) and (4.9) are actual asymptotic solutions and not mere estimates. Using (4.6), we can estimate the right-hand side of (4.9) to be of order $\epsilon^4|\mathbf{x}_1 - \mathbf{x}_4|^{4-d}\alpha_{12}$. Thus for $d > 4$ the maximum contribution comes from $|\mathbf{x}_1 - \mathbf{x}_4| \sim \epsilon^{-1}$. Also, since $|\mathbf{x}_1 - \mathbf{x}_3| \sim \epsilon^{-1}$, $|\mathbf{x}_3 - \mathbf{x}_4| \sim \epsilon^{-1}$ as $|\mathbf{x}_1 - \mathbf{x}_4|$ ranges over values of order ϵ^{-1}; by (4.8), α_{34} in (4.9) can be replaced by G_{34} to obtain

$$\alpha_{123} = n \int G_{14}\alpha_{24}G_{34}\, d\mathbf{x}_4 + o\Big(n \int G_{14}\alpha_{24}\alpha_{34}\, d\mathbf{x}_4\Big). \qquad (4.10)$$

This is equivalent to the statement of the last sentence of Sec. 3.

Now we substitute this in (2.4). Using (2.8), we have

$$\int\int \frac{\partial\psi_{13}}{\partial\mathbf{x}_1} G_{14}\alpha_{24}G_{34}\, d\mathbf{x}_4 = \int \frac{\partial G_{14}}{\partial\mathbf{x}_1} G_{14}\alpha_{24}\, d\mathbf{x}_4 - \int \frac{\partial\psi_{14}}{\partial\mathbf{x}_1} G_{14}\alpha_{24}\, d\mathbf{x}_4 = \frac{\partial}{\partial\mathbf{x}_1}\int H_{14}\alpha_{24}\, d\mathbf{x}_4, \qquad (4.11)$$

where

$$H(r) = \tfrac{1}{2}G^2(r) + K(r)$$

and (4.12)

$$K(r) = \int_r^\infty \frac{d\psi(r)}{dr} G(r)\, dr.$$

Thus $H(r)$ is a known function. Now (2.4) can be integrated to obtain

$$\alpha_{12} - n\int (\psi_{13} + H_{13})\alpha_{23}\, d\mathbf{x}_3 = \psi_{12}. \qquad (4.13)$$

This is a linear equation for α_{12}. The compressibility is proportional to

$$1 + \int \alpha_{12}\, d\mathbf{x}_2 = 1 + \frac{\int \psi_{12}\, d\mathbf{x}_2}{1 - n\int \psi_{12}\, d\mathbf{x}_2 - \int H_{12}\, d\mathbf{x}_2} = \frac{\int \psi_{12}\, d\mathbf{x}_2}{\mu - \int H_{12}\, d\mathbf{x}_2} + 1. \qquad (4.14)$$

When $\mu = \int H_{12}\, d\mathbf{x}_2 \equiv \mu_c$, this is infinite. It is easy to verify that $\int H_{12}\, d\mathbf{x}_2 \sim \epsilon^d$. From (2.3) we see that

$$\mu = 1 + n\sigma^3 - \frac{n}{\theta}\int v(\mathbf{r})\, dr - \tfrac{1}{2}\frac{n^2\epsilon^d}{\theta^2}\int v^2(\mathbf{r})\, dr + o(\epsilon^d). \qquad (4.15)$$

Thus μ is an analytic function of θ. Hence G, K, and therefore H are analytic functions of θ. Let θ_c be a solution to $\mu(\theta_c) = \mu_c$ so that $\mu = \mu_c + (\theta - \theta_c)\nu + o(\theta - \theta_c)$, $\theta > \theta_c$. It follows readily from (4.14) that $\int \alpha_{12}\, d\mathbf{x}_2 \sim (\theta - \theta_c)^{-1}$, which is the classical result. All the critical exponents in this case can be easily shown to be classical.

5. HIERARCHY FOR $d \leqslant 4$

In Sec. 3 we observed that for $d \leqslant 4$, when $\mu \sim \epsilon^{2d/(6-d)}$, the three-body term becomes significant in the two-body equation. To show this we used the first estimate for $\widetilde{\alpha}(\mathbf{k}_1, \mathbf{k}_2)$, namely, $n\widetilde{G}(\mathbf{k}_1 + \mathbf{k}_2)\widetilde{\alpha}(\mathbf{k}_1)\widetilde{\alpha}(\mathbf{k}_2)$ in (3.1), and noted that the last term is of order

$$n^3\widetilde{G}(\mathbf{k}_1)\widetilde{\alpha}(\mathbf{k}_1)\int \widetilde{\psi}(\mathbf{k}_2)\widetilde{G}(\mathbf{k}_1 + \mathbf{k}_2)\widetilde{\alpha}(\mathbf{k}_2)d\,\mathbf{k}_2 \sim \widetilde{\alpha}(\mathbf{k}_1)$$

when $\mu \sim \epsilon^{2d/(6-d)}$. In other words, in this region

$$\widetilde{G}(\mathbf{k}_1)\int \widetilde{\psi}(\mathbf{k}_2)\widetilde{G}(\mathbf{k}_1 + \mathbf{k}_2)\widetilde{\alpha}(\mathbf{k}_2)d\,\mathbf{k}_2 \sim 1. \qquad (5.1)$$

Furthermore, we showed that for $d < 4$ the major contribution to this integral comes from $|\mathbf{k}_2| \sim \epsilon\mu^{1/2}$. Similarly, we can consider the equation for $\widetilde{\alpha}(\mathbf{k}_1, \mathbf{k}_2)$ when $|\mathbf{k}_1|$ and $|\mathbf{k}_2|$ are $O(\epsilon\mu^{1/2})$ and show that when $\mu \sim \epsilon^{2d/(6-d)}$ the four-body term becomes significant in the three-body equation. More generally, suppose $|\mathbf{k}_i| = O(\epsilon\mu^{1/2})$ for $i = 1, 2, ..., s$, and consider the $s+1$-body contribution to the s-body equation, namely, $\widetilde{G}(\mathbf{k}_1 + \cdots + \mathbf{k}_s)\int \widetilde{\psi}(\mathbf{k}_{s+1})\widetilde{\alpha}(\mathbf{k}_1, ..., \mathbf{k}_{s+1})d\,\mathbf{k}_{s+1}$. The first estimate for the $s+1$-body function is [see Eq. (3.4)]

$$\widetilde{\alpha}(\mathbf{k}_1, ..., \mathbf{k}_{s+1}) \sim n\widetilde{G}(\mathbf{k}_1 + \cdots + \mathbf{k}_{s+1})\widetilde{\alpha}(\mathbf{k}_{s+1})\widetilde{\alpha}(\mathbf{k}_1 + \cdots + \mathbf{k}_s).$$

Here we are taking just one term of a sum as a representative for the asymptotic estimate. Inserting this in the integral, we see that the contribution to the s-body equation is of order

$$\widetilde{G}(\mathbf{k}_1 + \cdots + \mathbf{k}_s)\widetilde{\alpha}(\mathbf{k}_1, ..., \mathbf{k}_s) \times \int \widetilde{\psi}(\mathbf{k}_{s+1})\widetilde{G}(\mathbf{k}_1 + \cdots + \mathbf{k}_{s+1})\widetilde{\alpha}(\mathbf{k}_{s+1})d\,\mathbf{k}_{s+1}.$$

Since $|\mathbf{k}_i| \sim \epsilon\mu^{1/2}$, when $\mu \sim \epsilon^{2d/(6-d)}$ [see Eq. (5.1)] this is of order $\widetilde{\alpha}(\mathbf{k}_1, ..., \mathbf{k}_s)$. Thus in the s-body equation the $s+1$-body term must be retained and the major contribution due to the $s+1$-body term comes from $|\mathbf{k}_{s+1}| \sim \epsilon\mu^{1/2}$. Thus the hierarchy cannot be truncated. However, it is possible to derive a considerably simpler hierarchy, as shown below.

Consider the estimates (3.1)–(3.4) when all the $|\mathbf{k}_i|$'s are $O(\epsilon\mu^{1/2})$. Since $\widetilde{G}(\epsilon\mu^{1/2}) \sim \mu^{-1}$,

$$\widetilde{\alpha}(\mathbf{k}_1) \sim \widetilde{G}(\mathbf{k}_1) \sim \mu^{-1}, \qquad (5.2)$$

$$\widetilde{\alpha}(\mathbf{k}_1, \mathbf{k}_2) \sim \mu^{-1}\widetilde{\alpha}(\mathbf{k}_1)\widetilde{\alpha}(\mathbf{k}_2) \sim \mu^{-3}, \qquad (5.3)$$

$$\widetilde{\alpha}(\mathbf{k}_1, ..., \mathbf{k}_s) \sim \mu^{-1}\widetilde{\alpha}(\mathbf{k}_{s+1})\widetilde{\alpha}(\mathbf{k}_1, ..., \mathbf{k}_{s-1}) \sim \mu^{-2s+1} \quad \text{for } s \geqslant 1. \qquad (5.4)$$

Now consider the sum on the right-hand side of Eq. (3.4) for $\widetilde{\alpha}(\mathbf{k}_1, .., \mathbf{k}_s)$, namely, $\Sigma\, \Pi\, \widetilde{\alpha}(\{\mathbf{k}_i\})$. Here the sum is over distinct products $\Pi\, \widetilde{\alpha}(\{\mathbf{k}_i\})$, where the sets $\{\mathbf{k}_i\}$ partition $(\mathbf{k}_1, ..., \mathbf{k}_s)$. Suppose a product has n terms with arguments consisting of $p_1, p_2, ..., p_n$ $\mathbf{k}$'s, respectively, so that $p_1 + p_2 + \cdots + p_n = s$. From (5.4) the order of this product is μ^{-q}, where $q = 2s - n$. This is maximum when $n = 2$ since there should be at least two terms in a product. Thus all the terms that are products of more than two terms can be neglected since they will be asymptotically smaller than the products containing only two terms.

Similar estimates hold in the configuration space. The hierarchy takes the following form for $|\mathbf{x}_{ij}| \sim \epsilon^{-1}\mu^{-1/2}$:

$$\frac{\partial \alpha_{12}}{\partial \mathbf{x}_1} - n\int \frac{\partial \psi_{13}}{\partial \mathbf{x}_1}\alpha_{23}\,d\mathbf{x}_3 = \frac{\partial \psi_{12}}{\partial \mathbf{x}_1} + n\int \frac{\partial \psi_{13}}{\partial \mathbf{x}_1}\alpha_{123}\,d\mathbf{x}_3. \tag{5.5}$$

For $s>2$,

$$\frac{\partial \alpha_{\{s\}_1^s}}{\partial \mathbf{x}_1} - n\int \frac{\partial \psi_{1,s+1}}{\partial \mathbf{x}_1}\alpha_{\{s\}_2^{s+1}}\,d\mathbf{x}_{s+1} = n\int \frac{\partial \psi_{1,s+1}}{\partial \mathbf{x}_1}\sum \alpha_{\{m\}_2^s,s+1}\alpha_{\{n\}_2^s,s+1}\,d\mathbf{x}_{s+1} + n\int \frac{\partial \psi_{1,s+1}}{\partial \mathbf{x}_1}\alpha_{\{s+1\}_1^{s+1}}\,d\mathbf{x}_{s+1}, \tag{5.6}$$

where $\{m\}_2^s$ and $\{n\}_2^s$ are disjoint sets such that $\{m\}_2^s \cup \{n\}_2^s = \{s-1\}_2^s$. It follows from the estimates that

$$\alpha_{\{s\}_1^s} - n\int \psi_{1,s+1}\alpha_{\{s\}_2^{s+1}}\,d\mathbf{x}_{s+1} \sim \mu\alpha_{\{s\}_1^s} = o(\alpha_{\{s\}_1^s}). \tag{5.7}$$

This can also be seen by observing that since $|\mathbf{x}_{ij}|$ is assumed to be of order $\epsilon^{-1}\mu^{-1/2}$ for $i,j = 1, 2,...,s$, $i\neq j$, and $\psi_{1,s+1}$ has a range of order ϵ^{-1}, we can expand $\alpha_{\{s\}_2^{s+1}} = \alpha(\mathbf{x}_2, \mathbf{x}_3, ..., \mathbf{x}_{s+1})$ around the point $(\mathbf{x}_1,...,\mathbf{x}_s)$ and arrive at (5.7). Then the last term in (5.5) can be written as

$$n\int \frac{\partial \psi_{13}}{\partial \mathbf{x}_1}\psi_{14}\alpha_{234}\,d\mathbf{x}_4\,d\mathbf{x}_3 + o\left(n\int \frac{\partial \psi_{13}}{\partial \mathbf{x}_1}\alpha_{123}\,d\mathbf{x}_3\right) = \frac{n}{2}\frac{\partial}{\partial \mathbf{x}_1}\int \psi_{13}\psi_{14}\alpha_{234}\,d\mathbf{x}_4\,d\mathbf{x}_3 + o\left(n\int \frac{\partial \psi_{13}}{\partial \mathbf{x}_1}\alpha_{123}\,d\mathbf{x}_3\right) = \frac{n}{2}\frac{\partial}{\partial \mathbf{x}_1}\int \psi_{13}\alpha_{123}\,d\mathbf{x}_3 + o\left(n\int \frac{\partial \psi_{13}}{\partial \mathbf{x}_1}\alpha_{123}\,d\mathbf{x}_3\right).$$

Now the two-body equation can be integrated with the condition that $\alpha_{12}\to 0$ as $|\mathbf{x}_{12}|\to\infty$ to obtain

$$\alpha_{12} - n\int \psi_{13}\alpha_{23}\,d\mathbf{x}_3 = \psi_{12} + \frac{n}{2}\int \psi_{13}\alpha_{123}\,d\mathbf{x}_3 + o(\alpha_{12}). \tag{5.8}$$

Similarly, we have for $s>2$,

$$\alpha_{\{s\}_1^s} - n\int \psi_{1,s+1}\alpha_{\{s+1\}_2^{s+1}}\,d\mathbf{x}_{s+1} = n\int \psi_{1,s+1}\sum \alpha_{\{m\}_2^s,s+1}\,\alpha_{\{n\}_2^s,s+1}\,d\mathbf{x}_{s+1} + \frac{n}{2}\int \psi_{1,s+1}\alpha_{\{s+1\}_1^{s+1}}\,d\mathbf{x}_{s+1} + o(\alpha_{\{s\}_1^s}), \tag{5.9}$$

where $\{m\}_2^s$ and $\{n\}_2^s$ are disjoint and their union is $\{s-1\}_2^s$.

In terms of Fourier variables, (5.8) can be written as

$$\widetilde{\alpha}(\mathbf{k}_1) = \widetilde{G}(\mathbf{k}_1) + \frac{n}{2}\frac{1}{1-n\widetilde{\psi}(\mathbf{k}_1)} \times \int \widetilde{\psi}(\mathbf{k}_2)\widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2)d\,\mathbf{k}_2 + o(\widetilde{\alpha}(\mathbf{k}_1)). \tag{5.10}$$

Recall the following: When $|\mathbf{k}_1| \sim \epsilon\mu^{1/2}$, the major contribution to the integral on the right-side of (5.10) comes from $|\mathbf{k}_2| \sim \epsilon\mu^{1/2}$, and $\widetilde{\psi}(\mathbf{k})$ for $|\mathbf{k}| \sim \epsilon\mu^{1/2}$ is $(1-\mu)n^{-1} + o(k^2)$. Using these, we can rewrite (5.10) as

$$\widetilde{\alpha}(\mathbf{k}_1) = \widetilde{G}(\mathbf{k}_1) + \tfrac{1}{2}n\widetilde{G}(\mathbf{k}_1)\int \widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2)d\,\mathbf{k}_2 + o(\widetilde{\alpha}(\mathbf{k}_1)). \tag{5.11}$$

Similarly, for $s\geqslant 2$,

$$\widetilde{\alpha}(\mathbf{k}_1,...,\mathbf{k}_s) = n\widetilde{G}(\mathbf{k}_1+\cdots+\mathbf{k}_s)\sum \widetilde{\alpha}(\{\mathbf{k}_i\})\widetilde{\alpha}(\{\mathbf{k}_j\}) + \tfrac{1}{2}n\widetilde{G}(\mathbf{k}_1+\cdots+\mathbf{k}_s)\int \widetilde{\alpha}(\mathbf{k}_1+\cdots+\mathbf{k}_{s+1})d\,\mathbf{k}_{s+1} + o(\alpha(\mathbf{k}_1,...,\mathbf{k}_s)), \tag{5.12}$$

where the set $\{\mathbf{k}_i\}$ and $\{\mathbf{k}_j\}$ are disjoint and their union is $(\mathbf{k}_1,...,\mathbf{k}_s)$. For example, for $s=4$,

$$\begin{aligned}\widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4) = n\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3+\mathbf{k}_4)[&\widetilde{\alpha}(\mathbf{k}_1)\widetilde{\alpha}(\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4) + \widetilde{\alpha}(\mathbf{k}_2)\widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_4) + \widetilde{\alpha}(\mathbf{k}_3)\widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_4)\\ &+ \widetilde{\alpha}(\mathbf{k}_4)\widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) + \widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2)\widetilde{\alpha}(\mathbf{k}_3,\mathbf{k}_4)\\ &+ \widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_3)\widetilde{\alpha}(\mathbf{k}_2,\mathbf{k}_4) + \widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_4)\widetilde{\alpha}(\mathbf{k}_2,\mathbf{k}_3)]\\ &+ n\widetilde{G}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3+\mathbf{k}_4)\times\int \widetilde{\alpha}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4,\mathbf{k}_5)d\,\mathbf{k}_5.\end{aligned} \tag{5.13}$$

It should be pointed out that the hierarchy (5.11) and (5.12) is actually valid for all values of $|\mathbf{k}_1|$. This is because if $\epsilon\mu^{1/2} = o(|\mathbf{k}_1|)$, by (3.18) we have $\alpha(\mathbf{k}_1) = G(\mathbf{k}_1) + o(G(\mathbf{k}_1))$ and we had seen that the last term in the right-hand side of (5.11) is negligible. It becomes significant only for $|\mathbf{k}_1| = O(\epsilon\mu^{1/2})$.

6. SCALING OF CORRELATION FUNCTIONS FOR $d<4$

It may be observed that in the hierarchy (5.11), (5.12) the potential does not appear explicitly. Instead only its integral appears through μ. Solving for $\widetilde{G}(\mathbf{k})$ from (5.11) and substituting it in (5.12), we have for $s\geqslant 2$

$$\widetilde{\alpha}(\mathbf{k}_1,...,\mathbf{k}_s) = n\widetilde{\alpha}(\mathbf{k}_1+\cdots+\mathbf{k}_s)\times\left(1+n\int \widetilde{\alpha}(\mathbf{k}_1+\cdots+\mathbf{k}_s,\mathbf{k}_{s+1})d\,\mathbf{k}_{s+1}\right)^{-1}\times\sum \widetilde{\alpha}(\{\mathbf{k}_i\})\widetilde{\alpha}(\{\mathbf{k}_j\}) + \tfrac{1}{2}\int \widetilde{\alpha}(\mathbf{k}_1,...,\mathbf{k}_{s+1})d\,\mathbf{k}_{s+1}. \tag{6.1}$$

Now μ has also disappeared from the problem. All the correlations are given in terms of the pair function in an indirect way.

At the critical point, for small $|\mathbf{k}_1|$, $\widetilde{\alpha}(\mathbf{k}_1) \gg \widetilde{G}(\mathbf{k}_1)$. In fact $\widetilde{\alpha}(0) = \infty$, while $\widetilde{G}(\mathbf{k}_1) = \mu^{-1}$. Therefore $n\int\alpha(\mathbf{k}_1+\cdots+\mathbf{k}_s,\mathbf{k}_{s+1})d\,\mathbf{k}_{s+1} \gg 1$ and we can write this hierarchy for $s\geqslant 2$ as

$$\widetilde{\alpha}(\mathbf{k}_1,...,\mathbf{k}_s) = \widetilde{\alpha}(\mathbf{k}_1+\cdots+\mathbf{k}_s)\times\left(\int \widetilde{\alpha}(\mathbf{k}_1+\cdots+\mathbf{k}_s,\mathbf{k}_{s+1})d\,\mathbf{k}_{s+1}\right)^{-1}\times\left(\sum \widetilde{\alpha}(\{\mathbf{k}_i\})\widetilde{\alpha}(\{\mathbf{k}_j\}) + \tfrac{1}{2}\int \widetilde{\alpha}(\mathbf{k}_1,...,\mathbf{k}_{s+1})d\,\mathbf{k}_{s+1}\right). \tag{6.2}$$

It is interesting to observe that *none* of the physical parameters such as density, temperature, or μ appears explicitly in this chain of equations. If the pair correlation is given, all the other functions are determined, assuming, of course, that a solution exists for this hierarchy. Now let $\mathbf{k}_i = \lambda\,\mathbf{k}_i'$ and $\widetilde{\alpha}(\lambda\,\mathbf{k}_1',...,\lambda\,\mathbf{k}_s') = \lambda^{a_{s+1}}\beta(\mathbf{k}_1',...,\mathbf{k}_s')$ for $s\geqslant 1$. Substituting this in (6.2), we have for $s\geqslant 2$,

$$\lambda^{a_{s+1}}\beta(\mathbf{k}'_1,\dots,\mathbf{k}'_s) = \lambda^{a_2-a_3-d}\beta(\mathbf{k}'_1+\cdots+\mathbf{k}'_s) \times\left(\int\beta(\mathbf{k}'_1+\cdots+\mathbf{k}'_s,\mathbf{k}'_{s+1})d\,\mathbf{k}'_{s+1}\right)^{-1} \times\left[\sum\lambda^{a_m+a_{s+1-m}}\beta(\{\mathbf{k}'_i\})\beta(\{\mathbf{k}'_j\}) + \tfrac{1}{2}\lambda^{a_{s+2}+d}\int\beta(\mathbf{k}'_1,\dots,\mathbf{k}'_{s+1})d\,\mathbf{k}'_{s+1}\right]. \tag{6.3}$$

Here $2\leqslant m\leqslant s-2$. It is an easy matter to verify that if we set for $s\geqslant 3$,

$$a_s=\tfrac{1}{2}[s\,a_2-(s-2)d], \tag{6.4}$$

λ factors out of (6.3) completely. At the critical point, for small values of $|\mathbf{k}_1|$, if we assume that $\alpha(\lambda\,|\mathbf{k}_1|)=\lambda^{a_2}\widetilde{\alpha}(|\mathbf{k}_1|)$, we notice that the scaling

$$\alpha(\lambda\,\mathbf{k}_1,\dots,\lambda\,\mathbf{k}_s)=\lambda^{a_s}\alpha(\mathbf{k}_1,\dots,\mathbf{k}_s) \tag{6.5}$$

holds, where a_s is given by (6.4). This agrees with the scaling Wilson derived for correlation functions in a spin system.[2]

[1]G.V. Ramanathan, J. Math. Phys. **18**, 1928 (1977).
[2]K.G. Wilson and G. Kogut, Phys. Rep. C **12**, 76 (1974).
[3]E. Brezin, J.C. Le Guillou, and J. Zinn-Justin, in *Phase Transitions and Critical Phenomena* edited by C. Domb and M.S. Green (Academic, New York, 1976), Vol. 6.
[4]J. Hubbard and I. Schofield, Phys. Lett. A **40**, 245 (1972).
[5]D. Valls and J.A. Hertz, Phys. Rev. B **18**, 2367 (1978).
[6]D. Ruelle, *Statistical Mechanics* (Benjamin, New York 1969).
[7]M. Kac, G.E. Uhlenbeck, and P.C. Hemmer, J. Math. Phys. **4**, 216 (1963).
[8]P.C. Hemmer, J. Math. Phys. **5**, 75 (1964).
[9]P.C. Hemmer and J.L. Lebowitz, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M.S. Green (Academic, New York, 1976), Vol. 5b.
[10]J.S. Hoye and G. Stell, Mol. Phys. **32**, 195 (1976).
[11]In this work, by three-body and higher correlations we mean the excess to the generalized superposition approximation. These are the same as the "ϵ_n" functions defined in Ref. 8.
[12]G. Stell in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M.S. Green (Academic, New York, 1976), Vol. 5b.
[13]G.V. Ramanathan, J. Math. Phys. 7, 1507 (1966).
[14]S.G. Mikhlin, *Integral Equations* (Macmillian, New York, 1964).

Construction of two-dimensional quantum electrodynamics

Keiichi R. Ito
Research Institute for Mathematical Sciences, Kyoto University, Kyoto 606, Japan

(Received 10 April 1979; accepted for publication 25 June 1979)

We investigate quantum electrodynamics in two dimensions, $(\mathrm{QED})_2$, in the constructive method. We construct a cutoff Hamiltonian defined on the Fock space with an indefinite metric, and discuss its properties and renormalization counterterms. When the fermion is massless, this model is exactly solvable (known as the Schwinger model). We study the massless model through the Bogolyubov transformation (canonical linear transformation) extended to this Fock space, through which the properties of the Hamiltonian and the physical vacuum are clarified. When the mass of the fermion $M\neq 0$, we discuss the transformation of the renormalized Hamiltonian into the one of the so-called sine-Gordon model. The properties of the operator which implements this transformation are clarified; it is an unbounded operator which is isometric with respect to the indefinite metric.

1. INTRODUCTION

Two-dimensional quantum electrodynamics $[(\mathrm{QED})_2]$ $:\bar{\psi}\gamma^\mu\psi A_\mu:_2$, first investigated by Schwinger,[1] is considered as the first step to construct $(\mathrm{QED})_4$. In the case of QED-type models, we must use an indefinite metric formalism of the vector field which cancels the divergences in the perturbation and makes the theory manifestly covariant. Two-dimensional QED is super-renormalizable by the power-counting theorem[1-3] provided the indefinite metric formalism of the vector field (generalized Stückelberg formalism) is used.

This work starts with the study of the cutoff current $j^\mu_\sigma(x,0)=\int_\Lambda h_\sigma(x-y)j^\mu(y,0)\,dy$ in Sec. 2, where $j^\mu(y,0) = :\bar{\psi}(y,0)\gamma^\mu\psi(y,0):$, $\Lambda=[-L/2,L/2]\subset R$, and h_σ is a momentum cutoff function. We show that j^μ_σ is a densely defined symmetric operator and that j^μ_σ is written in terms of the associated boson when the fermion mass M vanishes. In Sec. 3, we construct a Fock space of the vector field of the Stückelberg formalism which is an indefinite formalism of the vector field. The Fock space has an indefinite inner product[3-7] $\langle\ ,\ \rangle=(\ ,\Theta\ )$, where Θ is a unitary and Hermitian operator on the Fock space.

Next we prove that both the bare Hamiltonian $H(L,\sigma)$ and the renormalized Hamiltonian $H(L,\sigma)-R_{L,\sigma}(A_\mu)-E(L,\sigma)=H_R(L,\sigma)$ ($R,E=$ counterterms) are symmetric with respect to the indefinite inner product (namely Θ-symmetric). Especially $H_S(L,\sigma)=H(L,\sigma)|_{M=0}$ and $H^R_S(L,\sigma)=H_R(L,\sigma)|_{M=0}$ are self-adjoint with respect to the indefinite inner product (namely Θ-self-adjoint). (See Secs. 4 and 5.)

Even if a Hamiltonian H is Θ-self-adjoint, the spectrum of H is not necessarily real. Further even if H is semibounded and the spectrum is real, the vacuum is sometimes outside the Fock space. But we first assume that $\Omega(L,\sigma)$ is the vacuum of $H_R(L,\sigma)$ (in the Fock space). Then $\rho_{L,\sigma}(\cdots) = \langle\Omega(L,\sigma),\ldots,\Omega(L,\sigma)\rangle$ is a normalized Θ-self-adjoint linear functional on the field algebra, which will be called a Lorentz state in this paper. We discuss whether or not the Lorentz states $\{\rho_{L,\sigma}(\cdots)\}$ converge on the field algebra when we take the limits $L,\sigma\to\infty$. This procedure is explicitly completed at least when $M=0$ (i.e., in the Schwinger model), and we will see that the Lorentz states $\{\rho_{L,\sigma}\}$ uniformly converge as L, $\sigma\to\infty$ (Sec. 6).

For this purpose, we investigate a Bogolyubov transformation which leaves canonical commutation relations with an indefinite metric invariant, and study its implementability by an operator on the Fock space. To obtain a representation from the limiting Lorentz state is one of the most difficult problems in QED, since the Lorentz states are not positive and not continuous on the C^*-algebra generated from $\{\exp i\Phi(f); f\in\mathscr{H}\}$, where Φ is the self-adjoint Segal's field. However, one representation will be explicitly obtained (following Klaiber).

When $M\neq 0$, two approaches will be discussed in Secs. 7 and 8. In Sec. 7, we consider Euclidean QED, while in Sec. 8, we study a transformation of $(\mathrm{QED})_2$ to a sine-Gordon model by *a generalized mass-shift transformation of the fermion*. We will see that $(\mathrm{QED})_2$-type models are isomorphic to two-component sine-Gordon models on the vacuum sector.

Notation:

$$g^{\mu\nu}=g^{\nu\mu}=\mathrm{diag}(1,-1),$$

$$\epsilon^{\mu\nu}=-\epsilon^{\nu\mu},\quad \epsilon^{01}=-\epsilon^{10}=\epsilon_{10}=-\epsilon_{01}=1,$$

$$\gamma^0=\gamma_0=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \gamma^1=-\gamma_1=\begin{pmatrix}0&1\\-1&0\end{pmatrix},$$

$$\gamma^5=\gamma^0\gamma^1=\begin{pmatrix}-1&0\\0&1\end{pmatrix},$$

$$x^\mu=(t,x)=(x^0,x^1)=(x_0,-x_1),$$

$$p^\mu=(p^0,p)=(p^0,p^1)=(p_0,-p_1),$$

$$\partial_\mu=\frac{\partial}{\partial x^\mu},\quad \not{p}=\gamma^\mu p_\mu=\gamma^0p_0+\gamma^1p_1=\gamma^0p^0-\gamma^1p^1,$$

$$\theta(x)=\begin{cases}1 & x\geqslant 0,\\ 0 & \text{otherwise},\end{cases}\qquad \epsilon(x)=\begin{cases}1 & x\geqslant 0,\\ -1 & \text{otherwise}.\end{cases}$$

2. FERMION FOCK SPACE AND PROPERTIES OF CURRENT

A. Two-dimensional fermion field operator

The free fermion field of mass M obeys the Dirac equation

 0022-2488/80/061473-22$01.00

$$(-i\gamma^\mu\partial_\mu + M)\psi(x,t) = 0,$$

and ψ is expanded as follows:

$$\psi_i(x,0) = \int \frac{dp}{(2\pi)^{1/2}} [u_i(p)c(p) + v_i(p)d^*(-p)]e^{ipx} \equiv \int \frac{dp}{(2\pi)^{1/2}} \hat{\psi}_i(p)e^{ipx},$$

where $i = 1$ or 2 and spinors u and v are, respectively, given by

$$u(p) = \begin{pmatrix} \nu(-p) \\ \nu(p) \end{pmatrix}, \quad v(p) = \begin{pmatrix} \nu(p) \\ -\nu(-p) \end{pmatrix}, \tag{2.1}$$

with

$$\nu(p) = \left(\frac{\omega(p)+p}{2\omega(p)}\right)^{1/2}, \quad \omega(p) = (p^2 + M^2)^{1/2}.$$

Then they satisfy the normalization conditions

$$(u(p),u(p)) = (v(p),v(p)) = 1, \qquad (u(p),v(p)) = 0. \tag{2.2}$$

Let $\mathscr{H}_e = \mathscr{H}_p = L^2(R,dx)$ and let $\mathscr{H} = \mathscr{H}_e \oplus \mathscr{H}_p$ (e and p mean electron and positron, respectively). The fermion Fock space is

$$\mathscr{F}_F \equiv \bigoplus_{n=0}^{\infty} \mathscr{F}_F^{(n)}, \quad \mathscr{F}_F^0 = C,$$

$$\mathscr{F}_F^n = AS_n[\otimes_n \mathscr{H}], \tag{2.3}$$

$$AS_n[f_1 \otimes \cdots \otimes f_n] \equiv \frac{1}{(n!)^{1/2}} \sum_\pi \operatorname{sgn}(\pi) f_{\pi(1)} \otimes \cdots \otimes f_{\pi(n)}, \tag{2.4}$$

where the Fock vacuum Ω is in $\mathscr{F}_F^0 = C$. The annihilation $[c(p)$ and $d(p)]$ and creation $\{c^*(p) = [c(p)]^*$ and $d^*(p) = [d(p)]^*\}$ operators obey the canonical anticommutation relations (CAR) $\{c(p),c^*(q)\} = \delta(p-q) = \{d(p),d^*(q)\}$, etc. Since $0 \leqslant \nu(p) \leqslant 1$, $\psi(f) = \int \psi(x,0) f\, dx$ is a bounded operator if $f \in L^2(R,dx)$.

In a periodic box (box $= \Lambda = [-L/2, L/2] \subset R$), $L^2(R,dx)$ is replaced by $L^2(\Lambda,dx)$. Further let $\Gamma = \{2\pi n/L; n = 0, \pm 1, \ldots\}$. By the Fourier transformation $\tilde{f}(p) \equiv \int_\Lambda f(x)\exp(-ipx)\,dx$ with $p \in \Gamma$, we can identify $L^2(\Lambda,dx)$ and $l^2(\Gamma) = \{\{\tilde{f}(p); p \in \Gamma\}; \Sigma|\tilde{f}(p)|^2 < \infty\}$. Let $k \in \Gamma$ and let $p \in [k - \pi/L, k + \pi/L)$. Then $c(p)$ is replaced by the following averaged operator in a periodic box:

$$c_\Lambda(k) = \frac{L}{2\pi} \int_{-\pi/L}^{\pi/L} c(k+l)\,dl. \tag{2.5}$$

We similarly define $c_\Lambda^*(k)$, $d_\Lambda(k)$, and $d_\Lambda^*(k)$. They obey the CAR in a periodic box:

$$\{c_\Lambda(p), c_\Lambda^*(q)\} = \{d_\Lambda(p), d_\Lambda^*(q)\} = (L/2\pi)\delta_{p,q}, \quad \text{etc.} \tag{2.6}$$

The expressions in this approximation are obtained from the usual ones by the following replacements:

$$\delta(p-q) \to \frac{L}{2\pi}\delta_{p,q}, \quad \int dp \to \frac{2\pi}{L}\sum_p,$$

$$\int_R dx \to \int_\Lambda dx.$$

For example, the generalized number operator N_τ changes as

$$N_\tau \to N_{\tau,\Lambda} \equiv \frac{2\pi}{L}\sum_p \omega(p)^\tau [c_\Lambda^*(p)c_\Lambda(p) + d_\Lambda^*(p)d_\Lambda(p)],$$

and especially, $N_{\tau=1,\Lambda}$ is the free Hamiltonian of the fermion in this approximation.

In terms of the Fourier components, the Fock space $\mathscr{F}_F(\mathscr{H})$ is given by

$$\oplus_{n=0}^{\infty} AS_n \otimes^n (D_e \oplus D_p),$$

where $D_e = D_p = l^2(\Gamma)$. Further

$$\psi_{i,\Lambda}(x,0) = \frac{\sqrt{2\pi}}{L}\sum_p [u_i(p)c_\Lambda(p) + v_i(p)d_\Lambda^*(-p)]e^{ipx} \equiv \frac{\sqrt{2\pi}}{L}\sum_p \hat{\psi}_{i,\Lambda}(p)e^{ipx}. \tag{2.7}$$

We use this approximation throughout the paper, and then we will omit the index Λ (or L) for simplicity when there is no danger of confusion.

When $M = 0$, there are some ambiguities in the expressions of the spinors $u(p)$ and $v(p)$. Throughout the paper, we use

$$u(p) = \begin{pmatrix} \theta(-p) \\ \theta(p) \end{pmatrix} \text{ for } p \neq 0, \quad \text{and } u(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$v(p) = \begin{pmatrix} \theta(p) \\ -\theta(-p) \end{pmatrix} \text{ for } p \neq 0, \quad \text{and } v(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{2.8}$$

B. Current operators in two dimensions

Let

$$\psi_\kappa(x,0) = \frac{\sqrt{2\pi}}{L}\sum_{|p|<\kappa} \hat{\psi}(p)e^{ipx},$$

$$W^{(+)\mu}(k;L,\kappa) = \frac{2\pi}{L}\sum_{p,|p|,|p-k|<\kappa} (u(p),\gamma^0\gamma^\mu v(p-k))c^*(p)d^*(k-p),$$

$$W^{(-)\mu}(k;L,\kappa) = [W^{(-)\mu}(-k;L,\kappa)]^*,$$

$$W^\mu(k;L,\kappa) = \frac{2\pi}{L}\sum_{p,|p|,|p-k|<\kappa} [(u(p),\gamma^0\gamma^\mu u(p-k)) \times c^*(p)c(p-k) - (v(p),\gamma^0\gamma^\mu v(p-k)) \times d^*(k-p)d(-p)] = [W^\mu(-k;L,\kappa)]^*. \tag{2.9}$$

Let a current with double momentum cutoffs κ and σ be

$$j^\mu_{\sigma,\kappa}(x,0) = \int_\Lambda h_\sigma(x-y){:}\bar{\psi}_\kappa(y,0)\gamma^\mu\psi_\kappa(y,0){:}\,dy, \tag{2.10}$$

where $\bar{\psi} = \psi^*\gamma^0$. We define

$$W^{(\#)\mu}(k;L) = \lim_{\kappa\to\infty} W^{(\#)\mu}(k;L,\kappa), \tag{2.11}$$

$$j^\mu_\sigma(x,0) = \lim_{\kappa\to\infty} j^\mu_{\sigma,\kappa}(x,0) = \frac{1}{L}\sum_k \tilde{h}_\sigma(k)\tilde{j}^\mu(k)e^{-ikx}, \tag{2.12}$$

where

$$\tilde{j}^\mu(k) = W^{(+)\mu}(k;L) + W^{(-)\mu}(k;L) + W^\mu(k;L).$$

The momentum cutoff functions $\{h_\sigma(x);\sigma\in R\}$ are real even functions such that $(2\pi/L)\Sigma_p|\tilde{h}_\sigma^2(p)|(|p|+1)<\infty$ (see lemma 2.1). Without loss of generality, we can choose a sequence of h_σ such that $\tilde{h}_\sigma(p)=\chi_\sigma(p)$, where $\chi_\sigma(p)$ is the characteristic function in the interval $[-\sigma,\sigma]\subset R$.

Lemma 2.1: (i) When $M=0$, $\{W^{(\pm)\mu}(k;L)\}$ are bounded operators whose norms are dominated by $(L/2\pi)|k|$. Further $\{W^{(\pm)\mu=0}(k;L)\}$ are bounded operators whose norms are dominated by

$$(L/2\pi)(|k|+MO(|k|)). \tag{2.13}$$

(ii) $\{W^{(\pm)\mu}(k;L)\}$ are operators such that

$$|W^{(\pm)\mu}(k;L)|^2\leqslant(L/2\pi)(|k|+c_0)(N^2+1)^2, \tag{2.14}$$

where N is the fermion number operator and c_0 is a constant of the order $M^2O(1)$ and independent of K and L

(iii) $\{W^\mu(k;L)\}$ are well defined (unbounded) operators:

$$|W^\mu(k;L)|^2\leqslant(N_\tau+c_1(\tau))^2,\quad c_1(\tau=0)=0, \tag{2.15}$$

where N_τ is the generalized fermion number operator and constant $c_1(\tau)$ with $\tau\neq0$ may depend on L and k.

Proof: (i) Since $\|c(p)\|=\|d(p)\|=(L/2\pi)^{1/2}$, the boundedness of $\{W^{(\pm)\mu}(k;L)\}_{M=0}$ follows from the property of the spinors:

$$\|W^{(\pm)\mu}(k;L)\|\leqslant\sum_p|(u(p),\gamma^0\gamma^\mu v(p-k))|=(L/2\pi)|k|.$$

Further since

$$(u(p),\gamma^0\gamma^\mu v(p-k))$$
$$=\pm\nu(-p)\nu(p-k)-\nu(p)\nu(k-p)\quad(\mu=0,1)$$
$$=\begin{cases}\dfrac{M}{2}\left(\pm\dfrac{1}{|p|}-\dfrac{1}{|k-p|}\right)+O(p^{-3})\ \text{as}\ p\to+\infty,\\[2ex] \dfrac{M}{2}\left(\pm\dfrac{1}{|p-k|}-\dfrac{1}{|p|}\right)+O(|p|^{-3})\ \text{as}\ p\to-\infty,\end{cases}$$

the same method shows the boundedness of $W^{(\pm)\mu=0}(k;L)$:

$$\sum_p|\nu(-p)\nu(p-k)-\nu(p)\nu(k-p)|$$
$$\leqslant(L/2\pi)[|k|+MO(|k|)].$$

(ii) Let

$$W^{(+)}=W^{(+)\mu}(k;L)=(2\pi/L)\sum_p f(p)c^*(p)d^*(k-p),$$

$$f(p)=\pm\nu(-p)\nu(p-k)-\nu(p)\nu(k-p).$$

We shall prove an inequality

$$\|W^{(+)}\omega\|^2\leqslant(\Sigma_p f^2(p))\|(N^2+1)\omega\|^2,$$

where $\omega\in\mathscr{D}_F$ ($=$ finite particle set). Then it suffices to assume that ω contains only m electrons (c-fermions) and n positrons (d-fermions). We set $L=2\pi$ for brevity.

Defining $P=(p_1,\dots,p_m)$ and $Q=(q_1,\dots,q_n)$ with

$p_1<\dots<p_m$, $q_1<\dots<q_n$, let

$$\omega=\sum_{p\in Z^m,Q\in Z^n}a(P;Q)\prod_{i=1}^m c^*(p_i)\prod_{j=1}^n d^*(q_j)\Omega,$$

where

$$\prod_{i=1}^m c^*(p_i)=c^*(p_1)\cdots c^*(p_m),\ \text{etc.}$$

Then

$$\|\omega\|^2=\sum_{p\in Z^m,Q\in Z^n}|a(P;Q)|^2$$

and

$$W^{(+)}\omega=\sum_{(p,P)\in Z^{m+1},Q\in Z^n}(-1)^m\prod_{i=0}^m c^*(p_i)\prod_{j=0}^n d^*(q_i)\Omega,$$

where $p_0=p$ and $q_0=p-k$. Thus we get

$$\|W^{(+)}\omega\|^2$$
$$=\sum_{\substack{(p,P)\in Z^{m+1},(p',P')\in Z^{m+1}\\ Q\in Z^n,Q'\in Z^n}}f(p)\bar{f}(p')a(P;Q)\bar{a}(P';Q')$$
$$\times\left\langle\Omega,\left[\prod_{i=0}^m c^*(p_i')\right]^*\left[\prod_{j=0}^m c^*(p_j)\right]\Omega\right\rangle$$
$$\times\left\langle\Omega,\left[\prod_{i=0}^n d^*(q_i')\right]^*\left[\prod_{j=0}^n d^*(q_j)\right]\Omega\right\rangle.$$

Since $p_1<\dots<p_m$, $p_1'<\dots<p_m'$ (and similar inequalities for q's and q''s), the Wick contractions for the vacuum expactation values of the c-fermion operators in the above equation yields the following m^2+1 contraction terms (except their signs):

$$\delta_{p,p'}\delta_{P,P'},\quad\delta_{p,p_k'}\delta_{p',p_l}\delta_{P\setminus p_l,P'\setminus p_k'},$$

where $\delta_{P,P'}=\delta_{p_1,p_1'}\times..\times\delta_{p_m,p_m'}$ and $P\setminus p_l=(p_1,\dots,p_{l-1},p_{l+1},\dots,p_m)$. Similar n^2+1 terms arise from the d-fermion operators. Thus $(m^2+1)(n^2+1)\leqslant((m+n)^2+1)^2$ terms arise from the contraction.

Now we take the sum over p,p',P,P',Q, and Q'. It is easy to see that each of the above terms is dominated by

$$\left(\sum_p f^2(p)\right)\left(\sum_{P,Q}|a(P;Q)|^2\right).$$

This is proved by Hölder inequality and by a trivial inequality $|\Sigma_i a_i b_i|\leqslant(\Sigma_i|a_i|)(\Sigma_j|b_j|)$. Thus we finally get

$$\|W^{(+)}\omega\|^2\leqslant\left(\sum_p f^2(p)\right)\|(N^2+1)\omega\|^2.$$

It is easily confirmed that this inequality does not depend on L (except the definition of Γ). Finally

$$\sum_p f^2(p)$$
$$=\sum_{p\in\Gamma}\tfrac{1}{2}\left(1-\frac{p(p-k)}{\omega\omega'}\mp\frac{M^2}{\omega\omega'}\right)\leqslant\frac{L}{2\pi}(|k|+c_0),$$

where $\omega=\omega(p),\omega'=\omega(p-k)$, and c_0 is a constant which is independent of L and k, and is of the order $M^2O(1)$.

(iii) First note that $\{W^\mu(k;L)\}$ do not mix electrons and positrons. Then when $\tau=0$, it suffices to prove the following inequality:

$$\|W(k)\omega\|\leqslant\|N'\omega\|,$$

with $\omega\in\mathscr{D}_F$, or equivalently the following inequality:

$$|W(k)|^2\leqslant(N')^2,$$

where

$$W(k)=\frac{2\pi}{L}\sum_p (u(p),\gamma^0\gamma^\mu u(p-k))c^*(p)c(p-k),$$

$$N'=\frac{2\pi}{L}\sum_p c^*(p)c(p).$$

Let

$$\omega=\sum_{n=0}^{\infty}\omega_n,\quad N'\omega_n=n\omega_n,\quad \phi=\sum_{n=0}^{\infty}\phi_n,\quad N'\phi_n=n\phi_n.$$

Then for $\omega\in D(N')$, using Hölder inequality and $|(u,\gamma^0\gamma^\mu u)|\leqslant 1$,

$$\|W(k)\omega\|=\sup_{\|\phi\|\leqslant 1}|(\phi,W(k)\omega)|=\sup_{\|\phi\|\leqslant 1}\sum_n|(\phi_n,W(k)\omega_n)|$$

$$\leqslant \sup_{\|\phi\|\leqslant 1}\sum_n\Big(\frac{2\pi}{L}\sum_p\|c(p)\phi_n\|^2\Big)^{1/2}\times\Big(\frac{2\pi}{L}\sum_p\|c(p)\omega_n\|^2\Big)^{1/2}$$

$$=\sup_{\|\phi\|\leqslant 1}\sum_n n\|\phi_n\|\|\omega_n\|$$

$$\leqslant \sup_{\|\phi\|\leqslant 1}\Big(\sum_n\|\phi_n\|^2\Big)^{1/2}\Big(\sum_m\|N'\omega_m\|^2\Big)^{1/2}$$

$$=\|N'\omega\|.$$

Finally, $c_1(\tau)$ is the minimal constant such that $N=N_{\tau=0}\leqslant N_\tau+c_1(\tau)$. Then $c_1(\tau)$ depends on L and M. □

From this lemma, we see that the current $j^\mu(f,0)=\int j^\mu(x,0)f(x)\,dx$ is a densely defined symmetric operator if $f(x)\in R$ and $(2\pi/L)\Sigma_p(|p|+1)|\tilde f(p)^2|<\infty$. [To check this, mimic the proof of Lemma 2.1 (ii).]

C. Associated bosons

The two-dimensional currents constructed from the massless free fermion can be written as derivatives of neutral scalar (or pseudoscalar) boson J (or $\tilde J$) called an associated boson (or an associated pseudoscalar boson):

$$j^\mu(x,t)=\partial^\mu J(x,t)/\sqrt{\pi}=-\epsilon^{\mu\nu}\partial_\nu\tilde J(x,t)/\sqrt{\pi}. \tag{2.16}$$

In fact following Uhlenbrock and Kaliber,[8-10] Let

$$A(p)=\frac{2\pi}{L}\frac{1}{\sqrt{|p|}}\Big[\theta(p)\sum_q:\hat\psi_1^*(q)\hat\psi_1(p+q): +\theta(-p)\sum_q:\hat\psi_2^*(q)\hat\psi_2(p+q):\Big]. \tag{2.17}$$

Then we define the associated bosons with sharp momentum cutoff σ by

$$J_\sigma(x,t)=-i\frac{\sqrt{2\pi}}{L}\sum_{\sigma>|p|>0}\frac{1}{\sqrt{2|p|}}[-A(p)e^{-ipx}+A(p)^*e^{ipx}]+\frac{\sqrt{\pi}}{L}[t(\Lambda_1+\Lambda_2)+x(\Lambda_1-\Lambda_2)],$$

$$\tilde J_\sigma(x,t)=-i\frac{\sqrt{2\pi}}{L}\sum_{\sigma>|p|>0}\frac{\epsilon(p)}{\sqrt{2|p|}}[-A(p)e^{-ipx}+A(p)^*e^{ipx}]-\frac{\sqrt{\pi}}{L}[t(\Lambda_1-\Lambda_2)+x(\Lambda_1+\Lambda_2)], \tag{2.18}$$

where $px=|p|t-px$, and the operators Λ_1 and Λ_2 are charges

$$\Lambda_1=\frac{2\pi}{L}\Big[\sum_{p<0}c^*(p)c(p)-\sum_{p<0}d^*(p)d(p)\Big],$$
$$\Lambda_2=\frac{2\pi}{L}\Big[\sum_{p>0}c^*(p)c(p)-\sum_{p>0}d^*(p)d(p)\Big]. \tag{2.19}$$

Lemma 2.1 shows that $\{A^\#(p);p\in\Gamma\setminus\{0\}\}$ are well-defined operators on $D(N)$. We can check[9,10]

$$[A(p),A(q)]\Phi=[A^+(p),A^+(q)]\Phi=0,$$
$$[A(p),A^+(q)]\Phi=(L/2\pi)\delta_{p,q}\Phi, \tag{2.20}$$
$$(\Phi,A(p)\Psi)=(A^+(p)\Phi,\Psi),$$

where $\Phi,\Psi\in\mathscr{D}_F$ (finite particle set in $\mathscr{F}_F$) and $A^+(p)=A(p)^*\upharpoonright\mathscr{D}_F$. Moreover

$$[\Lambda_1,\Lambda_2]=0,\quad [\Lambda_i,A^\#(p)]=0,\quad [H_{0:L},\Lambda_i]=0,$$
$$[\Lambda_i,c]=u_i(p)c^*(p),\quad [\Lambda_i,d^*(p)]=-|v_i(p)|d^*(p),$$

where $H_{0:L}=N_{\tau=1,\Lambda}$ is the free Hamiltonian of the fermion in the periodic box. Let

$$\mathscr{F}_F=\oplus_{\lambda=(\lambda_1,\lambda_2)}\mathscr{F}_F(\lambda),$$
$$\mathscr{F}_F(\lambda)=\{x\in\mathscr{F}_F;\Lambda_i x=\lambda_i x,\ i=1,2\}.$$

Let $\mathscr{D}_F(\lambda)=\mathscr{D}_F\cap\mathscr{F}_F(\lambda)$. The sector vacuum $\Omega(\lambda)$ is the vector in $\mathscr{D}_F(\lambda)$ that has the lowest energy $e(\lambda)$, and given by $U(\lambda)\Omega$ except the normalization where Ω is the Fock vacuum and

$$U(\lambda)=\Big[\delta_{\lambda_1,0}+\theta(\lambda_1)\prod_{\mu=1}^{\lambda_1}c^*\Big(-\frac{2\pi\mu}{L}\Big)+\theta(-\lambda_1)\prod_{\nu=0}^{|\lambda_1+1|}d^*\Big(-\frac{2\pi\nu}{L}\Big)\Big]\times\Big[\delta_{\lambda_2,0}+\theta(\lambda_2)\prod_{\mu'=1}^{\lambda_2-1}c^*\Big(\frac{2\pi\mu'}{L}\Big)+\theta(-\lambda_2)\prod_{\nu'=1}^{|\lambda_2|}d^*\Big(\frac{2\pi\nu'}{L}\Big)\Big].$$

Then

$$e(\lambda)=(\pi/L)(\lambda_1^2+\lambda_1+\lambda_2^2-\lambda_2)$$

and

$$A(k)\Omega(\lambda)=0\ \text{ for any }\ k\in\Gamma\setminus\{0\}.$$

Obviously

$$\mathscr{D}_J(\lambda)\subset\mathscr{D}_F(\lambda),$$

where $\mathscr{D}_J(\lambda)$ is the finite particle set of the boson obtained by operating $\{A^*(k);k\in\Gamma\setminus\{0\}\}$ on $\Omega(\lambda)$ cyclically. Uhlenbrock showed $\mathscr{D}_J(\lambda)=\mathscr{D}_F(\lambda)$. Then as the Fock spaces,

$$\mathscr{F}_J(\lambda)=\mathscr{F}_F(\lambda),$$

where $\mathscr{F}_J(\lambda)$ is the Fock space of the associated boson belonging to the charges (λ_1,λ_2): $\mathscr{F}_J(\lambda)=\overline{\mathscr{D}_J(\lambda)}$.

The free Hamiltonian $H_{0:L} = (2\pi/L)\Sigma|p|[c^*(p)c(p)+d^*(p)d(p)]$ is self-adjoint and nonnegative. Further

$$[H_{0:L},A^+(p)]\Phi=|p|A^+(p)\Phi$$

for any $\Phi\in\mathscr{D}_F$. Since $[H_{0:L},A_i]=[A^\#(k),A_i]=0$, we can define

$$H_{0:L}(\lambda)\equiv H_{0:L}\upharpoonright\mathscr{D}_F(\lambda),$$
$$A^\#(p;\lambda)\equiv A^\#(p)\upharpoonright\mathscr{D}_F(\lambda),$$

where $H_{0:L}(\lambda)$ is called a sector Hamiltonian. Thus we get

$$\left[H_{0:L}(\lambda)-\frac{2\pi}{L}\sum_p|p|A^+(p;\lambda)A(p;\lambda),A^\#(k;\lambda)\right]\Phi=0$$

for any $\Phi\in\mathscr{D}_F(\lambda)$. By the irreducibility of $\{A^\#(k;\lambda);k\in\Gamma\setminus\{0\}\}$, we obtain the Kronig's identity

$$H_{0:L}(\lambda)=\frac{2\pi}{L}\sum_p|p|A^+(p;\lambda)A(p;\lambda)+1_\lambda e(\lambda),$$
$$1_\lambda\equiv1\upharpoonright\mathscr{F}_F(\lambda). \tag{2.21}$$

3. VECTOR FIELD WITH AN INDEFINITE METRIC

A. The Proca field

Let $f^\mu(x)=f^\mu(x^1,...,x^s)\in\mathscr{S}(R^s)$ and let $\{f\}=\{f^0,f^1,...,f^s\}$. We define an inner product $(\ ,\ )_P$ (P means Proca):

$$(f,g)_P=\int\frac{d^sp}{(2\pi)^s2p^0}\tilde f^{\mu*}(p)\left(-g_{\mu\nu}+\frac{p_\mu p_\nu}{\mu^2}\right)\tilde f^\nu(p), \tag{3.1}$$

where μ is a (positive) constant, $p^0=(\mu^2+\Sigma_{i=1}^s p_i^2)^{1/2}$ and $p^\mu=(p^0,p^1,...,p^s)$. Since spec $(-g_{\mu\nu}+p_\mu p_\nu/\mu^2) = \{0,1,...,1,1+2p^2/\mu^2\}$, the inner product is not positive definite. Let $\mathscr{N}$ be the null space of the inner product and let $\mathscr{H}$ be a set of the above mentioned functions with the inner product $(\ ,\ )_P$. Thus we define $\mathscr{H}'\equiv\overline{\mathscr{H}/\mathscr{N}}$.

The Fock space of the Proca field of mass μ is then

$$\mathscr{F}_U=\oplus_{n=0}^\infty\mathscr{F}_U^{(n)},\quad \mathscr{F}_U^{(n)}=S_n[\otimes_n(\mathscr{H}')], \tag{3.2}$$

$$S_n[\{f_1\}\otimes\cdots\otimes\{f_n\}]=\frac{1}{n!}\sum_\pi\{f_{\pi(1)}\}\otimes\cdots\otimes\{f_{\pi(n)}\}. \tag{3.3}$$

For given $\{f\}\in\mathscr{H}'$, define the creation operator $U^*(f)$ by

$$U^*(f)\phi=(n+1)^{1/2}S_{n+1}[\{f\}\otimes\phi],\quad \phi\in\mathscr{F}_U^{(n)}.$$

The annihilation operator $U(f)$ is given by $[U^*(CF)]^*$, where C denotes the complex conjugation operator. Thus

$$[U(f),U(g)]=0,\quad [U(f),U^*(g)]=(Cf,g)_P.$$

The Proca field $U_\mu(x,0)$ has the following expression:

$$U_\mu(x,0)=\int\frac{d^sp}{[(2\pi)^s2p^0]^{1/2}}[\alpha_\mu(p)+\alpha_\mu^*(-p)]e^{ipx}. \tag{3.4}$$

The CCR read

$$[\alpha_\mu(p),\alpha_\nu(q)]=0,$$
$$[\alpha_\mu(p),\alpha_\nu^*(q)]=\delta(p-q)[-g_{\mu\nu}+p_\mu p_\nu/\mu^2] \tag{3.5}$$

with $p^\mu\alpha_\mu(p)=0$ and $\alpha_\mu(p)\Omega=0$. The generalized number operator is given by $(N_{\tau=1}=H_0(U_\mu))$

$$N_\tau=\int d^spp_0^\tau(p)[-\alpha_\mu^*(p)\alpha^\mu(p)]. \tag{3.6}$$

Especially in two dimensions,

$$\alpha_\mu(p)=\epsilon_{\mu\nu}p^\nu a(p)/\mu,$$

where $a(p)=\epsilon_{\mu\nu}p^\mu\alpha_\nu(p)/\mu$ obeys the CCR

$$[a(p),a(q)]=0,\quad [a(p),a^*(q)]=\delta(p-q). \tag{3.7}$$

Then

$$F_{\mu\nu}\equiv\partial_\mu U_\nu-\partial_\nu U_\mu=-\mu\epsilon_{\mu\nu}\phi,$$

$$\phi=\int\frac{dp}{(2\pi2p^0)^{1/2}}[-ia(p)e^{-ipx}+a^*(p)e^{ipx}]. \tag{3.8}$$

B. Fock space with an indefinite metric

Let $\mathscr{H}_{1/2}$ be a space with an inner product

$$(f,g)_{1/2}=\int\frac{d^sp}{(2\pi)^s2p_0'}f^*(p)g(p), \tag{3.9}$$

where $p_0'=(p_2+\mu'^2)^{1/2}$. The boson Fock space $\mathscr{F}_B$, the creation and annihilation operators $B^*(f)$, $B(f)$ are similarly defined.

For an operator A on $\mathscr{H}_{1/2}$, define $\Gamma(A)$ by

$$\Gamma(A)\mathscr{F}_B^{(n)}\subset\mathscr{F}_B^{(n)},\quad \Gamma(A)|\mathscr{F}_B^{(n)}=A\otimes\cdots\otimes A\ (n\ \text{times}). \tag{3.10}$$

For $\varphi=-1$, $\Gamma(\varphi)$ is unitary and Hermitian. We define an indefinite inner product $\langle\ ,\ \rangle=(\ ,\Theta)$ on $\mathscr{F}_B$. Let (Θ) be the adjoint with respect to $\langle\ ,\ \rangle$:

$$A^{(\Theta)}=\Theta A^*\Theta. \tag{3.11}$$

In the following, when there is no danger of confusion with $[B^*(Cf)]^*$, the annihilation operator $B(f)$ is defined by

$$[B^*(Cf)]^{(\Theta)}=\Theta[B^*(Cf)]^*\Theta=[B^*(\varphi Cf)]^*$$
$$=-[B^*(Cf)]^*. \tag{3.12}$$

Then the CCR with the indefinite metric read

$$[B(f),B^*(g)]=(Cf,\varphi g)_{1/2}=-(Cf,g)_{1/2},\ \text{etc.}$$

The gaugeon field B is now represented as

$$B(x,0)=\int\frac{d^sp}{[(2\pi)^s2p_0']^{1/2}}[b(p)+b^*(-p)]e^{ipx}. \tag{3.13}$$

The operator $N_\tau(B)=\int d^sp[-(p_0')^\tau b^*(p)b(p)]$ is nonnegative, self-adjoint and commutes with Θ. $H_0(B) = N_1(B)$ is the free Hamiltonian. In the following, B means μB.[11,12] Let

$$A_\mu\equiv U_\mu+(1/\mu^2)\partial_\mu B. \tag{3.14}$$

The Fock space $\mathscr{F}_A=\mathscr{F}_U\otimes\mathscr{F}_B$ has a indefinite inner product $\langle\ ,\ \rangle=(\ ,\Theta)$:

$$\Theta=(-1)^{N(B)},\quad N(B)=N_{\tau=0}(B). \tag{3.15}$$

The Hamiltonian $H_0(A)=H_0(U)+H_0(B)$ is self-adjoint as well as being Θ-self-adjoint, and spec $H_0(A)\geqslant0$. The time evolution is implemented by a Θ-unitary operator

$\exp[itH_0(A)]$ which is also unitary. The Feynman propagator of A_μ [3,11-13] is that of Stückelberg formalism: Especially of Landau gauge formalism for $\mu' = 0$ and of Feynman gauge formalism for $\mu' = \mu$.

4. INTERACTING HAMILTONIANS

A. Interacting Hamiltoniams with cutoffs

The free Hamiltonian $H_{0:L} = H_{0:L}(\psi) + H_{0:L}(U) + H_{0:L}(B)$ is not only self-adjoint but also Θ-self-adjoint (namely $H_{0:L}\Theta$ is self-adjoint) since $H_{0:L}$ commutes with Θ. Now let

$$H_I(L,\sigma,\kappa) = e\int_\Lambda j^\mu_{\sigma,\kappa}(x,0)A_\mu(x,0)\,dx$$
$$= V^c(L,\sigma,\kappa) + V^a(L,\sigma,\kappa) + V(L,\sigma,\kappa) \quad (4.1)$$

be an interacting Hamiltonian with double momentum cutoffs σ and κ, where

$$V^c(L,\sigma,\kappa) = e\frac{\sqrt{2\pi}}{L}\sum_k \tilde{h}_\sigma(k)W^{(+)\mu}(k;L,\kappa)\tilde{A}_\mu(-k),$$
$$V^a(L,\sigma,\kappa) = [V^c(L,\sigma,\kappa)]^{(\Theta)},$$
$$V(L,\sigma,\kappa) = e\frac{\sqrt{2\pi}}{L}\sum_k \tilde{h}_\sigma(k)W^\mu(k;L,\kappa)\tilde{A}_\mu(-k)$$
$$= [V(L,\sigma,\kappa)]^{(\Theta)}.$$

Here

$$A_\mu(x,0) = \frac{\sqrt{2\pi}}{L}\sum_k \tilde{A}_\mu(-k)e^{ikx}. \quad (4.2)$$

In the following we set $\tilde{h}_\sigma(p) = \chi_\sigma(p)$ for brevity. Let

$$V'^c(L,\sigma,\kappa) = e\frac{\sqrt{2\pi}}{L}\sum_{|k|\leqslant\sigma} W^{(+)\mu}(k;L,\kappa)\tilde{U}_\mu(-k),$$
$$V''^c(L,\sigma,\kappa) = e\frac{\sqrt{2\pi}}{L}\sum_{|k|\leqslant\sigma} W^{(+)\mu}(k,L,\kappa)[\mu^{-2}\tilde{B}_\mu(-k)], \quad (4.3)$$

$V^{\#a}(L,\sigma,\kappa) = [V^{\#c}(L,\sigma,\kappa)]^{(\Theta)}$, $(\# = {}'$ or $'')$.

$V'(L,\sigma,\kappa)$ and $V''(L,\sigma,\kappa)$ are similarly defined. Further let

$V'(L,\sigma,\kappa)$ [resp. $V''(L,\sigma,\kappa)$]

$$= \sum_{i=1}^{2} V'_i(L,\sigma,\kappa) \quad [\text{resp.}\ V''_i(L,\sigma,\kappa)],$$

where V'_1 (resp. V'_2) is a collection of terms in V' which include $\{a(k)\}$ (resp. $\{a^*(k)\}$). For example,

$$V'_1(L,\sigma,\kappa) = \frac{e}{\mu}\frac{\sqrt{2\pi}}{L}\sum_{|k|\leqslant\sigma}\frac{\epsilon_{\mu\nu}k^\nu W^\mu(k;L,\kappa)a(k)}{\sqrt{2p^0}},$$
$$k^\mu = (p^0(k),k).$$

Other operators, for example $V''_{1,2}$, $V'^{a,c}_{1,2}$, are similarly defined.

Since (Glimm and Jaffe [14]), for $\psi\in\mathscr{D}$,

$$\frac{2\pi}{L}p^0\|a(k)\psi\|^2 \leqslant (\psi,H_{0:L}\psi).$$

Lemma 4.1: Let $E \geqslant p^0(k)$. Then

$$\left(\frac{2\pi}{L}p^0(k)\right)^{1/2}\|a^\#(k)(H_{0:L}+E)^{-1/2}\| \leqslant 1.$$

Further since $(\epsilon H_{0:L}+E)^{-1} \leqslant (H_{0:L}+E/\epsilon)^{-1/2}(\epsilon E)^{-1/2}$ $(E/\epsilon \geqslant 1)$,

Lemma 4.1': If $E/\epsilon \geqslant \max\{p^0(k),1\}$, then

$$\left(\frac{2\pi}{L}p^0(k)\right)^{1/2}\|a^\#(k)(\epsilon H_{0:L}+E)^{-1}\| \leqslant (\epsilon E)^{-1/2}.$$

Similar inequalities hold for $\{b^\#(k)\}$. Let $E/\epsilon \geqslant \max\{p^0(k),p'^0(k),1\}$. Then by Lemma 4.1' and by the boundedness of $\{W^\mu(k;L,\kappa)\}$,

$$\|V'_1(L,\sigma,\kappa)(\epsilon H_{0:L}+E)^{-1}\| \leqslant (\epsilon E)^{-1/2}C'_1,$$

where C'_1 is a constant which may diverge as $L,\kappa\to\infty$. Similar inequalities hold for other operators V'_2, V''_i, etc. Let K be the sum of these constants. Then we have

Lemma 4.2: When $L,\kappa<\infty$, there are constants ϵ, E, and K such that

$$\|H_I(L,\sigma,\kappa)(\epsilon H_{0:L}+E)^{-1}\| \leqslant K/\sqrt{\epsilon E},$$

where K is independent of ϵ and E and diverges as $L,\kappa\to\infty$.

Lemma 4.2': When $L,\kappa<\infty$, $H_I(L,\sigma,\kappa)$ is infinitesimally small with respect to $H_{0:L}$ in the sense of Kato[15]: Namely for any $\psi\in D(H_{0:L})$ there are constants ϵ and E such that

$$\|H_I(L,\sigma,\kappa)\psi\| \leqslant \epsilon\|H_{0:L}\psi\| + E\|\psi\|, \quad \text{with } 0\leqslant\epsilon<1.$$

Proof: By Lemma 4.2,

$$\|H_I(L,\sigma,\kappa)\psi\|$$
$$\leqslant K(\epsilon E)^{-1/2}\|(\epsilon H_{0:L}+E)\psi\|$$
$$\leqslant K(\epsilon/E)^{1/2}\|H_{0:L}\psi\| + K(E/\epsilon)^{1/2}\|\psi\|.$$

Set $K(\epsilon/E)^{1/2} \equiv \epsilon$ and $K(E/\epsilon)^{1/2} \equiv E$ to complete the proof. □

Theorem 4.3: $H(L,\sigma,\kappa)$ is Θ-self-adjoint on $D(H_{0:L})$, and semibounded in the following sense:

$$\inf \text{Real spec}\, H(L,\sigma,\kappa) \geqslant -\max\{E/(1-\epsilon),E\}$$
$$= -E/(1-\epsilon).$$

Proof: Following Kato,[15] consider the resolvent

$$(H(L,\sigma,\kappa)-\zeta)^{-1}$$
$$= (H_0-\zeta)^{-1}[1+H_I(H_0-\zeta)^{-1}]^{-1}$$
$$= R(\zeta)\sum_{n=0}^{\infty}(-H_IR(\zeta))^n,$$

where we set $H_0 = H_{0:L}$, $H_I = H_I(L,\sigma,\kappa)$, and $R(\zeta) = (H_0-\zeta)^{-1}$. Since spec $H_0 \geqslant 0$, using Lemma 4.2' we get

$$\|H_IR(\zeta)\| \leqslant E\|R(\zeta)\| + \epsilon\|H_0R(\zeta)\|$$
$$\leqslant E\sup_{\lambda>0}|\lambda-\zeta|^{-1} + \epsilon\sup_{\lambda>0}|\lambda||\lambda-\zeta|^{-1}.$$

Since the Neuman series absolutely converges for ζ such that $\|H_IR(\zeta)\|<1$, the set of such $\zeta\in C$ is in the resolvent set of H_0+H_I. By a simple calculation, [15]

$$\inf \text{Real part of spec}\,(H_0+H_I)$$
$$\geqslant -\max\{E/(1-\epsilon),E\}.$$

Finally, since $\Theta = \Gamma(\varphi)$ is a unitary and Hermitian operator,

$$\|\Theta H_I\psi\| \leqslant E\,\|\psi\| + \epsilon\|\Theta H_0\psi\|. \quad (4.4)$$

This means that the symmetric operator ΘH_I is infinitesimally small with respect to the self-adjoint operator $H_0\Theta = \Theta H_0$ in the sense of Kato. Then the symmetric operator $\Theta[H_0 + H_I]$ is self-adjoint on $D(H_0)$, which means that $H_0 + H_I$ is Θ-self-adjoint. □

B. Total Hamiltonian

The investigation in Sec. 2 shows that the cutoff interacting Hamiltonian $H_I(L,\sigma) = \lim_{\kappa\to\infty} H_I(L,\sigma,\kappa)$ is a densely defined symmetric operator. Let (m_1,m_2,m_3) be the number of the Proca boson, c-fermion and d-fermion, respectively. Let $\mathscr{F}(m_1,m_2,m_3)\subset\mathscr{F}_U\otimes\mathscr{F}_F$ be a linear subspace which includes m_1 bosons, m_2 c-fermions, and m_3 d-fermions. Further let $V_i'^{\#}(L,\sigma) = \lim_{\kappa\to\infty} V_i'^{\#}(L,\sigma,\kappa)$. These are bounded when restricted on $\mathscr{F}(m_1,m_2,m_3)$:

$$\|V_i'(L,\sigma)\| \leqslant C_1(m_1+1)^{1/2}(m_2+m_3),$$

$$\|V_i'^{c,a}(L,\sigma)\| \leqslant C_2(m_1+1)^{1/2}((m_2+m_3)^2+1)$$

on $\mathscr{F}(m_1,m_2,m_3)$, where $C_1 = 0(\sigma^2)$ and $C_2 = 0(L\sigma^3)$.

First consider a Hamiltonian $H_{0:L} + V'^a(L,\sigma) + V'^c(L,\sigma)$. Since $\{W^{(\pm)\mu}(k;L)\}$ may not be bounded if $M\neq 0$, $V'^a + V'^c$ may not be infinitesimally small with respect to $H_{0:L}$. Next consider $H_{0:L} + V'(L,\sigma)$:

Lemma 4.4: $H_{0:L} + V'(L,\sigma)$ is essentially self-adjoint.

Proof: Note that $V'(L,\sigma)$ and $H_{0:L}$ do not change the number of th c- and d-fermions. Then it suffices to show that $H_{0:L} + V'(L,\sigma)$ is self-adjoint when restricted on $\mathscr{F}_U\otimes\mathscr{F}_F(m_2,m_3$. This follows from the infinitesimal smallness of $V'(L,\sigma)$ when restricted on $\mathscr{F}_U\otimes\mathscr{F}_F(m_2,m_3)$ □

Let $H^{(\Theta)}(L,\sigma)$ be the adjoint of $H(L,\sigma) = H_{0:L} + H_I(L,\sigma)$ with respect to $\langle\ ,\ \rangle$. Then $H^{(\Theta)}(L,\sigma)\supset H(L,\sigma)$ or equivalently $[H(L,\sigma)\Theta]^*\supset H(L,\sigma)\Theta$ since Θ is unitary and Hermitian. Thus to prove the Θ-self-adjointness of H, it suffices to prove the self-adjointness of $H\Theta$. By mimicking the proof of Lemma 4.4:

Lemma 4. 5: The symmetric operator

$$K = [H_{0:L} + V(L,\sigma)]\Theta$$

$$= [H_{0:L} + V'(L,\sigma) + V''(L,\sigma)]\Theta$$

is essentially self-adjoint.

As will be shown in Sec. 5, the renormalization counterterm is $R_{L,\sigma}(A_\mu) = -(e^2/2\pi)\int :A^2_{1,\sigma}(x,0): dx$, which is Θ-self-adjoint. Then $H(L,\sigma) - R_{L,\sigma}(A_\mu)$ ($= H_R(L,\sigma)$ except for $E(L,\sigma)$) is again Θ-symmetric and densely defined. We conjecture that both $H(L,\sigma)$ and $H_R(L,\sigma)$ are Θ-(essentially) self-adjoint. (For this problem, see.[14,16]) This will be discussed elsewhere

When the fermion mass M vanishes (the Schwinger model), the analysis of the Hamiltonian is very easy. Let

$$\mathscr{F} = \oplus_\lambda\mathscr{F}(\lambda),\quad \mathscr{F}(\lambda) = \mathscr{F}_U\otimes\mathscr{F}_B\otimes\mathscr{F}_F(\lambda). \quad (4.5)$$

Let $\mathscr{D}_U$, $\mathscr{D}_B$, and $\mathscr{D}_F$ be finite particle sets in $\mathscr{F}_U$, $\mathscr{F}_B$, and $\mathscr{F}_F$, respectively. Define

$$\mathscr{D} = \mathscr{D}_U\otimes\mathscr{D}_F\otimes\mathscr{D}_B,\quad \mathscr{D}(\lambda) = \mathscr{D}\cap\mathscr{F}(\lambda). \quad (4.6)$$

Let $H_S(L,\sigma) = H(L,\sigma)|_{M=0}$. Then the Kronig identity reads

$$H_S(L,\sigma) = \oplus_\lambda H_S(L,\sigma;\lambda),\quad H_S(L,\sigma;\lambda) = H_S(L,\sigma)|\mathscr{D}(\lambda). \quad (4.7)$$

Denoting the restrictions $a^*(p)|\mathscr{D}(\lambda)$, $A^*(p)|\mathscr{D}(\lambda)$, and $b^*(p)|\mathscr{D}(\lambda)$ as $a^+(p;\lambda)$, $A^+(p;\lambda)$, and $b^+(p;\lambda)$ (similar for the annihilation operators), we get

$$H_S(L,\sigma;\lambda) = \frac{2\pi}{L}\sum_{p>0} :\mathscr{A}^+(p;\lambda)\mathscr{H}(p;\sigma)\mathscr{A}(p;\lambda): + 1_\lambda e(\lambda) + \frac{\sqrt{\pi}}{L}\frac{m}{\mu}\tilde{h}_\sigma(0)(\lambda_1-\lambda_2)(a^+(0;\lambda) + a(0;\lambda)). \quad (4.8)$$

Here : : is the Wick ordering,

$\mathscr{A}^+(p;\lambda) = \{a^+(p;\lambda), A^+(p;\lambda), b^+(p;\lambda), a(-p;\lambda), A(-p;\lambda), b(-p;\lambda)\}$, $\mathscr{A}(p;\lambda) = [\mathscr{A}^+(p;\lambda)]^{(\Theta)}$, and $\mathscr{H}(p;\sigma)$ is a 6×6 Hermitian matrix:

$$\begin{pmatrix} \mathscr{H}_{11}(p;\sigma) & \mathscr{H}_{12}(p;\sigma) \\ \bar{\mathscr{H}}_{12}(-p;\sigma) & \bar{\mathscr{H}}_{11}(-p;\sigma) \end{pmatrix}, \quad (4.9)$$

where

$$\mathscr{H}_{11}(p;\sigma) = \begin{pmatrix} p_0 & \frac{m}{\mu}\tilde{h}_\sigma(p)\omega'(p) & 0 \\ \frac{m}{\mu}\tilde{h}_\sigma(p)\omega'(p) & |p| & 0 \\ 0 & 0 & -|p| \end{pmatrix},$$

$$\mathscr{H}_{12}(p,\sigma) = \begin{pmatrix} 0 & \frac{m}{\mu}\tilde{h}_\sigma(p)\omega(p) & 0 \\ \frac{m}{\mu}\tilde{h}_\sigma(p)\omega(p) & 0 & i\frac{m}{\mu}\tilde{h}_\sigma(p)|p| \\ 0 & i\frac{m}{\mu}\tilde{h}_\sigma(p)|p| & 0 \end{pmatrix}.$$

In addition

$$m = \frac{e}{\sqrt{\pi}},\quad \omega(p) = \frac{(p^0+|p|)p}{2\sqrt{p^0|p|}},$$

$$\omega'(p) = \frac{(p^0-|p|)p}{2\sqrt{p^0|p|}}, \quad (4.10)$$

and μ' is taken to be zero, which corresponds to the Landau gauge formalism of the vector field. Finally $\tilde{h}_\sigma(p) = \tilde{h}_\sigma(-p)\geqslant 0$ is assumed.

Note that[17] there is an antiunitary operator Γ such that

$$\Gamma^2 = 1,\quad \Gamma\zeta = \bar{\zeta}\,\Gamma\ (\zeta\in C),$$

$$\Gamma\mathscr{H}(p;\sigma) = \mathscr{H}(p;\sigma)\Gamma,\quad \Gamma T\Gamma = -T, \quad (4.11)$$

where $T = \varphi\oplus -\varphi = \mathrm{diag}(1,1,-1,-1,-1,1)$. In the present case, Γ is explicitly given as

$$C\begin{pmatrix} 0 & 1' \\ 1' & 0 \end{pmatrix}, \quad 1' = \mathrm{diag}(-1,1,1,).$$

Let $\mathscr{H}_I(p;\sigma) =$ off diagonal part of $\mathscr{H}(p;\sigma)$. Then

$$\sum_p \mathrm{Tr}\mathscr{H}_I^*(p;\sigma)\mathscr{H}_I(p;\sigma) < \infty \qquad (4.12)$$

if and only if

$$\sum_p |\tilde{h}_\sigma(p)p|^2 < \infty .$$

This is the condition so that the kernel of H_I is Hilbert–Schmidt. Let $K = [H_{0:L} + H_I(L,\sigma)]\Theta$ and note that Θ is unitary and hermitian. Since $H_{0:L}\Theta$ is self-adjoint and since $H_I(L,\sigma)\Theta$ is a symmetric operator whose kernel is Hilbert–Schmidt, K has null-deficiency indices [Theorem 6.1 in Ref. 18]: Namely K is essentially self-adjoint. By taking the closure, we conclude the Θ-self-adjointness of $H(L,\sigma)|_{M=0}$.

The above discussion applies also for the renormalized Hamiltonian $H_S^R(L,\sigma)$ since it is again a bilinear Hamiltonian (see Sec. 5):

Theorem 4.6: Let

$$\sum_p |\tilde{h}_\sigma(p)p|^2 < \infty .$$

Then both $H(L,\sigma)$ and $H_R(L,\sigma)$ are Θ-self-adjoint at least for $M = 0$.

5. RENORMALIZED HAMILTONIAN

First of all we must point out that the unrenormalized Hamiltonian $H(L,\sigma)$ is unphysical. For that purpose, we show that $H(L,\sigma)|_{M=0}$ is unphysical. In this case, consider the following Bogolyubov transformation [17–22]:

$$\mathscr{A}(p;\lambda) \to S(p;\sigma)\mathscr{A}(p;\lambda),$$
$$\mathscr{A}^+(p;\lambda) \to \mathscr{A}^+(p;\lambda)S^*(p;\sigma). \qquad (5.1)$$

Here 6×6 matrix $S(p;\sigma)$ is chosen so that the CCR (with an indefinite metric) are preserved:

$$S^*TS = STS^* = T, \quad \text{or} \quad S^{-1} = S^{(T)} \equiv TS^*T, \qquad (5.2)$$

where $T = \varphi \oplus -\varphi = \mathrm{diag}(1,1,-1,-1,-1,1)$. Note

$$\det(\mathscr{H} - Tx) = \det(S^*\mathscr{H}S - Tx),$$
$$\det(\mathscr{H} - Tx) = -\det(\mathscr{H}T - x). \qquad (5.3)$$

Then if $\mathscr{H}(p;\sigma)$ is diagonalized by suitable S, then the diagonal elements are real because of hermiticity of $\mathscr{H}$, and are the roots of $\det(\mathscr{H}T - x) = 0$. However, $\det(\mathscr{H}T - x) = (x^2 - p^2)[x^4 - (p_0^2 + p^2)x^2 + p^2(p_0^2 + m^2)]$ for $|p| \leqslant \sigma$, where $\tilde{h}_0 = \chi_0$. Thus there are imaginary roots for $|p| > \mu/2|m|$ and the real roots for $|p| \leqslant \mu/2|m|$ are not of Lorentz covariant $(x^2 - p^2)$form. Then $H_S(L,\sigma)$ is not semibounded or has no real spectrum though it is Θ-self-adjoint. [4–7,9,10] However it is possible and easy to find an S such that the resultant Hamiltonian is partly diagonalized and self-adjoint. (It is, of course, again Θ-self-adjoint.)

Thus we must consider the renormalization. See Fig. 1. Here $\pi_{\mathrm{cov}}^{\mu\nu}(k)$ is the covariant (gauge invariant) amplitude given by

$$e^2 \int \frac{\mathrm{Tr}(\not p + M)\gamma^\mu(\not p + \not k + M)\gamma^\nu}{(p^2 - M^2)((p+k)^2 - M^2)} \frac{d^2p}{(2\pi)^2}$$
$$= \frac{e^2}{\pi}(-g^{\mu\nu}k^2 + k^\mu k^\nu)\int_0^1 \frac{x(1-x)}{M^2 - x(1-x)k^2}dx$$
$$= \frac{e^2}{\pi}\left(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}\right)[1 - 4M^2 I(k^2)], \qquad (5.4)$$

where $I(k^2) = \int_1^0 [(4M^2 - k^2) + y^2k^2]^{-1}\,dy$. The integration by d^2p converges if we calculate it in the gauge invariant way[3,23,24] which ensures the conservation of the current and that the gaugeon field decouples from the theory, at the second order of the perturbation. We say that the theory satisfies the stability condition if $4M^2 > \mu^2$. For $4M^2 < \mu^2$, we define $I(k^2 = \mu^2)$ (for example) by the principal integration by y.[14]

Corresponding to vector boson→vector boson, gaugeon→gaugeon, and vector boson→gaugeon, we denote these second order self-energies by $\delta E_{vv}^{\mu\nu}(k)$, $\delta E_{gg}^{\mu\nu}(k)$, and $\delta E_{vg}^{\mu\nu}(k)$, respectively, ($\mu,\nu$ denote the polarizations of the external bosons and k denotes the momentum ($|k| \leqslant \sigma$ is assumed)).

First of all, by the classical perturbation theory, [14]

$$\delta E_{vv}^{\hat\mu\hat\nu}(k)$$
$$= -\langle \delta^{\hat\mu}(k)|[H_I(L,\sigma)(H_{0:L} - p_0)^{-1}H_I(L,\sigma)]|\delta^{\hat\nu(k)}\rangle_{\mathrm{conn.}}$$

where conn. means that we omit the disconnected (i.e., vacuum) diagrams. Denoting the Wick contraction by, [3] we get

$$\delta E_{vv}^{\hat\mu\hat\nu}(k)$$
$$= -\frac{e^2}{2\pi}\sum_{k',k''}\left\{\frac{2\pi}{L}\langle\delta^{\hat\mu}(k)|\tilde{A}_\mu(-k')\tilde{A}_\nu(-k'')|\delta^{\hat\nu}(k)\rangle\right\}$$
$$\times \delta_{k,-k'}\delta_{k,k''}\{\pi_{(1)}^{\mu\nu}(p_0(k),k) + \pi_{(2)}^{\mu\nu}(p_0(k),k)\}, \qquad (5.5)$$

where

$$\pi_{(1)}^{\mu\nu}(p_0(k),k)$$
$$= \frac{2\pi}{L}\langle\Omega|W^{(-)\mu}(-k;L)$$
$$\times (H_{0:L} - p_0(k))^{-1}W^{(+)\nu}(k;L)|\Omega\rangle,$$

$$\pi_{(2)}^{\mu\nu}(p_0(k),k)$$
$$= \frac{2\pi}{L}\langle\Omega|W^{(-)\nu}(k;L)$$
$$\times (H_{0:L} + p_0(k))^{-1}W^{(+)\mu}(-k;L)|\Omega\rangle.$$

We define

$$\pi^{\mu\nu}(p_0(k),k) = \frac{e^2}{2\pi}\sum_{i=1}^2 \pi_i^{\mu\nu}(p_0(k),k).$$

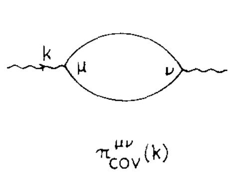

FIG. 1. Gauge invariant amplitude.

The gaugeon self-energy is given by replacing $p_0(k)$ by $p_0'(k)$.

Now

$$\pi^{\mu\nu}_{(^1_2)}(p_0(k),k) = \frac{2\pi}{L}\sum_p S^{\mu\nu}(p,p\pm k) \times(\omega(p)+\omega(p\pm k)\mp p_0(k))^{-1},$$

where

$$S^{\mu\nu}(p,p') = S^{\nu\mu}(p,p') = (v(p),\gamma^0\gamma^\mu u(p'))(u(p'),\gamma^0\gamma^\nu v(p)) = \frac{1}{2}\Big[-g^{\mu n}\frac{M^2}{\omega\omega'}+\delta^{\mu\nu}\Big(1+\frac{p}{\omega}\Big) \times\Big(1+\frac{p'}{\omega'}\Big)+\frac{p'}{\omega'}-\frac{p}{\omega}\Big],$$

with $\omega=\omega(p)$, $\omega'=\omega(p')$. Note $S^{\mu\nu}(p,p\pm k)=0(p^{-2})$ as $|p|\to\infty$. (See also Lemma 2.1.)

We first consider the massless fermion case, i.e., in the Schwinger model limit, noting that $p/\omega(p)=\epsilon(p)$ for $p\neq 0$, we have

$$\pi^{\mu\nu}_{(1)}(p_0(k),k)=\epsilon(\hat{k}^\mu\hat{k}^\nu)\frac{|k|}{|k|-p_0(k)},$$

$$\pi^{\mu\nu}_{(2)}(p_0(k),k)=\epsilon(\tilde{k}^\mu\tilde{k}^\nu)\frac{|k|}{|k|+p_0(k)},$$

where

$$\hat{k}^\mu=(|k|,k),\qquad \tilde{k}^\mu=(|k|,-k).$$

Thus

$$\pi^{\mu\nu}(p_0(k),k)=\frac{e^2}{\pi}\begin{cases}-k^2/\mu^2 & \text{for } \mu=\nu,\\ -kp_0(k)/\mu^2 & \text{for } \mu\neq\nu.\end{cases}$$

On the other hand, from the covariant perturbation theory,

$$\pi^{\mu\nu}_{\rm cov}(p_0(k),k)=\frac{e^2}{\pi}\Big(g^{\mu\nu}-\frac{k^\mu k^\nu}{k^2}\Big)\Big|_{k^0=p_0(k)},$$

namely

$$\pi^{\mu\nu}_{\rm cov}(p_0(k),k)=\pi^{\mu\nu}(p_0(k),k)+\frac{e^2}{\pi}(-1)\delta^{\mu 1}\delta^{\nu 1}. \tag{5.6}$$

Next we consider the amplitude ($v\to g$) following this line. Thus in order to renormalize the amplitudes following the covariant perturbation theory, we have

$$R_{L,\sigma}(A_\mu)=-\frac{m^2}{2}\int_\Lambda :A^2_{1,\sigma}(x,0):dx,\quad m=e/\sqrt{\pi}. \tag{5.7}$$

The case of $M\neq 0$ will be similarly discussed (by approximating the theory by $L=\infty$ theory to calculate the energies), and we see that $R_{L,\sigma}(A_\mu)$ is independent of M.

Finally we consider the vacuum energy which should be a real quantity chosen so that $\inf \operatorname{spec} H_R(L,\sigma)=0$. In the present case, however, the spectrum of $H_R(L,\sigma)$ is not necessarily real since the indefinite formalism is used. In such a case, the vacuum energy renormalization may be meaningless. But if $\operatorname{spec} H_R(L,\sigma)\subset R$, then we can expect that the

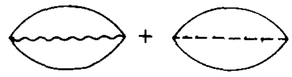

FIG. 2. The lowest order vacuum energy.

divergent terms in $E(L,\sigma)$ as $\sigma\to\infty$ is the second order vacuum energy $E_2(L,\sigma)$ as a conclusion of the power-counting theorem.[2,3]

The second order vacuum energy $E_2(L,\sigma)$ is given by Fig. 2, and

$$-\langle\Omega|H_I(L,\sigma)H^{-1}_{0;L}H_I(L,\sigma)|\Omega\rangle = -e^2\frac{L}{(2\pi)^2}\Big[\frac{2\pi}{L}\sum_{|k|<\sigma}\Big(\frac{2\pi}{L}\sum_p S^{\mu\nu}(p,p-k) \times\widetilde{U}_{\mu\nu}(k)(p_0(k)+\omega(p)+\omega(p-k))^{-1} + \frac{2\pi}{L}\sum_p S^{\mu\nu}(p,p-k)\widetilde{B}_{\mu\nu}(k) \times(p_0'(k)+\omega(p)+\omega(p-k))^{-1}\Big)\Big], \tag{5.8}$$

where

$$\widetilde{U}_{\mu\nu}(k)=-\frac{g_{\mu\nu}\mu^2-k_\mu k_\nu}{2\mu^2 p_0(k)},\quad k^\mu=(p^0(k),k),$$

$$\widetilde{B}_{\mu\nu}(k)=-\frac{k'_\mu k'_\nu}{2\mu^2 p_0'(k)},\quad k'^\mu=(p_0'(k),k).$$

Though the integration by p coverges absolutely (see Lemma 2.1), the integration by k diverges logarithmically as $\sigma\to\infty$ (after the integration by p). $E_2(L,\sigma)$ diverges like const $L\sigma^2$ in the Proca formalism.

The difference $\widetilde{E}(L,\sigma)=E(L,\sigma)-E_2(L,\sigma)$ (renormalized vacuum energy) may converge even for $\sigma\to\infty$ as well as the cases in $(\phi^4)_3$, $(Y)_2$[25–29]:

(i) $\widetilde{E}(L,\sigma)$ converges for $\sigma\to\infty$?

(ii) $\lim_{L,\sigma\to\infty}\frac{1}{L}\widetilde{E}(L,\sigma)$ exists ?

These questions will be affirmatively answered for $M=0$ in Sec. 6.2. For use in that section, we calculate $E_2(L,\sigma)$ in the Schwinger model limit:

$$E_2(L,\sigma) = -L\frac{m^2}{8\pi\mu^2}\Big(\frac{2\pi}{L}\sum_{|k|<\sigma}k^2(p_0^{-1}(k)-p_0'^{-1}(k))\Big). \tag{5.9}$$

6. THE SCHWINGER MODEL

A. General aspect

We study postulates which QED-type models should satisfy. In such a formalism, we require that negative probability states do not appear in physical world.[2,3,30,31]

Postulate [0]: There is a Hilbert space $\mathscr{H}$ equipped with an indefinite sesquilinear form $\langle\ ,\ \rangle=(\ ,\Theta)$, where Θ is unitary and Hermitian. The Hilbert space $\mathscr{H}$ contains at least a linear subspace $\mathscr{H}'$ such that

$$\langle x,x\rangle\geqslant 0\quad\text{for}\quad x\in\mathscr{H}'.$$

Let

$$\mathscr{H}''=\{x\in\mathscr{H}';\ \langle x,x\rangle=0\}.$$

Then the physical subspace $\mathscr{H}_{\rm phys}$ is defined by $\overline{(\mathscr{H}'/\mathscr{H}'')}$.

Postulate [1]: There is a Θ-unitary representation of the Poincare group, and we denote it by $U(a;\Lambda)$. Then

[1.1] $U(a;\Lambda)\mathscr{H}'\subset\mathscr{H}'$, $\quad U(a;\Lambda)\mathscr{H}''\subset\mathscr{H}''$.

[1.2] Let $\{P^\mu\}$ be the generators of the space-time translation operator $U(a;1)$. Then these are Θ–self-adjoint and leave $\mathscr{H}'$ and $\mathscr{H}''$ invariant. $\{P^\mu \restriction \mathscr{H}_{\rm phys}\}$ always have real spectrum. Especially $P^0 \restriction \mathscr{H}_{\rm phys}$ has real nonnegative spectrum. Spec $\{P^\mu \restriction \mathscr{H}_{\rm phys}\}$ is always in the forward light cone.

[1.3] There is, at least, one vector Ω in $\mathscr{H}_{\rm phys}$ such that $P^\mu\Omega = 0$.

Therefore, we must consider whether or not the renormalized Hamiltonian satisfies the postulates except those destroyed by the cutoffs.

(1) Θ self-adjointness: $H_R(L,\sigma)^{(\Theta)} = H_R(L,\sigma)$.

(2) Let $\mathscr{H}^0\subset\mathscr{F}$ be the domain of $H_R(L,\sigma)$. Then from Postulates [1] and [2], one may expect that there is a direct orthogonal sum $\mathscr{H}^0 = \mathscr{H}^0_{\rm phys}\oplus\mathscr{H}^0_{\rm unphys}$ (with respect to $\langle\ ,\ \rangle$) such that

$$H_R\mathscr{H}^0_{\rm phys(unphys)}\subset\mathscr{H}^0_{\rm phys(unphys)},$$

$$\text{spec } H_R \restriction \mathscr{H}^0_{\rm phys(unphys)}\geqslant 0,$$

$$\langle x,x\rangle\geqslant 0 \text{ for } x\in\mathscr{H}^0_{\rm phys}.$$

Unfortunately such invariant subspaces cannot be in the Fock space in general, and this requirement is not completely true. See Sec. 6.3 and the final remark in that section.

The time evolution is expected to be implemented by the would-be Θ-unitary group $U(t) = \exp[itH_R(L,\sigma)]$. But $U(t)$ is not Θ-unitary and bonded in general. Then we define $\alpha_t(\cdot\cdot)$:

$$\alpha_t(A) = \exp it\delta_{H_R}.A$$

$$= \sum\frac{(it)^n}{n!}[H_R[H_R,\dots[H_R,A]\cdots]].$$

Formally α_t satisfies

$$\alpha_0 = 1,\quad \alpha_t\alpha_s = \alpha_{t+s},\quad [\alpha_s(A)]^{(\Theta)} = \alpha_s(A^{(\Theta)}).$$

Let $\Phi(f,t) = \alpha_t(\Phi(f,0))$, and let $j^\mu(x,0)$ be a conserved current provided that the time evolution is implemented by a free Hamiltonian. However,

$$\partial_\mu j^\mu(x,t) = \alpha_t(\partial_1 j^1(x,0) + i[H_R,j^0(x,0)])$$
$$= \alpha_t(i[H_I - R_{L,\sigma},j^0(x,0)])\neq 0.$$

Thus $\alpha_t(j^\mu(x,0))$ is not conserved in general, and we define suitable Heisenberg field operators which appear in the field equations by improved methods when this phenomenon takes place.

(3) Conservation of current: $\partial^\mu j_\mu(x,t) = 0$.

(4) Maxwell equation:

$$(\Box+\mu^2)A^\mu(x,t) - (1-\mu'^2/\mu^2)\partial^\mu B(x,t) - ej^\mu(x,t) = 0.$$

If B is a massless free field, then $\partial^\mu A_\mu = 0$ (Landau gauge).

Since the gaugeon field B interacts with matter fields in a nontrivial way, it is not obvious that the gaugeon field is represented as a free field. But as far as the vector field couples to a conserved current, this is ensured by the covariant perturbation theory.[3,24,32,33] If there exist a suitable space such that B is represented on a Fock space which is left invariant by the time evolution, we denote such a space as $\mathscr{F}_{\rm phys}$. This may be completely different from the original Fock space (even if $L,\sigma<\infty$). Thus we set $\mathscr{F}_{\rm phys} = \mathscr{H}$. In this case,

$$\langle\Phi|[(\Box+\mu^2)A^\mu - ej^\mu]|\Psi\rangle = 0 \quad\text{for}\quad \Phi,\Psi\in\mathscr{H}_{\rm phys}.$$

Further,[3,11,12,30]

$$\Phi\in\mathscr{H}' \text{ if and only if } B^{(-)}(x,t)\Phi = 0,$$

where $B^{(-)}$ is the negative frequency part (annihilation part) of B. This is the Lorentz condition.

(5) Dirac equation:

$$(i\partial\!\!\!/ - eA\!\!\!/)\Psi(x,t) = 0.$$

(For the definition of $A\!\!\!/\Psi$, see Sec. 6.5.)

B. Diagonalization of $H^R_S(L,\sigma)$

The renormalized Hamiltonian reads

$$H^R_S(L,\sigma;\lambda) = \frac{2\pi}{L}\sum_{p>0}:\mathscr{A}^+(p;\lambda)\mathscr{H}_R(p;\sigma)\mathscr{A}(p;\lambda): + 1_\lambda(e(\lambda) - E(L,\sigma)) + (p=0\ \text{term}), \tag{6.1}$$

$$\mathscr{H}_R(p;\sigma) = \text{diag}\{p^0,|p|,-|p|,p^0,|p|,-|p|\}\quad\text{for}\quad |p|>\sigma, \tag{6.2}$$

$$\mathscr{H}_R(p;\sigma) = \begin{pmatrix}\mathscr{H}_{R,11}(p) & \mathscr{H}_{R,12}(p)\\ \overline{\mathscr{H}}_{R,12}(-p) & \overline{\mathscr{H}}_{R,11}(-p)\end{pmatrix}\quad\text{for}\quad |p|\leqslant\sigma,$$

with

$$\mathscr{H}_{R,11}(p) = \begin{pmatrix} p^0\left(1+\frac{m^2}{2\mu^2}\right) & \frac{m}{\mu}\omega' & i\frac{m^2}{2\mu^2}\kappa\\ \frac{m}{\mu}\omega' & |p| & 0\\ -i\frac{m^2}{2\mu^2}\kappa & 0 & |p|\left(-1+\frac{m^2}{2\mu^2}\right)\end{pmatrix},$$

$$\mathscr{H}_{R,12}(p) = \begin{pmatrix} p^0\frac{m^2}{\mu^2} & -\frac{m}{\mu}\omega & + & i\frac{m^2}{2\mu^2}\kappa\\ \frac{m}{\mu}\omega & 0 & & i\frac{m}{\mu}|p|\\ -i\frac{m^2}{2\mu^2}\kappa & i\frac{m}{\mu}|p| & & \frac{m^2}{2\mu^2}|p|\end{pmatrix}.$$

here

$$\kappa = \omega + \omega' = p^0p/\sqrt{p^0|p|}, \tag{6.3}$$

and for more general h_σ, H_S^R is given by replacing m by $m\tilde{h}_\sigma$.

Let

$$\mathscr{H}_R(p;\sigma)Tu_i = \omega_i u_i, \tag{6.4}$$

where ω_i is one of the roots of $\det[\mathscr{H}_R T - x] = 0$ and u_i is the corresponding eigenvector normalized as $(u_i, Tu_j) = (T)_{ij}$.

Thus $\{\omega_i; i = 1,...,6\} = \{\pm q^0, \pm |p|, \pm |p|\}$, with

$$q^0 = (p^2 + \bar{\mu}^2)^{1/2}, \quad \bar{\mu} = (\mu^2 + m^2)^{1/2}. \tag{6.5}$$

We obtain $S(p;\sigma)$ and $\mathscr{H}_0^S(p;\sigma) \equiv S^*(p;\sigma) \times \mathscr{H}_R(p;\sigma) S(p;\sigma)$ as follows:

$$(1)\ |p| > \sigma: S(p;\sigma) = 1 \quad \mathscr{H}_R(p;\sigma) = \mathscr{H}_0^S(p;\sigma), \tag{6.6}$$

(2) $|p| \leqslant \sigma$:

$$S(p;\sigma) = T(u_1, u_2, u_3, \Gamma u_1, \Gamma u_2, \Gamma u_3) = \begin{pmatrix} S_+(p) & S_-(p) \\ \bar{S}_-(-p) & \bar{S}_+(-p) \end{pmatrix}, \tag{6.7}$$

where

$$S_+(p) = \begin{pmatrix} \dfrac{\bar{\mu}^2 p^0 + \mu^2 q^0}{2\mu\bar{\mu}\sqrt{p^0 q^0}} & -\dfrac{m\kappa}{2\bar{\mu}p^0} & 0 \\ \dfrac{mp}{2\mu\bar{\mu}\sqrt{q^0|p|}} & \dfrac{\mu^2 + \bar{\mu}^2}{2\mu\bar{\mu}} & -i\dfrac{m}{2\mu} \\ \dfrac{im^2 p}{2\mu\bar{\mu}\sqrt{q^0|p|}} & i\dfrac{m}{2\bar{\mu}} & 1 \end{pmatrix},$$

$$S_-(p) = \begin{pmatrix} -\dfrac{\bar{\mu}^2 p^0 - \mu^2 q^0}{2\mu\bar{\mu}\sqrt{p^0 q^0}} & \dfrac{m\kappa}{2\bar{\mu}p^0} & 0 \\ -\dfrac{mp}{2\bar{\mu}\sqrt{q^0|p|}} & -\dfrac{m^2}{2\mu\bar{\mu}} & -i\dfrac{m}{2\mu} \\ -\dfrac{im^2}{2\mu\bar{\mu}\sqrt{q^0|p|}} & \dfrac{im}{2\bar{\mu}} & 0 \end{pmatrix},$$

and

$$\mathscr{H}_0^S(p;\sigma) = \mathrm{diag}\{q^0, |p|, -|p|, q^0, |p|, -|p|\}. \tag{6.8}$$

Note

$$U{:}\mathscr{A}^+\mathscr{H}_R\mathscr{A}{:}U^{-1} = {:}\mathscr{A}^+\mathscr{H}_0^S\mathscr{A}{:} + \frac{L}{2\pi}\sum_{i=1}^{3}[\mathscr{H}^S_{0,3+i,3+i} - \mathscr{H}_{R,3+i,3+i}]\varphi_{ii}.$$

Thus the vacuum energy $E(L,\sigma)$ should be chosen as

$$\sum_{0<p\leqslant\sigma}\left[q^0 - p^0 - \frac{m^2}{2\mu^2}(p^0 - |p|)\right]. \tag{6.9}$$

Then from (5.9) with $p_0' = |p|$, $\tilde{E}(L,\sigma) = E(L,\sigma) - E_2(L,\sigma)$ is

$$\sum_{0<p\leqslant\sigma}\left(q^0 - p^0 - \frac{m^2}{2\mu^2}(p^0 - |p|) + \frac{m^2}{2\mu^2}(p^2 p_0^{-1} - |p|)\right),$$

which converges as $\sigma \to \infty$. Thus we get

Theorem 6.1: The renormalized vacuum energy $\tilde{E}(L,\sigma)$ converges to $\tilde{E}(L,\infty)$ as $\sigma \to \infty$, and the renormalized vacuum energy per unit volume converges as $L,\sigma \to \infty$:

$$\lim_{L,\sigma\to\infty}\frac{\tilde{E}(L,\sigma)}{L} = \frac{1}{2\pi}\int_0^\infty dp\left[q^0 + \frac{m^2}{2\mu^2}p^2 p_0^{-1} - \left(1 + \frac{m^2}{2\mu^2}\right)p_0\right].$$

Finally we obtain the transformation rules of the operators $\{\phi_\sigma'(x,0), J_\sigma'(x,0), B_\sigma'(x,0), U_{\mu,\sigma}'(x,0), j_{\mu,\sigma}'(x,0), B_{\mu,\sigma}'(x,0)\}$, where σ means the sharp momentum cutoff and (′) means that the zero frequency parts are omitted. Let $U(L,\sigma;\lambda)$ be an operator on $\mathscr{F}(\lambda)$ which implements (5.1) and let $U(L,\sigma) = \otimes_\lambda U(L,\sigma;\lambda)$. Then (we omit σ, (′) and (L,σ) for simplicity):

$$U\phi U^{-1} = \frac{\bar{\mu}}{\mu}\tilde{\phi}, \quad UBU^{-1} = B,$$

$$UJU^{-1} = \frac{\bar{\mu}}{\mu}J + \frac{m}{\mu^2}B,$$

$$UU_0U^{-1} = \tilde{U}_0, \quad UU_1U^{-1} = \tilde{U}_1 + \frac{e}{\mu\bar{\mu}}j_1,$$

$$UJ_0U^{-1} = \frac{e\mu}{\pi\bar{\mu}}\tilde{U}_0 + \frac{\mu}{\bar{\mu}}j_0, \quad Uj_1U^{-1} = \frac{\bar{\mu}}{\mu}j_1 + \frac{m}{\mu^2}B_1, \tag{6.10}$$

$$UB_0U^{-1} = -m^2\tilde{U}_0 + \frac{e\mu}{\bar{\mu}}j_0, \quad UB_1U^{-1} = B_1.$$

Here $\tilde{U}_\mu$ is the Proca field of mass $\bar{\mu} = (m^2 + \mu^2)^{1/2}$ and $\tilde{\phi}$ is a scalar field of mass $\bar{\mu}$ defined by $-\bar{\mu}\epsilon_{\alpha\beta}\tilde{\phi} = \partial_\alpha\tilde{U}_\beta - \partial_\beta\tilde{U}_\alpha$. Note that there is a mixing between J and B, which is closely related to the so-called Higgs mechanism.

The Maxwell equation and the conservation of the current do not hold among these transformed operators, which means that these operators cannot be the Heisenberg operators which appear in the field equations.

C. Properties of $U(L,\sigma;\lambda)$, $U(L,\sigma)$ and the physical vacuum (I)

Each $\mathscr{F}_i$ $(=\mathscr{F}_U, \mathscr{F}_J, \mathscr{F}_B)$ is the Fock space constructed from the one particle space $\mathscr{H}_i$. Let $\mathscr{H}_+ = \mathscr{H}_U \oplus \mathscr{H}_J$, $\mathscr{H}_- = \mathscr{H}_B$ and let $\varphi = P_+ - P_-$ be a unitary and hermitian operator on $\mathscr{H} = \mathscr{H}_+ \oplus \mathscr{H}_-$, where P_+ and P_- are projections onto $\mathscr{H}_+$ and $\mathscr{H}_-$, respectively. We define an indefinite product $\langle\ ,\ \rangle = (\ ,\varphi)$ in $\mathscr{H}$ as before. Further we can define the Fock space $\mathscr{F}(\mathscr{H}) = \mathscr{F}(\mathscr{H}_+) \otimes \mathscr{F}(\mathscr{H}_-)$ and the creation operators $\{a^*(f); f\in\mathscr{H}\}$ as before.

When $\Theta = \Gamma(\varphi) = 1$, the annihilation operator is defined by $a(f) = [a^*(Cf)]^*$, where C is the complex conjugation operator. In the present case, by unitary and hermitian $\Theta = \Gamma(\varphi)$, we define an indefinite inner product $\langle\ ,\ \rangle = (\ ,\Theta)$ on the Fock space, which we denote $\{\mathscr{F},$

$\langle\ ,\ \rangle\}$. In this case we define the annihilation operator by

$$a(f) = [a^*(Cf)]^{(\Theta)} = [a^*(Cf)]^*\Theta = [a^*(\varphi Cf)]^*.$$

$\{a(f), a^*(g)\}$ obey the CCR with an indefinite metric:

$$[a^*(f),a^*(g)] = [a(f),a(g)] = 0,$$
$$[a(f),a^*(g)] = (Cf,\varphi g) = \langle Cf,g\rangle. \quad (6.11)$$

We define the Θ-self-adjoint (Segal's) field by

$$\Phi_\varphi(f) = (1/\sqrt{2})[a^*(f) + [a^*(f)]^{(\Theta)}]. \quad (6.12)$$

Then the CCR reduce to

$$[\Phi_\varphi(f),\Phi_\varphi(g)] = i\,\mathrm{Im}(f,\varphi g) = -i\,\mathrm{Re}(f,\varphi Jg), \quad (6.13)$$

where J is the multiplication operator of $i = (-1)^{1/2}$. Conversely

$$a^*(f) = [1/(2)^{1/2}][\Phi_\varphi(f) - i\Phi_\varphi(Jf)],$$
$$[a^*(f)]^{(\Theta)} = [1/(2)^{1/2}][\Phi_\varphi(f) + i\Phi_\varphi(Jf)],$$

and the Fock vacuum Ω is defined by $[a^*(f)]^{(\Theta)}\Omega = 0$ for any f.

Let (B_+, B_-) be a pair of bounded operators on $\mathscr{H}$. Then $T_B = B_+ + CB_-$ $(= B)$ is called an (invertible) φ-symplectic operator (or φ-Bogolyubov transformation) if and only if

$$JT_B^{(\varphi)}J^{-1}T_B = T_BJT_B^{(\varphi)}J^{-1} = 1, \quad (6.14)$$

where $T_B^{(\varphi)}$ is the adjoint of T_B with respect to $\mathrm{Re}\langle\ ,\ \rangle$: $T_B^{(\varphi)} = \varphi T_B^*\varphi$, $T_B^* = B_+^* + B_-^*C$. Hence the transformation

$$\Phi_\varphi(f) \to \Phi_\varphi(T_Bf) \equiv \pi_B(\Phi_\varphi(f)) \quad (6.15)$$

leaves the CCR invariant. We study a linear transformation U_B^{-1} on $\mathscr{F}$ which implements

$$U_B^{-1}P(\Phi_\varphi(f_1),...,\Phi_\varphi(f_n))\Omega = \pi_{B^{-1}}(P)\Omega_B, \quad (6.16)$$

for any polynomial P of the fields and any test function f, where $\pi_{B^{-1}}(P) = P(\Phi_\varphi(T_B^{-1}f_1),...,\Phi_\varphi(T_B^{-1}f_n))$ and $\Omega_B = U_B^{-1}\Omega$.

Definition 6.1: A φ-symplectic operator T_B is called

(i) Θ-unitarily implementable if there is a Θ-unitary (i.e., bijective Θ-isometric) operator U_B^{-1} such that $\pi_B(P) = U_BPU_B^{-1}$,

(ii) weakly Θ-unitarily implementable if there are a Θ-isometric (not necessarily bounded) operator U_B^{-1} and a cyclic vector $\Omega_B \in \mathscr{F}$ such that (6.16) holds,

(iii) Θ-unitarily quasi-implementable if $\det[1 + B_-^{(\varphi)} \times B_-]$ converges to a nonzero positive finite constant, where $B_- = C(T_B - JT_BJ^{-1})/2$.

First of all, using $[\Phi_\varphi(T_B^{-1}f) + i\Phi_\varphi(T_B^{-1}Jf)]\Omega_B = 0$, we get:

Theorem 6.2 [34]: If $\Omega_B \in \mathscr{F}$, then

(i) B_- is of Hilbert–Schmidt class,

(ii) $(-\infty, 0]$ is in the resolvent set of $B_+^{(\varphi)}B_+ = 1 + B_-^{(\varphi)}B_-$.

Theorem 6.3: The necessary and sufficient condition for T_B to be Θ-unitarily implementable is that $[B_\pm, \varphi] = 0$ and $B_- \in$H.S. (H.S. means the Hilbert–Schmidt class).

Proof: If T_B is Θ-unitarily implementable, then $\|U_B\| = \|U_B^{-1}\| < \infty$, since $U_B^{-1} = \Theta U_B^*\Theta$. Let $f \in \mathscr{H}_+ = P_+\mathscr{H}$. Then $\Phi_\varphi(f)$ is self-adjoint. $\Phi_\varphi(T_Bf) = \Phi_\varphi(P_+T_Bf) + \Phi_\varphi(P_-T_Bf)$ is a normal operator because $\Phi_\varphi(P_+g)$ and $i\Phi_\varphi(P_-g)$ are mutually commuting self-adjoint operators for any $g \in \mathscr{H}$. Now

$$U_Be^{i\Phi_\varphi(f)}U_B^{-1} = e^{i\Phi_\varphi(P_+T_Bf)}e^{i\Phi_\varphi(P_-T_Bf)},$$

where $\|\exp[i\Phi_\varphi(f)]\| = \|\exp[i\Phi_\varphi(P_+T_Bf)]\| = 1$ and $\|\exp[i\Phi_\varphi(P_-T_Bf)]\| = \infty$ whenever $P_-T_Bf \neq 0$. Then if T_B is Θ-unitarily implementable, $P_-T_BP_+ = 0$. By considering $\Phi_\varphi(f)$ with $f \in \mathscr{H}_- = P_-\mathscr{H}$, we also get $0 = P_+T_BP_-$ whenever T_B is Θ-unitarily implementable.

Thus we can restrict ourselves to usual symplectic operators which commute with φ. A result of Shale [17–22,34] can apply for the rest of the proof. □

If $B_- \in$H.S., then $B_-^{(\varphi)}B_- = B_+^{(\varphi)}B_+ - 1$ is of trace class. Then Theorems 6.2 and 6.3 mean: quasi ⊃ weakly ⊃ unitarily.

Definition 6.2: Let K be a real linear subspace of $\mathscr{H}$ such that $\mathscr{H} = K \oplus JK$, where the direct orthogonal sum refers the orthogonality with respect to both $\mathrm{Re}(\ ,\ )$ and $\mathrm{Re}\langle\ ,\ \rangle$. A φ-self-adjoint φ-symplectic operator S is called a generalized φ-scaling if S leaves K and JK invariant:

$$S = \begin{pmatrix} h & 0 \\ 0 & h^{-1} \end{pmatrix} \quad \text{on } K \oplus JK. \quad (6.17)$$

Theorem 6.4 [34]: Under the conditions of Theorem 6.2,

$$T = UH, \quad (6.18)$$

where U is φ-unitary and H is a φ-self-adjoint φ-symplectic operator with its spectrum in the right half plane. Further this decomposition is unique. If T leaves K and JK in definition 6-2 invariant, then H equals a generalized φ-scaling S.

This is proved [34] by applying an integral formula owing to Dunford and Schwartz [15,25] which gives $H = (T^{(\varphi)}T)^{1/2}$ with its spectrum in the right half plane. (Note that $T^{(\varphi)}T$ is a φ-self-adjoint φ-symplectic operator.)

Given a generalized φ-scaling S, let $T_B = V_1SV_2$ with V_i φ-unitary. Then, since $\{V_i\}$ are bounded operators,

$$B_- \in \text{H.S.} \longleftrightarrow S_- \in \text{H.S.},$$
$$\det[1 + B_-^{(\varphi)}B_-] = \det[1 + S_-^{(\varphi)}S_-].$$

Further if V is φ-unitary, then $\Gamma(V)\Phi_\varphi(f)\Gamma(V^{-1}) = \Phi_\varphi(Vf)$. Then

$$U_B = \Gamma(V_1)U_S\Gamma(V_2) \quad (6.19)$$

for $T_B = V_1SV_2$. Since $\|V\| \geqslant 1$ if V is φ-unitary, $\Gamma(V)$ is unbounded Θ-isometric in general. If $T_B = VS$, then $\Omega_B = \Omega_S$ since $\Gamma(V)\Omega = \Omega$.

We introduce the Q-space method [14,22,34,36]: Namely introduce the following unitary operator W:

$$W\mathscr{F} = L^2(Q,d\mu_0), \quad W\Omega = 1, \quad Q = R^\infty,$$
$$d\mu_0 = \prod_{i=1}^{\infty}(\pi)^{-1/2}\exp[-q_i^2]dq_i, \quad (6.20)$$

$$W\Phi_\varphi(e_i)W^{-1} = \begin{cases} q_i, & e_i \in K_+, \\ -iq_i, & e_i \in K_-, \end{cases}$$

$$W\Phi_\varphi(Je_i)W^{-1} = \begin{cases} i(-\partial_i + q_i), & e_i \in K_+, \\ -\partial_i + q_i, & e_i \in K_-, \end{cases}$$

where $\partial_i = \partial/\partial q_i$ and $\{e_i\}$ is complete orthonormal basis in K with respect to both (,) and ⟨ , ⟩, and $K_\pm = P_\pm K$.

In terms of the q-variables, Ω_S satisfies

$$\left[\psi^*(h^{-1})^{tr}\psi\{q\} + \psi^* h^{tr}\psi\left\{\frac{\partial}{\partial q} - q\right\}\right]\Omega_S(q) = 0$$

with $\langle \Omega_S, \Omega_S \rangle = 1$. Here $h_{ij} = (e_i, he_j)$ and $\psi = P_+ - iP_-$ is a unitary operator such that $\psi^2 = \varphi$ and $C\psi C = \psi^* = \psi^{-1}$. Then we obtain a formal vacuum $\Omega_S(q)$:

$$\Omega_S(q) = [\det(\alpha)]^{1/4}\exp[-\tfrac{1}{2}(q,(\alpha - 1)q)]. \qquad (6.21)$$

Here $\alpha = \psi^*(h^{-2})^{tr}\psi = \psi h^{-2}\psi^*$ is a φ-self-adjoint symmetric matrix since h is φ-self-adjoint and real (namely $ChC = \bar{h} = h$). Obviously,

$$S_- \in \text{H.S.} \longleftrightarrow \alpha - 1 \in \text{H.S.} \longleftrightarrow \alpha_r - 1, \alpha_i \in \text{H.S.},$$

where α_r and $i\alpha_i$ are the self-adjoint (= real) and skew-self-adjoint (= imaginary) parts of α, respectively. Note that $\varphi\alpha_r\varphi = \alpha_r$ and $\varphi\alpha_i\varphi = -\alpha_i$. Then $\det(\alpha) > 0$ whenever $\alpha_r > 0$.

Theorem 6.5 [34]: Let $S_- \in$H.S. and let $\alpha_r > c$ for some $c > 0$. Then

(i) $\Omega_S \in L^2(Q, d\mu_0)$,

(ii) $\langle \Omega, \Omega_S \rangle = \det^{-1/4}[1 + S_-^{(\varphi)}S_-]$ is positive nonvanishing finite,

(iii) $\Omega_{S^{-1}} \in L^2(Q, d\mu_0)$.

Theorem 6.6 [34]: Under the assumption of the above theorem,

(i) Ω_S and $\Omega_{S^{-1}}$ are cyclic for $\mathscr{F}$,

(ii) $\langle \Omega_S, P\Omega_S \rangle = \langle \Omega, \pi_S(P)\Omega \rangle$.

Theorem 6.7: For a generalized φ-scaling S, the vector Ω_S which satisfies

$$\rho_S(P) = \langle \Omega, \pi_S(P)\Omega \rangle = \langle \Omega_S, P\Omega_S \rangle \qquad (6.22)$$

cannot be in the Fock space if inf spec $\alpha_r < 0$.

Proof: $\alpha = h^{-2}$ takes the following from on $JK = JK_+ \oplus JK_-$:

$$\begin{pmatrix} (\alpha_r)_{++} & (\alpha_i)_{+-} \\ (\alpha_i)_{-+} & (\alpha_r)_{--} \end{pmatrix}.$$

Namely $\alpha_r = (\alpha_r)_{++} \oplus (\alpha_r)_{--}$ on $JK_+ \oplus JK_-$. First assume that inf spec $(\alpha_r)_{++} = -\lambda$ $(\lambda > 0)$. Let $f \in JK_+$ be the eigenvector belonging to the eigenvalue $-\lambda$. Then for $f \in JK_+$, $\|\exp i\Phi_\varphi(f)\| = 1$ and

$$\rho_S(\exp i\Phi_\varphi(f)) = \langle \Omega, \exp i\Phi_\varphi(Sf)\Omega \rangle = \exp[-\tfrac{1}{4}\langle Sf, Sf \rangle]$$

$$= \exp[-\tfrac{1}{4}(f, (\alpha_r)_{++} f)] = \exp\left[\frac{\lambda}{4}\|f\|^2\right].$$

If $\lambda > 0$, then the right hand can be made arbitralily large, which contradicts $|\langle \Omega_S, \exp i\Phi_\varphi(f)\Omega_S \rangle| \leqslant \|\Omega_S\|^2$. The case of inf spec $(\alpha_r)_{--} < 0$ is similarly discussed. □

The proof of Theorem 6.7 suggests that the Lorentz state defined by $\langle \Omega, \pi_B(\ldots)\Omega \rangle$ cannot be continuous on the C^*-algebra generated by $\{\exp i\Phi(f); f \in \mathscr{H}\}$ in general $(\Phi(f) = \Phi_{\varphi = 1}(f))$.

The following is proved in Ref. 34 (see also Refs. 14, 25, 35–38):

Theorem 6.8: Under the assumption of theorem 6-6, $\Omega_S \in D(N)$, where $N = \Sigma_i a^*(e_i)a(e_i)$ is the number operator.

We turn to the model: $\mathscr{F}(\mathscr{H}) = \otimes_{p>0}\mathscr{F}(p) \otimes \mathscr{F}(p = 0)$, where $\mathscr{F}(p)$ is the Fock space constructed from the one-particle Hilbert space $\mathscr{H}(p) = C^6$ with an indefinite inner product $(\ ,\varphi)$ with $\varphi \equiv \varphi \oplus \varphi = \text{diag}(1,1,-1,1,1,-1)$.

Let $B_\pm(\sigma) = \oplus_{0<p\leqslant\sigma} B_\pm(p)$, where

$$B_+(p) = \begin{pmatrix} S_+^*(p) & 0 \\ 0 & S_+^*(-p) \end{pmatrix},$$

$$B_-(p) = \begin{pmatrix} 0 & S_-^*(-p) \\ S_-^*(p) & 0 \end{pmatrix}. \qquad (6.23)$$

Let

$$T(\sigma) = B_+(\sigma) + CB_-(\sigma) = \oplus_{0<p\leqslant\sigma} T(p), \qquad (6.24)$$

where $T(p) = B_+(p) + CB_-(p)$ is a φ-symplectic operator on $\mathscr{H}(p)$. $T(\sigma)$ leaves $K_\varphi \equiv K_+ \oplus iK_-$ and JK_φ invariant and takes the following form on $\mathscr{H} = K_\varphi \oplus JK_\varphi$:

$$\begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} = \begin{pmatrix} B_+(\sigma) + \varphi B_-(\sigma) & 0 \\ 0 & B_+(\sigma) - \varphi B_-(\sigma) \end{pmatrix}. \qquad (6.25)$$

Then T has the decomposition $T = US$ in Theorem 6.4. Let $\alpha = \alpha(\sigma) = T_2^{(\varphi)}T_2$. Then in the Q-space

$$\Omega_B = \Omega(L,\sigma) = \det{}^{1/4}(\sigma)\exp[-\tfrac{1}{2}(q,(\alpha - 1)q)]. \qquad (6.26)$$

It is easily seen that (replace m by $\tilde{h}_\sigma m$)

$$c^{-1} \geqslant \alpha(\sigma)_r \geqslant c \longleftrightarrow |m\tilde{h}_\sigma(p)/\mu| < 1,$$

$$B_- \in \text{H.S.} \longleftrightarrow \sum |\tilde{h}_\sigma^2(p)| < \infty. \qquad (6.27)$$

In the following, we set $\tilde{h}_s = \chi_\sigma$. Then $\Omega_B \in \mathscr{F}$ if $\sigma < \infty$ and $|m/\mu| < 1$. In this case

$$Z(L,\sigma) \equiv \langle \Omega, \Omega_B \rangle = \prod_{0<p\leqslant\sigma} \det{}^{-1}(S_+(p)) \qquad (6.28)$$

and $\det^{-1}(S_+(p))$

$$= \frac{8\mu^2\bar{\mu}^2\sqrt{p^0 q^0}}{(4\mu^2 + m^2)(\bar{\mu}^2 p^0 + \mu^2 q^0) + 2m^2 m^2|p| - m^2|p|}$$

are nonvanishing finite for any $\mu, |m| \geqslant 0$ and p. Further $\det^{-1}(S_+(p)) = 1 + m^2/8p^2 + O(p^{-4})$ as $p \to \infty$. Then the overlap between the vacua uniformly converges to a nonvanishing finite value $Z(L)$ (from below) as $(\sigma_0 <)\sigma \to \infty$.

Theorem 6.9: (i) Let $\sigma < \infty$ and $|m/\mu| < 1$. Then $T(\sigma)$ is weakly Θ-unitarily implementable

(ii) $T(\sigma)$ is Θ-unitarily quasi-implementable for $\mu, |m| \geqslant 0$ and $\sigma \leqslant \infty$ (provided $L < \infty$).

Theorem 6.10: Define $\alpha_p(L,\sigma) = -\log Z(L,\sigma)/L$. Then $\lim_{L,\sigma\to\infty} \alpha_p(L,\sigma) \equiv \alpha_p$ exists (p means "periodic boundary condition"):

$$\alpha_p = \frac{1}{2\pi}\int_0^\infty dp \log\det[S_+(p)].$$

$Z(L)$ and $Z(L,\sigma)$ are necessarily in [0,1] when $\Theta = 1$, and is just the case seen in the models $P(\phi)_d, (Y)_d$, etc.[25,29,35,36,39] But in the present model, because of the indefinite metric, $Z(L)$ and $Z(L,\sigma)$ are not necessarily smaller than 1: In fact $\det^{-1}[S_+(p)] > 1$ for large p. Further α_p is positive if $\Theta = 1$. In the present model, however, this is not necessarily positive.

Though $\Omega(L,\sigma) \in \mathscr{F}$ and $\langle \Omega(L,\sigma), \Omega(L,\sigma)\rangle = 1$ if $\sigma < \infty$ and $|m/\mu| < 1$, $\Omega(L,\infty) \notin \mathscr{F}$. In fact $B_-(\infty) \notin$H.S., and moreover

$$\lim_{\sigma\to\infty} \|\Omega(L,\sigma)\|_2^2 = \lim_{\sigma\to\infty} \det{}^{1/4}[1 + (\alpha_r^{-1/2}\alpha_i\alpha_r^{-1/2})^2] = \infty.$$

On the other hand, the formal overlap $\det^{-1/4}[1 + B^{(\varphi)}_- B_-]$ converges to a nonvanishing finite value $Z(L)$ because $B^{(\varphi)}_-(p)B_-(p) =$ nilpotent matrix $+0(p^{-2})$. As is well known, if $\Theta = 1$, then the three notions in definition 6-1 are equivalent to each other. But for $\Theta \neq 1$, even if the formal overlap $\det^{-1/4}[1 + B^{(\varphi)}_- B_-]$ converges to a nonvanishing finite value, Ω_B is not necessarily in the Fock space.

Consider

$$\frac{d}{dt}\phi(t) = -iH_S^R(L,\sigma)\phi(t), \quad \phi(0) = \phi_0 \in \mathscr{F}.$$

Thus we formally have $\phi(t) = U(-t)\phi_0$ with $U(t) = \exp[itH_S^R]$. $U(t)$ is formally a one-parameter Θ-unitary group, but this cannot be well defined: $U(t)$ is unbounded in general. Instead, we consider

$$\alpha_t(\cdots) = \exp it\delta_{H_S^R(L,\sigma)}, \tag{6.29}$$

which implements the following transformation:

$$\alpha_t(\Phi_\varphi(f)) = \Phi_\varphi(T(t)f),$$

where $T(t) = T_B^{-1}K_0(t)T_B$ and $K_0(t)$ is a one-parameter unitary group defined by (H_0^S the diagonalized Hamiltonian)

$$\exp it\delta_{H_S^R(L,\sigma)} \cdot \Phi_\varphi(f) = \Phi_\varphi(K_0(t)f).$$

K_0 is a one-parameter unitary group commuting with φ. Then $T(t)$ is a one-parameter group of φ-symplectic operators. $T(t)_- \in$H.S. for all $t \in R$ whenever $B_- \in$H.S. because of the boundedness of B_+ and $K_0(t)$. However, even if T_B is weakly Θ-unitarily implementable, $T(t)$ is not always weakly Θ-unitarily implementable. This is Θ-unitarily quasi-implementable in general.[34]

Remarks: (1) One-parameter Θ-unitary group $U(t)$ is called stable if $\|U(t)\| \leqslant M < \infty$ for all $t \in R$. The necessary and sufficient condition for a Θ-self-adjoint operator H to be a generator of one-parameter stable Θ-unitary group is that H is similar to a self-adjoint operator H_0,[4–7,15] which is equivalent to that $D(H)$ is decomposed as $D(H)^+ \oplus D(H)^-$ where $D(H)^\pm$ are uniformly definite subspaces. This is not expected in physics [in fact either $U(L,\sigma)$ or $U(L,\sigma)^{-1}$ is unbounded (theorem 6-3)]. (2) If $|m/\mu| < 1$, then $H_S^R(L,\sigma)$ has the spectra resolution in terms of Θ-self-adjoint projections $\{E(\lambda); \lambda \geqslant 0\}$. But their ranges are not uniformly definite. If $|m/\mu| > 1$, then $\{E(\lambda)\}$ are not defined on $\mathscr{F}$.

D. Properties of $U(L,\sigma)$, $U(L,\sigma;\lambda)$ and the physical vacuum (II)

The topology of $\mathscr{H} = \mathscr{H}_+ \oplus \mathscr{H}_-$ [$\oplus$: orthogonality for $\langle\ ,\ \rangle = (\ ,\varphi)$], is defined by $(x,x)_1 = \langle x_+, x_+\rangle - \langle x_-, x_-\rangle$ ($x_\pm \in \mathscr{H}_\pm$), while the topology of $B^{-1}\mathscr{H} = B^{-1}\mathscr{H}_+ \oplus B^{-1}\mathscr{H}_-$ (B: φ-unitary) is defined by $(x,x)_2 = \langle x_+, x_+\rangle - \langle x_-, x_-\rangle$ ($x_\pm \in B^{-1}\mathscr{H}_\pm$). Since $B^{-1} = \varphi B^*\varphi$, there are positive constants μ_1 and μ_2 such that $\mu_1 \|x\|_1 \leqslant \|x\|_2 \leqslant \mu_2\|x\|_1$, which means that $\mathscr{H}$ and $B^{-1}\mathscr{H}$ are equivalent.[4]

On the other hand next consider $\mathscr{F}$ and $U_B^{-1}D(U_B^{-1})$ with $U_B^{-1} = \Gamma(B^{-1})$. Since $\|B\| = \|B^{-1}\| > 1$ if B does not commute with φ, $\Gamma(B)$ and $\Gamma(B^{-1})$ are unbounded operators. Therefore, $\mathscr{F}$ and $\mathscr{F}(B^{-1}\mathscr{H})$ are not equivalent. This is the reason why the set of weakly Θ-unitarily implementable B does not form a group and is not invariant by φ-unitary operators.

Definition 6.2: Let $\mathfrak{A}$ be a C^*-algebra with an identity. Let $\mathfrak{A}_0 \subset \mathfrak{A}$ be a Banach algebra containing the identity which has an algebraic involution Θ defined by $A^{(\Theta)} = \Theta A^*\Theta$. Here Θ is a unitary and Hermitian operator in $\mathfrak{A}$. Then the Banach algebra $\mathfrak{A}_0$ is called a Lorentz algebra.

In the present model, the Lorentz algebra $\mathfrak{A}_0$ is given from the C^*-algebra generated by $\{\exp i\Phi(f), \psi(g), \psi^*(g)\}$ by taking the new adjoint $A^{(\Theta)} = \Theta A^*\Theta$.

If $\Omega(L,\sigma) \in \mathscr{F}$, a (continuous) linear functional $\rho_{L,\sigma}(\ldots) = \langle \Omega(L,\sigma), \ldots, \Omega(L,\sigma)\rangle$ satisfies the following:

(1) $\rho_{L,\sigma}(1) = 1$,

(2) $\rho_{L,\sigma}(aA + bB) = a\rho_{L,\sigma}(A) + b\rho_{L,\sigma}(B)$,

(3) $\overline{\rho_{L,\sigma}(A)} = \rho_{L,\sigma}(A^{(\Theta)})$

for all elements $A, B, \cdots$, and constants a, b.

Definition 6.3: A linear functional ρ on a Lorentz algebra $\mathfrak{A}_0$ is called a Lorentz state if it satisfies (1)–(3). If $\mathfrak{A}_0$ is in the C^*-algebra consisting of the bounded operators on a Hilbert space and if ρ is given as $\langle\omega, \ldots, \omega\rangle$ with $\omega \in \mathscr{F}$, then ρ is called a vector Lorentz state.

Unfortunately, however, see the proof of Theorem 6.7, which implies that the Lorentz state cannot be continuous in general on the C^*-algebra generated by $\{\exp i\Phi(f), \psi(g), \psi^*(g)\}$.

The following conjecture is due to Hansen[40]:

Conjecture 6.1: Let $\mathfrak{A}_0$ be a Banach algebra ($\ni 1$) with an algebraic involution $\dagger$ such that $\|A^\dagger\| = \|A\|$. Let ρ be a continuous Lorentz state on $\mathfrak{A}_0$ such that $\rho(1) = 1$, $\overline{\rho(A)} = \rho(A^\dagger)$. Then there exists a representation π_ρ of $\mathfrak{A}_0$ as closable operators on a Hilbert space $\mathscr{F}$, which satisfies:

(i) There is an invariant dense domain $\mathscr{D} \subset \mathscr{F}$ such that $\pi_\rho(A) \upharpoonright \mathscr{D}$ is closable;

(ii) there are cyclic vector Ω and a unitary and Hermitian operator Θ such that $\pi_\rho(A^\dagger)\subset\Theta\pi_\rho^*(A)\Theta,\rho(A) = (\Theta\Omega,\pi_\rho(A)\Omega)$;

In addition, can we prove the following?

(iii) $(\Omega,\Omega) = (\Theta\Omega,\Omega) = 1$;

(iv) the representation is unique up to a Θ-isometric operator (in some sense[4-7]).

Our preceding investigation strongly suggests that (iii) and (iv) also hold.

The present state is continuous only for $|m/\mu| < 1$. Then even if the conjecture holds, it applies only for the small coupling region such that the state is continuous. This means that it is not appropriate to consider a Lorentz state as a (continuous) linear functional on the Lorentz algebra. Thus it may be rather appropriate to consider the state as a linear functional on the field algebra which has no topology.

In the present model, the representation π_B is explicitly obtained and there is no problem: New Hilbert space (equipped with an indefinite metric) is again a Fock space:

$$\mathscr{H} = \mathscr{F}_{\rm phys} = \mathscr{F}_{\bar{U}} \otimes \mathscr{F}_F \otimes \mathscr{F}_B,$$

$$\mathscr{H}_{\rm phys} = \mathscr{F}_{\bar{U}} \otimes \mathscr{F}_F,$$

where $\mathscr{F}_{\bar{U}}$ is the Fock space of the Proca field of mass $\bar{\mu}$. In QED-type models where an indefinite metric is used, we require that the ghost particle (gaugeon) should be represented as a free field whether or not the model is exactly solvable. But the representation may mix the field operators, and then we must reconstruct the observable algebra from these field operators.

E. Heisenberg fermion field operator

For simplicity, we transform $S(p)$ into a real matrix by a unitary transformation

$$a(p)\to a_1(p)\equiv a(p), \quad A(p)\to a_2(p)\equiv A(p),$$

$$b(p)\to a_3(p)\equiv ib(p).$$

We denote the resultant matrix again by $S(p)$, and define

$$a_i^{(\pm)}(p)\equiv a_i(p) \pm a_i^*(-p).$$

We must obtain $U(L,\sigma) = \otimes_{0<p<\sigma} U(p)$ as an explicit function of the generators

(i) $-\theta_i[a_i^*(p)a_i^*(-p) - \text{h.c.}], \quad i = 1,2,3,$

(ii) $-\tau_k[(a_i^*(p)a_j^*(-p) - \text{h.c.}) \pm (p\to -p)],$

(iii) $-\varphi_k[(a_i^*(p)a_j(p) - \text{h.c.}) \pm (p\to -p)],$

where h.c. denotes the adjoint with respect to $\langle\ ,\ \rangle$ and (i,j,k) = permutations of (1,2,3). We want to obtain $U(p)$ in the form $U(p)V(p)W(p)$ where $U(p) = U_1(p)U_2(p)U_3(p)$, $V(p) = V_{12}(p)V_{23}(p)V_{31}(p)$, and $W(p) = W_{12}(p)W_{23}(p)W_{31}(p)$.

Here

$$U_i(p) = \exp\{(2\pi/L)[\text{expression (i)}]\},$$

$$V_{ij}(p) = \exp\{(2\pi/L)[\text{expression (ii)}]\},$$

and

$$W_{ij}(p) = \exp\{(2\pi/L)[\text{expression (iii)}]\}.$$

The sign $(+)$ and $(-)$ in V_{ij} and W_{ij} are determined by $S(p)$:

$$V_{12}(-),\quad V_{23}(+),\quad V_{31}(-),$$

$$W_{12}(-),\quad W_{23}(+),\quad W_{31}(-).$$

The angles $\{\theta_i,\tau_i,\varphi_i\}$ are also determined by $U(p)\mathscr{A}(p)U^{-1}(p) = S(p)\mathscr{A}(p)$ [in the sense of definition 6-1 (ii)], but not uniquely determined in general. As is shown in the Appendix, there is a unique real analytic solution $\{\theta_i(m),\tau_i(m),\varphi_i(m)\}$ in a neighborhood of $m = 0$ such that $\theta_i(0) = \tau_i(0) = \varphi_i(0) = 0$. Since $U_1(p)\cdots W_{31}(p)$ is a product of the exponential mappings of the generators in a neighborhood of $S = 1$, we only expect that the present expression generates S only in a neighborhood of $S = 1$. General solution will be obtained as a product of $\{U_l\}$, where each U_l generates S in a neighborhood of $S = 1$.

Now in the sense of definition 6.1 (ii),

$$\begin{aligned}\Psi_i(x,0)&\equiv U(L,\sigma)\psi_i(x,0)U(L,\sigma)^{-1}\\ &= Z^{-1/2}(L,\sigma)\exp[i\chi_{i,\sigma}^{(+)}(x,0)]\psi_i(x,0)\\ &\quad\times\exp[i\chi_{i,\sigma}^{(-)}(x,0)]\\ &\equiv Z^{-1/2}(L,\sigma){:}\exp[i\chi_{i,\sigma}(x,0)]\psi_i(x,0){:},\end{aligned} \tag{6.30}$$

where

$$\chi_{i,\sigma}(x,0) = \alpha_1 B'_\sigma(x,0) + \alpha_2 J'_\sigma(x,0) + (\gamma_5)_{ii}[\alpha_3\tilde{J}'_s(x,0) + \alpha_4\tilde{\phi}'_\sigma(x,0)],$$

with

$$\begin{aligned}&\alpha_1 = -e/\mu^2,\quad \alpha_2 = \sqrt{\pi}(1-\bar{\mu}/\mu),\\ &\alpha_3 = \sqrt{\pi}(1-\mu/\bar{\mu}),\quad \alpha_4 = e/\bar{\mu},\end{aligned} \tag{6.31}$$

and $Z^{-1/2}(L,\sigma)$ is given by

$$\exp\Big[-\frac{1}{L}\sum_{0<p<\sigma}\Big([-\alpha_1^2\mu^2 + \alpha_2^2 - 2\sqrt{\pi}\alpha_2 + \alpha_3^2 - 2\sqrt{\pi}\alpha_3]\frac{1}{2p} + \alpha_4^2\frac{1}{2q_0}\Big)\Big]. \tag{6.32}$$

Further $(+)$ and $(-)$ denotes the creation and annihilation parts of χ, respectively (see Appendix).

First note that Ψ_i is an analytic function of $m = e/(\pi)^{1/2}$, and can be defined for any (real) m though $U_1(p)...W_{31}(p)$ is not defined for large $|m|$. Further

(1) $(\alpha_2 - \sqrt{\pi})(\alpha_3 - \sqrt{\pi}) = \pi,$

(2) $-\alpha_1^2\mu^2 + \alpha_2^2 - 2\sqrt{\pi}\alpha_2 = 0,$

(3) $-(\alpha_3^2 - 2\sqrt{\pi}\alpha_3) = \alpha_4^2 = e^2/\bar{\mu}^2,$

where the first condition is the necessary and sufficient condition so that $\{\Psi_i\}$ obey the CAR and spin 1/2 transformation [see Ref. 8]. In this work, $\{\Psi_i\}$ are obtained through the Θ-isometric operator $U(L,\sigma)$, and this condition is automatically satisfied.

Let $P_{\rm phys}$ be the self-adjoint projection from $\mathscr{H}$ to $\mathscr{H}_{\rm phys}$ that is a closed subspace of $\mathscr{H}$ which does not contain any gaugeon. Let

$$\Psi_i^P(x,0) = P_{\rm phys}\Psi_i(x,0)P_{\rm phys}.$$

Then

$$\Psi_i^p(x,0) = Z^{-1/2}(L,\sigma){:}\ \exp[i\chi_{i,\sigma}^p(x,0)]\psi_i(x,0){:},$$

where

$$\chi_{i,\sigma}^p(x,0) = \alpha_2 J'_\sigma(x,0) + (\gamma_5)_{ii}[\alpha_3 \widetilde{J}'_\sigma(x,0) + \alpha_4 \tilde{\phi}'_\sigma(x,0)].$$

We investigate the Heisenberg vector field $\widetilde{A}_\mu$:

$$(i\partial\!\!\!/ - e\widetilde{A\!\!\!/})(x,t) = 0,$$

where

$$\Psi(x,t) = \exp[itH_0^S(L,\sigma)]\Psi(x,0) \times \exp[-itH_0^S(L,\sigma)].$$

Since $A\!\!\!/\Psi$ is not well defined in general in the limit of $L,\sigma = \infty$, we define the Dirac equation by

$$[i\partial\!\!\!/]\Psi - eZ^{-1/2} \times \begin{pmatrix} e^{i\chi_2'{}^{(+)}}[(\widetilde{A}_0 - \widetilde{A}_1)^{(+)}\psi_2 + \psi_2(\widetilde{A}_0 - \widetilde{A}_1)^{(-)}]e^{i\chi_2'{}^{(-)}} \\ e^{i\chi_1'{}^{(+)}}[(\widetilde{A}_0 + \widetilde{A}_1)^{(+)}\psi_1 + \psi_1(\widetilde{A}_0 + \widetilde{A}_1)^{(-)}]e^{i\chi_1'{}^{(-)}} \end{pmatrix} = 0,$$

where we omit L,σ and ($'$) for brevity. Thus we get (see the Appendix and Ref. 41)

$$\widetilde{A}_\mu(x,t) = \frac{1}{\mu^2}\partial_\mu B(x,t) + \frac{m}{\mu\bar{\mu}}\partial_\mu J(x,t) + \widetilde{U}_\mu(x,t).$$

This should be considered as the Heisenberg field operator which appears in the field equations, and is a vector field of mass $\bar{\mu}$ in the Landau gauge formalism, which is consistent with the convariant perturbation theory. But $P_{\rm phys}A_\mu P_{\rm phys}$ is not a Proca field.

The gaugeon field B is invariant by the interaction and the Maxwell equation means that the improved (conserved) current is

$$ej_\mu(x,t) = \frac{m\mu}{\bar{\mu}}\partial_\mu J(x,t) - m^2\widetilde{U}_\mu(x,t).$$

F. Infinite volume limit of the Schwinger model

We first consider the fermion Wightman functions for $L,\sigma < \infty$. Let

$$W_{L,\sigma}(x_1,\dots,x_n;y_1,\dots,y_n) \equiv \langle \Psi_{i_1}^*(x_1)\cdots\Psi_{i_n}^*(x_n)\Psi_{j_1}(y_1)\cdots\Psi_{j_n}(y_n)\rangle_{L,\sigma},$$

where we denote the cutoffs by the subscript. Then by the standard method (Ref. 8, see also the appendix), we get

$$W_{L,\sigma} = [Z(L,\sigma)]^{-n}\exp\left[\frac{e^2}{\bar{\mu}^2}F_{L,\sigma}(x_1,\dots,x_n;y_1,\dots,y_n)\right] \times \langle\psi_{i_1}^*(x_1)\dots\psi_{i_n}^*(x_n)\psi_{j_1}(y_1)\dots\psi_{j_n}(y_n)\rangle, \tag{6.34}$$

where i_l and j_l are 1 or 2 and

$$F_{L,\sigma} = \sum_{l,m}(\gamma_5)_{i_l}(\gamma_5)_{j_m}[\Delta_{L,\sigma}(x_l - y_m) - D_{L,\sigma}(x_l - y_m)] - \sum_{l<m}(\gamma_5)_{i_l}(\gamma_5)_{j_m}[\Delta_{L,\sigma}(x_l - x_m) - D_{L,\sigma}(x_l - x_m)] - \sum_{l<m}(\gamma_5)_{i_l}(\gamma_5)_{j_m}[\Delta_{L,\sigma}(y_l - y_m) - D_{L,\sigma}(y_l - y_m)]. \tag{6.35}$$

Here $(\gamma_5)_i = (\gamma_5)_{ii}$ and

$$\Delta_{L,\sigma}(x) = \frac{1}{L}\sum_{\substack{p\in\Gamma\\ |p|\leqslant\sigma,\ p\neq 0}}\frac{1}{2q_0}\exp[-iq_0t + ipx],$$
$$D_{L,\sigma}(x) = \frac{1}{L}\sum_{\substack{p\in\Gamma\\ |p|\leqslant\sigma,\ p\neq 0}}\frac{1}{2|p|}\exp[-i|p|t + ipx]. \tag{6.36}$$

Since

$$Z(L,\sigma)^{-1} = \exp\left(-\frac{e^2}{\bar{\mu}^2}[\Delta_{L,\sigma}(0) - D_{L,\sigma}(0)]\right),$$

and since the number of x_l such that $(\gamma_5)_{i_l} = 1$ is equal to that of y_l such that $(\gamma_5)_{j_l} = 1$ (otherwise, the free fermion's Wightman function in W vanishes)

$$W_{L,\sigma} = Z(L,\sigma,\kappa)^{-n}\exp\left(\frac{e^2}{\bar{\mu}^2}F_{L,\sigma,\kappa}(x_1,\dots,x_n;y_1,\dots,y_n)\right) \times(\text{free fermion's Wightman function}),$$

where $F_{L,\sigma,\kappa}$ is given by replacing $D_{L,\sigma}(x)$ in $F_{L,s}$ by

$$D_{L,\sigma,\kappa}(x) = D_{L,\sigma}(x) - D_{L,\kappa}(0), \tag{6.37}$$

and

$$Z(L,\sigma,\kappa)^{-1} = \exp\left[-\frac{e^2}{\bar{\mu}^2}[\Delta_{L,\sigma}(0) - D_{L,\sigma,\kappa}(0)]\right]. \tag{6.38}$$

For fixed (t,x) such that $t^2 - x^2 \neq 0$

$$D_{L,\sigma,\kappa}(x) \to D_\kappa(x) = \frac{1}{2\pi}\int\frac{dp}{2|p|}[e^{-ipx} - \theta(\kappa - |p|)],$$
$$\Delta_{L,\sigma}(x) \to \Delta(x) = \frac{1}{2\pi}\int\frac{dq}{2q_0}[e^{-iqx}], \tag{6.39}$$
$$Z(L,\sigma,\kappa)^{-1} \to Z(\kappa) = \exp\left(-\frac{e^2}{2\pi\bar{\mu}^2}\log\frac{2\kappa}{\bar{\mu}}\right) = \left(\frac{2\kappa}{\bar{\mu}}\right)^{-m^2/2\bar{\mu}^2}$$

as $L,\sigma,\to\infty$, where $px = |p|t - px$ and $qx = q_0t - qx$. Thus $W_{L,\sigma}$ uniformly converges to

$$Z(\kappa)^{-n}\exp[(e^2/\bar{\mu}^2)F_\kappa(x_1,\dots x_n;y_1,\dots,y_n)] \times(\text{free fermion's Wightman function}) \tag{6.40}$$

as $L,\sigma\to\infty$ provided that $(x_l - x_m)^2 \neq 0$, $(y_l - y_m)^2 \neq 0$ for $l\neq m$ and $(x_l - y_m)^2 \neq 0$ for any l and m, where F_κ is defined by $\lim_{\kappa\to\infty}F_{L,\sigma,\kappa}$. Here κ is an artificial infrared cutoff (due to Klaiber[8]) and these Wightman functions themselves do not depend on κ.

These Wightman functions are equivalent to those of Klaiber[8] by setting $a = \alpha_2^2 - 2\sqrt{\pi}\alpha_2 = 0$, $b = \alpha_3^2 - 2\sqrt{\pi}\alpha_3 = -e^2/\bar{\mu}^2$ in his calculation except $Z(\kappa)$ and Δ. The reason why "a" vanishes is that there is a cancellation between $\alpha_2^2 - 2\sqrt{\pi}\alpha_2 = e^2/\mu^2$ and $-\alpha_1^2\mu^2 = -e^2/\mu^2$. Since $-\pi < b \leqslant 0$ for $\mu > 0$, and since $b = -\pi$ for $\mu = 0$, these Wightman functions have the cluster property for $\mu > 0$, and not for $\mu = 0$.

Next we consider the Wightman functions of $\Psi_i^p = P_{\rm phys}\Psi_i P_{\rm phys}$, which are considered as the expectation values of the state $\rho_{L,\sigma{\rm phys}} = \langle\Omega,\pi_{B,{\rm phys}}(..)\Omega\rangle$, where $\pi_{B,{\rm phys}}$

is defined by replacing $\pi_B(\Phi)$ by $P_{\rm phys}\ \pi_B(\Phi)P_{\rm phys}$ with Φ any field operator. we denote them $\{W^p_{L,\sigma}\}$. By an easy calculation we have

$$W^p_{L,\sigma} = Z(L,\sigma)^{-n} \exp[(e^2/\bar{\mu}^2)F_{L,\sigma}(x_1,..,x_n;y_1,..,y_n) + (e^2/\mu_2)F^B_{L,\sigma}(x_1,...,x_n;y_1,...,y_n)] \times(\text{free fermion's Wightman function}), \quad (6.41)$$

where

$$F^B_{L,\sigma}(x_1,...,x_n;y_1,...,y_n) = -\sum_{i<j}[D_{L,\sigma}(x_i-x_j)+D_{L,\sigma}(y_i-y_j)] + \sum D_{L,\sigma}(x_i-y_j). \quad (6.42)$$

Replacing $D_{L,\sigma}(x)$ by $D_{L,\sigma,\kappa}(x) = D_{L,\sigma}(x) - D_{L,\kappa}(0)$ in $F_{L,\sigma}$ and $F^B_{L,\sigma}$ as before, and denoting the resultant functions as $F_{L,\sigma,\kappa}$ and $F^B_{L,\sigma,\kappa}$ we finally have

$$W^p_{L,\sigma} = Z(L,\sigma,\kappa)^{-n}\exp\left(\frac{ne^2}{\mu^2}D_{L,\kappa}(0)\right) \times\exp\left(\frac{e^2}{\bar{\mu}^2}F_{L,\sigma,\kappa}+\frac{e^2}{\mu^2}F^B_{L,\sigma,\kappa}\right) \times(\text{free fermion's Wightman function}).$$

Since $D_{L,\kappa}(0)\to C\log L\kappa$ (C:positive constant) as $L\to\infty$, $W^p_{L,\sigma}$ do not converge for $L,\sigma\to\infty$. Instead consider

$$\exp\left(-\frac{e^2}{2\mu^2}D_{L,\kappa}(0)\right)\Psi^p_i(x) = \exp\left(\frac{e^2}{2\mu^2}D_{L,\sigma,\kappa}(0)\right)\exp[i\chi^p_{i,\sigma}(x)]\psi_i(x).$$

Then the corresponding Wightman functions uniformly converge to (nontrivial) Lorentz covariant distribution functions which satisfy the clustering for $\mu>0$.

Remark: The Wightman functions for $\exp[i\chi^p_{i,\sigma}]\psi_i$ themselves converge to zero as $L,\sigma\to\infty$. This is due to the fact that the present model is not super-renormalizable if the Proca formalism is used. Moreover the so-called short-distance behavior of W^p is different from that of W.

The Wightman functions obviously converge uniformly even if U_μ, j_μ, and $\partial_\mu B$ are included. Then we get:

Theorem 6.11: Both the Lorentz states $\{\rho_{L,\sigma}\}$ and the physical states $\{\rho_{L,\sigma,\rm phys}\}$ uniformly converge on the field and observable algebras, respectively. The limiting states satisfy the clustering for $\mu>0$.

For the physical state, we can apply the Wightman reconstruction theorem.[8,41] Further even for the limiting Lorentz state, we can obtain a representation of the field algebra. This is done by mimicking the Klaiber's method. To write down the representation, we introduce field operators $J^{(\pm)}_\kappa(x)$, $\tilde{J}^{(\pm)}_\kappa(x)$, and $B^{(\pm)}_\kappa(x)$ with an infrared cutoff κ as follows:

$$J^{(-)}_\kappa(x) = \int\frac{dp}{\sqrt{2\pi 2|p|}}\,iA(p)[e^{-ipx}-\theta(\kappa-|p|)],$$

$J^{(+)}=[J^{(-)}]^*$, and $\tilde{J}^{(\pm)}_\kappa$ and $B^{(\pm)}_\kappa$ are similarly defined. Further we define

$$d^\pm_\kappa(x) = \mp\frac{i}{2\pi}\int_{-\kappa}^{\kappa}\frac{dp}{2|p|}[e^{\pm ipx}-1],$$

$$\tilde{d}^\pm_\kappa(x) = \mp\frac{i}{2\pi}\int_{-\kappa}^{\kappa}\frac{dp}{2|p|}\epsilon(p)[e^{\pm ipx}-1].$$

Thus we define

$$\Psi_i(x) = C(\kappa)\exp[i\chi^{(+)}_i(x)]\psi_i(x)\exp[i\chi^{(-)}_i(x)],$$

where

$$\chi^{(\pm)}_i(x) = \alpha_1 B^{(\pm)}_\kappa(x) + \alpha_2 J^{(\pm)}_\kappa(x) + (\gamma_5)_{ii} \times[\alpha_3\tilde{J}^{(\pm)}_\kappa(x)+\alpha_4\tilde{\phi}^{(\pm)}(x)] + \alpha_1^2\mu^2 Q d^\pm_\kappa(x) + \alpha_2 \times(\sqrt{\pi}-\alpha_2)Qd^\pm_\kappa(x) + \alpha_3(\sqrt{\pi}-\alpha^2)\tilde{Q}d^\pm_\kappa(x) + (\gamma_5)_{ii} \times\left[\alpha_2(\sqrt{\pi}-\alpha_3)\tilde{Q}d^\pm_\kappa(x)+\alpha_3(\sqrt{\pi}-\alpha_3)Q\tilde{d}^\pm_\kappa(x)\right].$$

Here $Q=\Lambda_1+\Lambda_2$, $\tilde{Q}=\Lambda_1-\Lambda_2$, and $C(\kappa)$ is a finite constant. Though χ_i include noncovariant terms $d^\pm_\kappa$ and $\tilde{d}^\pm_\kappa$, they cancel each other when we calculate the Wightman functions, and recover the Wightman functions already calculated.

Set $\alpha_1=0$ in χ_i. Then we get the operator solution of the positive metric formalism.[41]

7. MORE ABOUT THE RENORMALIZED HAMILTONIAN

When $M\neq 0$, what shall we do? We turn to the renormalized Hamiltonian $H_R(L,\sigma)$. It seems impossible to obtain Ω_B or π_B explicitly in this case. But for the moment assume that all the (cutoff) Wightman functions are obtained:

$$W_{L,\sigma}(x_1,...,x_n) = \langle\Omega(L,\sigma),\Phi_1(x_1)\cdots\Phi_n(x_n)\Omega(L,\sigma)\rangle,$$

where $\Phi_i(x_i)$ are arbitrary field opeators and $\Omega(L,\sigma)$ is the vacuum vector (if it exists) in $\mathscr{F}$ of the renormalized Hamiltonian.

Thus we apply the reconstruction theorem (Conjecture 6.1). Further if (iii) and (iv) also hold as well as the conjecture, the cyclic vector Ω in the conjecture will be the physical vacuum in $\mathscr{H}_{\rm phys}$. Since the renormalization ensures the conservation of the current, we expect that the gaugeon field B is represented as a free boson field with an indefinite metric on the new Hilbert space $\mathscr{H}$. Thus $H_R(L,\sigma)$ is represented as a self-adjoint operator in which the gaugeon part is separated as a free Hamiltonian of the gaugeon field B.

But how can we obtain $\Omega(L,\sigma)$ or $W_{L,\sigma}$? In fact even if $H_R(L,\sigma)$ is Θ-self-adjoint $\Omega(L,\sigma)$ is sometimes outside the Fock space, and further even if $H_R(L,\sigma)$ is Θ-self-adjoint, it may not have eigenvectors of real eigenvalues.

One possible method is to use the euclidean method to calculate $S_{L,\sigma}(x_1,...,x_n)$ which are the analytic continuation of the Wightman functions $W_{L,\sigma}(x_1,...,x_n)$ to the euclidean region (the so-called Schwinger functions). Since the present model includes the fermion fields, we will be obliged to use the semi-Euclidean method owing to Seiler[Ref. 42; see also Refs. 26–29] or use the Euclidean fermion fields owing to Osterwalder and Schrader.[43,44] Further we use the Euclidean Markov vector field of the Stückelberg formalism.[13]

Since the indefinite metric is by-passed in the euclidean region, the analysis may be rather easy. At the present stage, however, the author can say nothing about this approach, but this will be discussed elsewhere [Ref. 23; see also Refs.

45,46].

Remark: We can prove that the gaugeon part decouples from the matter fields in $H_R(L,\sigma)$ as the free Hamiltonian by suitable Θ-isometric transformation (see next section). This fact may imply that spec $H_R(L,\sigma)\subset R$ and $\Omega(L,\sigma)\in\mathscr{F}$ (possibly for sufficiently small $|e|$).

8. MASS-SHIFT TRANSFORMATION AND SINE-GORDON MODELS

We consider a Bogolyubov transformation which changes the bare mass of the fermion and apply it to the renormalized Hamiltonian. This technique is familiar in the $P(\phi)_2$ models.[35,36] Here this technique shows that the $(\mathrm{QED})_2$-models are related to the (quantum) sine-Gordon models.[47–49]

A. Mass-shift transformation of fermion

As before, we consider the theory in a periodic box Λ, and let κ be the sharp momentum cutoff (M denotes the mass) of ψ_M. Consider $V(L,\kappa)$ which implements

$$V(L,\kappa)\psi_{M,\kappa}(x,0)V(L,\kappa)^{-1}=\psi_{M',\kappa}(x,0) \tag{8.1}$$

or equivalently which implements (for $|p|\leqslant\kappa$)

$$V(L,\kappa)c(p)V(L,\kappa)^{-1}=\cos\theta_p c(p)+\sin\theta_p d^*(-p), \tag{8.2}$$

$$V(L,\kappa)d^*(-P)V(L,\kappa)^{-1}$$
$$=-\sin\theta_p c(p)+\cos\theta_p d^*(-p),$$

where $\cos\theta_p=v(p)v'(p)+v(-p)v'(-p)$, $\sin\theta_p=-v(p)v'(-p)+v'(p)v(-p)$, and v' is the v-function [see Eq. (2.1)] for mass M'.

Lemma 8.1: Let $j^\mu_{(M,\kappa)}=:\bar\psi_{M,\kappa}(x,0)\gamma^\mu\psi_{M,\kappa}(x,0):$. Then

(i) $Vj^\mu_{(M,\kappa)}V^{-1}=j^\mu_{(M',\kappa)}$,

(ii) $V:\bar\psi_{M,\kappa}\psi_{M,\kappa}:V^{-1}=:\bar\psi_{M',\kappa}\psi_{M',\kappa}:+c_0$,

(iii) $V:j^\mu_{(M,\kappa)}j_{\mu(M,\kappa)}:V^{-1}=:j^\mu_{(M',\kappa)}j_{\mu(M',\kappa)}:+c_1$
$\times:\bar\psi_{M',\kappa}\psi_{M',\kappa}:+c_2$,

where the constants c_0,c_1, and c_2 diverge as $\kappa\to\infty$.

Proof: Let $\psi_{M,\kappa}(x,0)=\psi,\psi_{M',\kappa}(x,0)=\psi'$ ($\hat\psi$ and $\hat\psi'$ for the Fourier components). Note that: $\hat\psi_i^*(p)\hat\psi_j(p')$: $=\hat\psi_i^*(p)\hat\psi_j(p')-(\hat\psi_i^*(p)\hat\psi_j(p'))=\hat\psi_i^*(p)\hat\psi_j(p')-(L/2\pi)\times\delta_{p,p'}v_i(p)v_j(p')$. Thus $V:\bar\psi\Gamma\psi:V^{-1}=:\bar\psi'\Gamma\psi':+(1/L)\times\Sigma_{|p|\leqslant\kappa}[(v'(p),\gamma^0\Gamma v'(p))-(v(p),\gamma^0\Gamma v(p))]$, where the last constant is zero for $\Gamma=\gamma^\mu$ and $(L)^{-1}\Sigma_{|p|<\kappa}[M/\omega(p)-M'/\omega'(p)]\equiv c_0$ for $\Gamma=1$. This proves (i) and (ii).

Since $:\bar\psi\gamma^\mu\psi:=\bar\psi\gamma^\mu\psi$-constant being independent of M, it suffices to consider how

$$:(\bar\psi\gamma^\mu\psi)(\bar\psi\gamma_\mu\psi):=4:\psi_1^*\psi_1\psi_2^*\psi_2:$$
$$=4[\psi_1^*\psi_1\psi_2^*\psi_2-(\psi_1^*\psi_1)\psi_2^*\psi_2$$
$$-(\psi_2^*\psi_2)\psi_1^*\psi_1-(\psi_1^*\psi_2)\psi_1^*\psi_2$$
$$+(\psi_1^*\psi_2)\psi_2^*\psi_1$$
$$+(\psi_1^*\psi_1)(\psi_2^*\psi_2)-(\psi_1^*\psi_2)(\psi_2^*\psi_1)]$$

transforms under the operation of V. Since

$$(\psi_1^*\psi_1)=(\psi_2^*\psi_2)=\frac{1}{L}\sum_{|p|<\kappa}\frac{\omega+p}{2\omega}=\frac{1}{2L}\sum_{|p|<\kappa}1\equiv\Delta_1(\kappa),$$

$$-(\psi_1\psi_2^*)=(\psi_1^*\psi_2)=-\frac{M}{L}\sum_{|p|<\kappa}\frac{1}{2\omega}\equiv-\Delta_2(\kappa,M),$$

we have

$$:\psi_1^*\psi_1\psi_2^*\psi_2:=\psi_1^*\psi_1\psi_2^*\psi_2-\Delta_1(\kappa)(\psi_1^*\psi_1+\psi_2^*\psi_2)$$
$$-\Delta_2(\kappa,M)(\psi_1^*\psi_1+\psi_2^*\psi_2)+\Delta_1^2(\kappa)-\Delta_2^2(\kappa,M),$$

which proves

$$V:\psi_1^*\psi_1\psi_2^*\psi_2:V^{-1}=:\psi_1'^*\psi_1'\psi_2'^*\psi_2':$$
$$+[\Delta_2(\kappa,M')-\Delta_2(\kappa,M')]:\bar\psi'\psi':$$
$$-(\Delta_2(\kappa,M)-\Delta_2(\kappa,M'))^2. \qquad \square$$

Let $H_{0:L}(\psi_M)$ be the free Hamiltonian of the fermion of mass M and let

$$H_{0:L}(M,\delta M;\kappa)=H_{0:L}(\psi_M)+M\int_\Lambda:\bar\psi_{M,\kappa}(x,0)\psi_{M,\kappa}(x,0):dx.$$

This Hamiltonian is diagonalized by $V(L,\kappa)$ which implements $V(L,\kappa)\psi_{M,\kappa}V(L,\kappa)^{-1}=\psi_{M+\delta M,\kappa}$. In fact

$$V[H_{0:L}(M,\delta M;\kappa)]V^{-1}$$
$$=\frac{2\pi}{L}\Big\{\sum_{|p|>\kappa}\omega(p)+\sum_{|p|<\kappa}\omega'(p)\Big\}[c^*(p)c(p)$$
$$+d^*(p)d(p)]+E(L,M,\delta M;\kappa),$$

where $\omega'(p)=(p^2+M'^2)^{1/2}$ with $M'=M+\delta M$ and

$$E(L,M,\delta M;\kappa)=-2\sum_{0<p<\kappa}\Big[\omega(p)-\omega'(p)+\frac{M\delta M}{\omega(p)}\Big].$$

Note $\cos\theta_p=1+0(1)(\delta M)^2p^{-2}+0(p^{-4})$ as $|p|\to\infty$. Then $\{\theta_p\}\in l^2(\Gamma)$ in two dimensions which implies[17–21,50]

Theorem 8.2: $V(L,\kappa)$ is unitary for $\kappa\leqslant\infty$ provided $L<\infty$. Especially $V(L)=V(L,\infty)$ is unitary though $E(L,M,\delta M;\kappa)$ diverges as $\kappa\to\infty$.

B. Pre-sine-Gordonization

In the following, $\psi(x,t)$ implies the massless free fermion field. Then

$$\sigma_{L,\kappa}(x)\equiv\psi_{1,\kappa}^*\psi_{2,\kappa}(x):=\psi_{1,\kappa}^*(x)\psi_{2,\kappa}(x). \tag{8.3}$$

The operator $\sigma_{L,\kappa}(x)$ cannot be any operator (even if suitably smeared) in the limit of $\kappa\to\infty$. But the following[51,52]

$$\langle\sigma_{L,\kappa}^*(x_1)\cdots\sigma_{L,\kappa}^*(x_n)\sigma_{L,\kappa}(y_1)\cdots\sigma_{L,\kappa}(y_n)\rangle$$
$$=\Big[\sum_\pi\mathrm{sgn}(\pi)\prod_{i=1}^n\langle\psi_{2,\kappa}^*(x_i)\psi_{2,\kappa}(y_{\pi(i)})\rangle\Big]$$
$$\times\Big[\sum_\pi\mathrm{sgn}(\pi)\prod_{j=1}^n\langle\psi_{1,\kappa}^*(x_j)\psi_{1,\kappa}(u_{\pi(j)})\rangle\Big],$$

where we used $\{\psi_1,\psi_2^*\}=0$ and $\{\psi_i,\psi_j\}=0$, and π means permutation of $(1,\ldots,n)$. Thus in the limit of $L,\kappa\to\infty$,

$$\text{l.h.s.}=(4\pi^2)^{-n}\frac{\Pi_{i<j}[-(x_i-x_j)^2]\Pi_{k<l}[-(y_k-y_l)^2]}{\Pi_{i,j}[-(x_i-y_j)^2+\epsilon]}, \tag{8.4}$$

where $x^2=x_\mu x^\mu=x_0^2-x_1^2$ and ϵ is an infinitesimally small positive constant.

Our idea is essentially due to Coleman,[53,54] and is summarized in Table I.

TABLE I. Transformations of various Hamiltonians.

	$H_{S.G}(L,\sigma,\kappa;M) \equiv H_0^S(,\sigma) + M\int U(L,\sigma):\bar{\psi}_\kappa\psi_\kappa:U(L,\sigma)^{-1}\,dx - E(L,\sigma,\kappa)$		
	$\downarrow$	$\downarrow$	$U(L,\sigma)$
$\kappa\to\infty$	$\downarrow$	$H_S^R(L,\sigma) + M\int:\bar{\psi}_\kappa\psi_\kappa:dx - E(L,\sigma,\kappa)$	
	$\downarrow$	$\downarrow$	$V(L,\kappa),\kappa\to\infty$
	$H_{S.G}(L,\sigma;M)$ $\longleftrightarrow$	$H_R(\psi_M;L,\sigma)$	
$L,\sigma\to\infty$	$\downarrow$	$\downarrow$	$L,\sigma\to\infty$
	$H_{S.G}(M)$ $\longleftrightarrow$	$H_R(\psi_M;\infty,\infty)$	

In this table, $H_R(\psi_M;L,\sigma)$ is $H_R(L,\sigma)$ with the bare fermion mass M, and $E(L,\sigma,\kappa)$ is a constant (if possible) chosen so that inf spec $H_{\rm S.G}(L,\sigma,\kappa;M)=0$.

Let

$$\Sigma_{L,\sigma,\kappa}(x,0) = U(L,\sigma)\sigma_{L,\kappa}(x,0)U^{-1}(L,\sigma). \tag{8.5}$$

Then

$$\lim_{\kappa\to\infty}\Sigma_{L,\sigma,\kappa}(x,0) = \Psi_1^*(x,0)\Psi_2(x,0)$$
$$= Z'(L,\sigma)^{-1}\exp[2i\alpha_3\widetilde{J}_\sigma^{\prime(+)}(x) + 2i\alpha_4\tilde{\phi}_\sigma^{(+)}(x)]\sigma_L(x)\exp[(+)\to(-)],$$

where

$$Z'(L,\sigma)^{-1} = \exp[-2(\alpha_3^2 - 2\sqrt{\pi}\alpha_3)D_{L,\sigma}(0) - 2\alpha_4^2\Delta_{L,\sigma}(0)]$$

and $\sigma_L = \lim_{\kappa\to\infty}\sigma_{L,\kappa}$. Thus we get for $\Sigma_{L,\sigma} = \lim_{\kappa\to\infty}\Sigma_{L,\sigma,\kappa}$,

$$\langle\Sigma_{L,\sigma}^*(x_1)\cdots\Sigma_{L,\sigma}^*(x_n)\Sigma_{L,\sigma}(y_1)\cdots\Sigma_{L,\sigma}(y_n)\rangle = Z'(L,\sigma)^{-2n}\exp[F'_{L,\sigma}] \times\langle\sigma_L^*(x_1)\cdots\sigma_L^*(x_n)\sigma_L(y_1)\cdots\sigma_L(y_n)\rangle, \tag{8.6}$$

$$F'_{L,\sigma} = \sum_{i,j}4\{[\alpha_3^2 - 2\sqrt{\pi}\alpha_3]D_{L,\sigma}(x_i - y_j) - \alpha_4^2\Delta_{L,\sigma}(x_i - y_j)\}$$
$$- \sum_{i<j}4\{[\alpha_3^2 - 2\sqrt{\pi}\alpha_3]D_{L,\sigma}(x_i - x_j) - \alpha_4^2\Delta_{L,\sigma}(x_i - x_j)\}$$
$$- \sum_{i<j}4\{[\alpha_3^2 - 2\sqrt{\pi}\alpha_3]D_{L,\sigma}\times(y_i - y_j) - \alpha_4^2\Delta_{L,\sigma}(y_i - y_j)\}. \tag{8.7}$$

As before, we rewrite as follows:

$$Z'(L,\sigma,\kappa)^{-2n}\exp[F'_{L,\sigma,\kappa}]\langle\sigma_L^*(x_1)\cdots\sigma_L^*(x_n)\sigma_L(y_1)\cdots\sigma_L(y_n)\rangle, \tag{8.8}$$

where

$$Z'(L,\sigma,\kappa)^{-1} = \exp\{-2[\alpha_3^2 - 2\sqrt{\pi}\,\alpha_3]D_{L,\sigma,\kappa}(0) - 2\alpha_4^2\Delta_{L,\sigma}(0)\} \tag{8.9}$$

and $F'_{L,\sigma,\kappa}$ is given by replacing $D_{L,\sigma}$ by $D_{L,\sigma,\kappa}$ in $F'_{L,\kappa}$. Since $\alpha_4^2 = -\alpha_3^2 + 2(\pi)^{1/2}\alpha_3$ in the (QED)$_2$-models, $Z'(L,\sigma,\kappa)^{-1}$ converges to

$$Z(\kappa)^{-2} = (2\kappa/\tilde{\mu})^{-m^2/\tilde{\mu}^2} \tag{8.10}$$

as $L,\sigma\to\infty$. Since $\Sigma_{L,\sigma}$ does not contain the gaugeon field B'_σ, we see that B'_σ decouples from the Hamiltonian by this transformation which we call "pre-sine-Gordonization."

C. (QED)$_2$ and the sine-Gordon equations

Let φ have a mass $\tilde{\mu}$ and let $\widetilde{\Sigma} = :\exp[i(\beta_1)^{1/2}\varphi + i(\beta_2)^{1/2}\tilde{\phi}]:$. Then

$$K^{m+n}\langle\widetilde{\Sigma}(x_1)\cdots\widetilde{\Sigma}(x_m)\widetilde{\Sigma}^*(y_1)\cdots\widetilde{\Sigma}^*(y_n)\rangle$$
$$= K^{m+n}\exp\Big\{\sum_{i,j}[\beta_1\widetilde{\Delta}(x_i - y_j) + \beta_2\Delta(x_i - y_j)] - \sum_{i<j}[\beta_1\widetilde{\Delta}(x_i - x_j) + \beta_2\Delta(x_i - x_j)] - \sum_{i<j}[\beta_1\widetilde{\Delta}(y_i - y_j) + \beta_2\Delta(y_i - y_j)]\Big\},$$

where $\widetilde{\Delta}$ is the Δ-function of mass $\tilde{\mu}$. Note

$$\widetilde{\Delta}(x) = D_\kappa(x) + (1/4\pi)\log(2\kappa/\tilde{\mu})^2 + O(\tilde{\mu}^2x^2) \quad \text{as} \quad \tilde{\mu}\to 0,$$
$$D_\kappa(x) = -(1/4\pi)\log(e^\gamma\kappa)^2|x^2|$$

($\gamma = 0.57\ldots$, Euler's constant).

Using the first equation, we get

$$\text{l.h.s.} = K^{m+n}\exp\{-(\beta_1/4\pi)[(m-n)^2/2 - (m+n)]\log(2\kappa/\tilde{\mu})^2 + \{\widetilde{\Delta}\to D_\kappa\} + O(\tilde{\mu}^2) + \Delta - \text{terms})\}.$$

Thus if we set $K = \delta(\tilde{\mu}/2\kappa)^{\beta_1/8\pi}$,

$$\lim_{\tilde{\mu}\to 0}K^{m+n}\langle\widetilde{\Sigma}(x_1)\cdots\widetilde{\Sigma}(x_m)\widetilde{\Sigma}^*(y_1)\cdots\widetilde{\Sigma}^*(y_n)\rangle$$
$$= 0 \quad \text{for } m\neq n,$$
$$= (\delta)^{2n}\exp\Big\{\beta_1\Big[\sum D_\kappa(x_i - y_j) - \sum\{D_\kappa(x_i - x_j) + D_\kappa(y_i - y_j)\}\Big] + \beta_2\Big[\sum\Delta(x_i - x_j) - \sum\{\Delta(x_i - x_j) + \Delta(y_i - y_j)\}\Big]\Big\}$$

for $m = n$. Comparing these expressions with those of Σ, we see that these expressions are equivalent to each other if we set

$$\beta_1 = 4(\alpha_3 - \sqrt{\pi})^2 = 4\pi(\mu^2/\tilde{\mu}^2),$$
$$\beta_2 = 4\alpha_4^2 = 4\pi(m^2/\tilde{\mu}^2), \tag{8.11}$$

$\delta \equiv \delta(\kappa) = Z(\kappa)^{-2}(e^{2\gamma}\kappa^2/4\pi^2)^{1/2} = (\bar{\mu}/2\kappa)^{m^2/\bar{\mu}^2}(e^\gamma\kappa/2\pi)$.

Thus $\beta_1 + \beta_2 = 4\pi$ (May's critical point[55]) in the present QED-type models, and $\lim_{\mu\to 0}\delta = (m/4\pi)e^\gamma$ (independent of κ).

Let $:\,:_\kappa$ be an artificial Wick ordering such that the D_κ functions arise from the Wick contraction. Then we formally have

$$M\Sigma(x) = M\delta(\kappa)\colon \exp\left[i\sqrt{\beta_1}\,\varphi + i\sqrt{\beta_2}\,\tilde{\phi}\,\right]\colon_\kappa$$

on the charge zero (i.e., vacuum) sector, and we also have

$$M\,[\Sigma(x) + \Sigma^*(x)] = 2M\delta(\kappa)\colon \cos\left[\sqrt{\beta_1}\,\varphi + \sqrt{\beta_2}\,\tilde{\phi}\,\right]\colon_\kappa$$

on the vacuum sector. This is nothing but the interacting Hamiltonian density of the two-component sine-Gordon model.

Let $V(x) = z\colon\exp[i(\beta_1)^{1/2}\varphi + i(\beta_2)^{1/2}\tilde{\phi}\,]\colon_\kappa$ with $z = M\delta(\kappa)$. Then we have

$$\begin{aligned}\Xi_n&(x_1,\ldots,x_n;y_1,\ldots,y_n)\\ &\equiv\langle V(x_1)\cdots V(x_n)V^*(y_1)\cdots V^*(y_n)\rangle\\ &= z^{2n}\exp\Big[\beta_1\Big[\sum_{i,j}D_\kappa(x_i-y_j)\\ &\quad-\sum_{i<j}\{D_\kappa(x_i-x_j)+D_\kappa(y_i-y_j)\}\Big]\\ &\quad+\beta_2\Big[\sum_{i,j}\Delta(x_i-y_j)\\ &\quad-\sum_{i<j}\{\Delta(x_i-x_j)+\Delta(y_i-y_j)\}\Big]\Big].\end{aligned}\tag{8.12}$$

Fröhlich noted that, in the Euclidean region, this is nothing but the generalized partition function of $2n$-particles with n (+)-charged and n (−)-charged particles whose interactions are defined by the potentials D_κ (Coulomb potential in two dimensions) and Δ (Yukawa potential) with inverse temperatures β_1 and β_2, and by the fugacity z [$\equiv M\delta(\kappa)$]. Then he could prove the existence of the (quantum) sine-Gordon model for sufficiently small (β_1,β_2) (he only considered the case that $\beta_1 = 0$ or $\beta_2 = 0$[47,52,56,57]) using known results in the study of "stability of matters consisting of electrons and nucleons".[55,58–60]

In the present model, however, β_1 and β_2 are somewhat large : $\beta_1 + \beta_2 = 4\pi$, and their methods cannot directly apply without nontrivial boson-mass renormalization (the model is again super-renormalizable). In other words, the sine-Gordonized $(\mathrm{QED})_2$-models lie on the May's critical point at which the system collapses to one point by Yukawa–Coulomb force.

Though we have already renormalized the Hamiltonian of $(\mathrm{QED})_2$, the Hamiltonian of sine-Gordonized $(\mathrm{QED})_2$ seems to require the nontrivial renormalization, which is very pathological. The author does not know whether or not the sine-Gordonized $(\mathrm{QED})_2$-model exists (after the renormalization if necessary) and coincides with the $(\mathrm{QED})_2$-model on the vacuum sector. This will be discussed elsewhere.

ACKNOWLEDGMENTS

The author would like to thank Professor H. Araki and Professor N. Nakanishi for their patient guidance and encouragement. He also thanks Dr. A. Kishimoto for helpful discussions, and Professor J. Fröhlich who kindly sent his works prior to thier publication.

APPENDIX : CONSTRUCTION OF $U(L,\sigma)$ AND Ψ

We first obtain how $\{a_i^{(\pm)}(\pm p)\}$ transform by the Θ-isometric operators U_i, V_{ij}, and W_{ij}. This is easy and left to the reader. For example:

$$U_i a_i^{(\pm)} U_i^{-1} = (\cosh\theta_i \pm \sinh\theta_i)a_i^{(\pm)}, \quad i = 1,2,$$
$$V_{12}a_1^{(+)}(\pm p)V_{12}^{-1} = \cosh\tau_3 a_1^{(+)}(\pm p) - \sinh\tau_3 a_2^{(-)}(\pm p),$$
$$V_{23}a_2^{(+)}(\pm p)V_{23}^{-1} = \cos\tau_1 a_2^{(+)}(\pm p) + \sin\tau_1 a_3^{(+)}(\pm p), \text{ etc.}$$

Let $U = UVW$ as in Sec. (6.5). We calculate the angles $\{\theta_i, \tau_i, \varphi_i\}_{i=1}^3$ such that (1) $U(p)\mathscr{A}(p)U^{-1}(p) = S(p)\mathscr{A}(p)$, (2) $\theta_i(m=0) = \tau_i(m=0) = \varphi_i(m=0) = 0$, and (3) they are holomorphic in a neighborhood of $m = 0$. And finally we calculate $\Psi_i(x,0) = U(L,\sigma)\,\psi_i(x,0)U(L,\sigma)^{-1}$.

First note that

$$\begin{aligned}&[V_{23}V_{31}W]a_1^{(\pm)}(p)[V_{23}V_{31}W]^{-1}\\ &= [\cos\tau_2\cosh\varphi_2\cos\varphi_3 + \cos\tau_2\sinh\varphi_1\sinh\varphi_2\cos\varphi_3\\ &\quad\pm\sin\tau_2\cosh\varphi_1\sinh\varphi_2\,]a_1^{(\pm)}(p) + [\cos\tau_1\cosh\varphi_2\sin\varphi_3\\ &\quad-\cos\tau_1\sinh\varphi_1\sinh\varphi_2\cos\varphi_3\\ &\quad\pm\sin\tau_1\cos\tau_2\cosh\varphi_1\sinh\varphi_2\\ &\quad+\sin\tau_1\sin\tau_2(\cosh\varphi_2\cos\varphi_3\\ &\quad+\sinh\varphi_1\sinh\varphi_2\sin\varphi_3)]a_2^{(\mp)}(p)\\ &\quad+\{-\sin\tau_1\cos\tau_2\cosh\varphi_1\sinh\varphi_2\\ &\quad\pm[\cos\tau_1\sin\tau_2(\cosh\varphi_2\cos\varphi_3\\ &\quad+\sinh\varphi_1\sinh\varphi_2\sin\varphi_3) - \sin\tau_1(\cosh\varphi_2\sin\varphi_3\\ &\quad-\sinh\varphi_1\sinh\varphi_2\cos\varphi_3)]\}a_3^{(\pm)}(p).\end{aligned}$$

Next we shall operate $V_{12}(\cdots)V_{12}^{-1}$. Since this operator mixes only $a_1^{(\pm)}$ and $a_2^{(\pm)}$ and since U_i do not mix these operators, the coefficients of $a_3^{(\pm)}$ must vanish because $Ua_1^{(\pm)}U^{-1}$ do not contain $a_3^{(\pm)}$. Thus

$$\cos\tau_1\cos\tau_2\cosh\varphi_1\sinh\varphi_2 = 0,$$
$$\cos\tau_1\sin\tau_2\cos\varphi_3 - \sin\tau_1\sin\varphi_3 = 0,$$

for any $m = e/(\pi)^{1/2}$, which implies

$$\varphi_2 = 0, \quad \sin\tau_2 = \tan\tau_1\tan\varphi_3.$$

In this case, $[VW]a_+^{(\pm)}[VW]^{-1}$ equals

$$\begin{aligned}&[\cos\tau_2\cosh\tau_3\cos\varphi_3 \mp \sinh\tau_3(\cos\tau_1\sin\varphi_3\\ &\quad+\sin\tau_1\sin\tau_2\cos\varphi_3)]a_1^{(\pm)} + [\cosh\tau_3(\cos\tau_1\sin\varphi_3\\ &\quad+\sin\tau_1\sin\tau_2\cos\varphi_3)\mp\sinh\tau_3\cos\tau_2\cos\varphi_3\,]a_2^{(\mp)}.\end{aligned}$$

Further since $Ua_1^{(-)}U^{-1}$ does not contain $a_2^{(+)}$,

$$\begin{aligned}&\cosh\tau_3(\cos\tau_1\sin\varphi_3 + \sin\tau_1\sin\tau_2\cos\varphi_3)\\ &\quad+\sinh\tau_3\cos\tau_2\cos\varphi_3 = 0,\end{aligned}$$

i.e.,

$$\tanh\tau_3 = -\sin\tau_1\tan\tau_2 - \cos\tau_1\tan\varphi_3\sec\tau_2.$$

Defining $t_i = \exp(\theta_i)$ with $i = 1,2$, we finally get:

$$\begin{aligned}Ua_1^{(+)}U^{-1} = t_1\,&[\cos\tau_2\cosh\tau_3\cos\varphi_3 - \sinh\tau_3(\cos\tau_1\sin\varphi_3\\ &+\sin\tau_1\sin\tau_2\cos\varphi_3)]a_1^{(+)}\end{aligned}$$

TABLE II. Coefficients of $\{a_i^{(\pm)}(\pm p)\}$ in χ_i.

$a_1^{(+)}(p)$:	$t_1\left[\cosh\tau_3(-\sin\tau_2\sinh\varphi_1-\cos\tau_2\sin\varphi_3\cosh\varphi_1)-\sinh\tau_3\,[\cos\tau_1\cosh\varphi_1\cos\varphi_3+\sin\tau_1(\cos\tau_2\sinh\varphi_1-\sin\tau_2\sin\varphi_3\cosh\varphi_1)]\right]$
$a_1^{(-)}(p)$:	$t_1^{-1}\left[\cosh\tau_3(\mp\sin\tau_2\sinh\varphi_1\pm\cos\tau_2\sin\varphi_3\cosh\varphi_1)\pm\sinh\tau_3\{-\cos\tau_2\cosh\varphi_1\cos\varphi_3+\sin\tau_1(\cos\tau_2\sinh\varphi_1+\sin\tau_2\sin\varphi_3\cosh\varphi_1)\}\right]$
$a_1^{(+)}(-p)$:	$t_1\left[\cosh\tau_3(\sin\tau_2\sinh\varphi_1+\cos\tau_2\sin\varphi_3\cosh\varphi_1)+\sinh\tau_3\,[\cos\tau_1\cosh\varphi_1\cos\varphi_3+\sin\tau_1(\cos\tau_2\sinh\varphi_1-\sin\tau_2\sin\varphi_3\cosh\varphi_1)]\right]$
$a_1^{(-)}(-p)$:	$t_1^{-1}\left[\cosh\tau_3(\mp\sin\tau_2\sinh\varphi_1\pm\cos\tau_2\sin\varphi_3\cosh\varphi_1)\pm\sinh\tau_3\,[-\cos\tau_2\cosh\varphi_1\cos\varphi_3+\sin\tau_1(\cos\tau_2\sinh\varphi_1+\sin\tau_2\sin\varphi_3\cosh\varphi_1)]\right]$
$a_2^{(+)}(p)$:	$\pm(1-t_2\cos\tau_1\cosh\tau_3\cosh\varphi_1\cos\varphi_3)\pm t_2\cosh\tau_3\sin\tau_1(\cos\tau_2\sinh\varphi_1+\sin\tau_2\sin\varphi_3\cosh\varphi_1)\pm t_2\sinh\tau_3(\sin\tau_2\sinh\varphi_1-\cos\tau_2\sin\varphi_3\cosh\varphi_1)$
$a_2^{(-)}(p)$:	$-(1-t_2^{-1}\cos\tau_1\cosh\tau_3\cosh\varphi_1\cos\varphi_3)+t_2^{-1}\cosh\tau_3\sin\tau_1(\cos\tau_2\sinh\varphi_1+\sin\tau_2\sin\varphi_3\cosh\varphi_1)+t_2^{-1}\sinh\tau_3(\sin\tau_2\sinh\varphi_1+\cos\tau_2\sin\varphi_3\cosh\varphi_1)$
$a_2^{(+)}(-p)$:	$\mp(1-t_2\cos\tau_1\cosh\tau_3\cosh\varphi_1\cos\varphi_3)\mp t_2\cosh\tau_3\sin\tau_1(\cos\tau_2\sinh\varphi_1+\sin\tau_2\sin\varphi_3\cosh\varphi_1)+t_2\sinh\tau_3(\sin\tau_2\sinh\varphi_1-\cos\tau_2\sin\varphi_3\cosh\varphi_1)$
$a_2^{(-)}(-p)$:	$-(1-t_2^{-1}\cos\tau_1\cosh\tau_3\cosh\varphi_1\cos\varphi_3)+t_2\cosh\tau_3\sin\tau_1(\cos\tau_2\sinh\varphi_1-\sin\tau_2\sin\varphi_3\cosh\varphi_1)+t_2^{-1}\sinh\tau_3(\sin\tau_2\sinh\varphi_1+\cos\tau_2\sin\varphi_3\cosh\varphi_1)$
$a_3^{(+)}(p)$:	$\mp t_3\,[\cos\tau_1(\cos\tau_2\sinh\varphi_1+\sin\tau_2\sin\varphi_3\cosh\varphi_1)+\sin\tau_1\cosh\varphi_1\cos\varphi_3]$
$a_3^{(-)}(p)$:	$t_3^{-1}\,[\cos\tau_1(\mathrm{cs}\tau_2\sinh\varphi_1-\sin\tau_2\sin\varphi_3\cosh\varphi_1)-\sin\tau_1\cosh\varphi_1\cos\varphi_3]$
$a_3^{(+)}(-p)$:	$\pm t_3\,[\cos\tau_1(\cos\tau_2\sinh\varphi_1+\sin\tau_2\sin\varphi_3\cosh\varphi_1)+\sin\tau_1\cosh\varphi_1\cos\varphi_3]$
$a_3^{(-)}(-p)$:	$t_3^{-1}\,[\cos\tau_1(\cos\tau_2\sinh\varphi_1-\sin\tau_2\sin\varphi_3\cosh\varphi_1)+\sin\tau_1\cosh\varphi_1\cosh\varphi_3]$

$$-2\,t_2^{-1}\sinh\tau_3\cos\tau_2\cos\varphi_3\,a_2^{(-)},$$

$$Ua_1^{(-)}U^{-1}$$
$$=t_1^{-1}\,[\cos\tau_2\cosh\tau_3\cos\varphi_3+\sinh\tau_3(\cos\tau_1\sin\varphi_3+\sin\tau_1\sin\tau_2\cos\varphi_3)]a_1^{(-)}.$$

The coefficients of $Ua_i^{(\pm)}U^{-1}$ with $i=2$ and 3 are obtained in a similar way. By comparering $S(p)$ with these coefficients, we get the equations which the angles satisfy ($\varphi_2=0$):

$$\cosh\varphi_1=\frac{\cos\tau_1\cos\tau_2}{s_1},\quad \sinh\varphi_1=-\epsilon\frac{s_2}{s_1},$$
$$\cos\varphi_3=\epsilon\frac{\sin\tau_1}{s_2},\quad \sin\varphi_3=\frac{\cos\tau_1\sin\tau_2}{s_2},\tag{A1}$$
$$\cosh\tau_3=\epsilon\frac{\sin\tau_1\cos\tau_2}{s_3},\quad \sinh\tau_3=-\frac{\sin\tau_2}{s_3},$$

where $\epsilon=\mathrm{sgn}(m)=\mathrm{sgn}(e)$ and

$$s_1=(2\cos^2\tau_1\cos^2\tau_2-1)^{1/2},$$
$$s_2=(1-\cos^2\tau_1\cos^2\tau_2)^{1/2}=(1-s_1^2)^{1/2}/\sqrt{2},$$
$$s_3=(2\cos^2\tau_2-\cos^2\tau_2\cos^2\tau_2-1)^{1/2}.$$

Further defining $t_i=\exp[\theta_i]$ with $i=1,2$ and $t_3=\exp[-\theta_3]$,

$$\begin{aligned}
&(1)\quad s_1t_1=1,\\
&(2)\quad s_1s_3/s_2t_2=\bar{\mu}/\mu,\\
&(3)\quad t_2/s_1t_1=(p^0/q^0)^{1/2},\\
&(4)\quad 2s_2\cos\tau_1\cos\tau_2=|m|/\mu,\\
&(5)\quad \frac{2\sin\tau_1\sin\tau_2\cos\tau_2}{t_2s_2s_3}=-\frac{mp}{\bar{\mu}\sqrt{p_0|p|}}.
\end{aligned}\tag{A2}$$

Independent parameters are $s_1,s_3,t_1\sim t_3$, and then we can solve these equations. But for large $|m|$, there are no real solutions. In fact from (4), $2\cos^2\tau_1\cos^2\tau_2=1+(1-(m/\mu)^2)^{1/2}$. The reason is clarified in Sec. 6.5 and not essential. Now

(1) $[\psi_i(x,0),A(p)]=|p|^{-1/2}u_i(p)e^{-ipx}\psi_i(x,0)$,

(2) $[\psi_i(x,0),A^*(p)]=|p|^{-1/2}u_i(p)e^{ipx}\psi_i(x,0)$,

by the definitions of $A(p)$ and $A^*(p)$.

We first consider $W_{23}\psi_i(x,0)W_{23}^{-1}$:

$$W_{23}=\Pi_{0<p\leqslant\sigma}W_{23}(p),\quad W_{23}(p)=\exp[\varphi_1(p)g_{23}(p)],$$
$$g_{23}(p)=-\frac{\pi}{L}[(a_2^*(p)a_3(p)-\text{h.c.})+(p\rightarrow -p)].$$

Let $F=W_{23}(p)\psi_i(x,0)W_{23}^{-1}$. Then we get

$$\dot F\equiv\frac{d}{d\varphi_1}F=W_{23}(p)[g_{23}(p),\psi_i(x,0)]W_{23}^{-1}$$
$$=[\dot\chi_i^{(+)}+\dot\chi_i^{(-)}]F,$$

where

$$\chi_1^{(+)}=\frac{\pi}{L}|p|^{-1/2}\{-(\cosh\varphi_1-1)2\,a_2^*(-p)\times e^{ipx}-\sinh\varphi_i2a_3^*(-p)e^{ipx}\},$$
$$\chi_2^{(+)}=\chi_1^{(+)}(p\rightarrow -p),\quad \chi_i^{(-)}=-[\chi_i^{(+)}]^*.$$

Since $[\chi,\dot\chi]=0$ with $\chi=\chi_i^{(+)}+\chi_i^{(-)}$, we can integrate this equation under the condition $F(\varphi_1=0)=\psi_i(x,0)$:

$$F=\exp[\chi_i]\psi_i(x,0)=z^{-1/2}\exp[\chi_i^{(+)}]\psi_i(x,0)\exp[\chi_i^{(-)}],$$
$$z^{-1/2}=\exp[4\pi(e^{\varphi_1}-1)/L\,|p|\,].$$

We can similarly obtain $\Psi_i(x,0)=[UVW]\psi_i(x,0)[UVW]^{-1}$ in the form of $\exp[(\pi/L)|p|^{-1/2}\chi_i]\psi_i(x,0)$ or in the form of (6.30). Here $\chi_i=\chi_i(p)$ is a linear combination of $\{a_i^{(\pm)}(\pm p)\}_{i=1}^{3}$, and their coefficients are in Table II.

In Table II, we omit $e^{\pm ipx}$ for brevity and $p>0$ is assumed. In addition, the upper signs in ($\pm,\mp$) correspond to $i=1$, and the lower signs to $i=2$. We rewrite these coefficients in terms of $m,\mu,\bar\mu$, and p. This is done by using (1)–(5) after eliminating φ_1,φ_3, and τ_3. The results are the following:

$a_1^{(+)}(\pm p)$:0,

$$a_1^{(-)}(\pm p): \pm \frac{\sin\tau_1 \cos\tau_2 \sin\tau_2}{t_1 s_1 s_2 s_3} = \pm\left(-\frac{mp}{\bar{\mu}\sqrt{q_0|p|}}\right),$$

$a_2^{(+)}(p)$: $\pm(1 - s_2 t_2/s_1 s_3) = \pm(1-\mu/\bar{\mu})$,

$a_2^{(+)}(-p)$: $\mp(1-\mu/\bar{\mu})$,

$a_2^{(-)}(\pm p)$: $-(1 - s_1 s_3/s_2 t_2) = -(1-\bar{\mu}/\mu)$,

$a_3^{(+)}(\pm p)$:0,

$$a_3^{(-)}: -\frac{2\epsilon s_2 \cos\tau_1 \cos\tau_2}{s_1 t_1} = -\frac{m}{\mu}.$$

Thus by multiplying $(\pi/L)|p|^{-1/2}e^{\pm ipx}$, we have

$$\frac{\pi}{L}|p|^{-1/2}\chi_i(p)$$
$$= i\frac{\sqrt{2\pi}}{L}\left\{\frac{e}{\mu}\frac{i}{\sqrt{2|p|}}[a_3^{(-)}(p)e^{ipx}\right.$$
$$+ a_3^{(-)}(-p)e^{ipx}] + \sqrt{\pi}\left(1-\frac{\bar{\mu}}{\mu}\right)$$
$$\times\frac{i}{\sqrt{2|p|}}[a_2^{(-)}(p)e^{ipx}$$
$$+ a_2^{(-)}e^{-ipx}] + (\gamma_5)_{ii}\sqrt{\pi}\left(1-\frac{\mu}{\bar{\mu}}\right)\frac{i}{\sqrt{2|p|}}$$
$$\times[a_2^{(+)}(p)e^{ipx} - a_2^{(+)}(-p)e^{-ipx}]$$
$$+ (\gamma_5)_{ii}\frac{e}{\bar{\mu}}\frac{-i}{\sqrt{2q^0}}$$
$$\left.\times[a_1^{(-)}(p)e^{ipx} + a_1^{(-)}(-)e^{-ipx}]\right\},$$

where we have used $(\gamma_5) = \mathrm{diag}(-1,1)$. Then using the explicit expressions for $B'_\sigma, J'_\sigma, \tilde{J}'_\sigma$, and $\tilde{\phi}'_s$, we have the final expression in Sec. 6.5. The wave function renormalization constant $Z(L,\sigma)$ arises from the relabelling (Wick-type reordering) as before:

$$\exp[i\chi_i(x,0)]\psi_i(x,0)$$
$$= Z(L,\sigma)^{-1/2}{:}\exp[i\chi_i(x,0)]\psi_i(,0){:}.$$

$Z(L,\sigma)$ is easily obtained by means of the following equations:

(1) $e^{A+B} = e^A e^B \exp(-\frac{1}{2}[A,B]) = e^B e^A \times\exp(-\frac{1}{2}[B,A])$,

if $[A,B] = c1$, with $c\in C$,

(2) $\psi e^A = e^{-\lambda}e^A\psi$ if $[A,\psi] = \lambda\psi$, with $\lambda\in C$.

These equations are frequently used in order to investigate the fermion Wightman functions and the sine-Gordonization.

Finally remark that

$$\frac{d}{dx}e^{f(x)} = \left(\frac{d}{dx}f(x)\right)e^{f(x)} = e^{f(x)}\left(\frac{d}{dx}f(x)\right)$$

if $[f(x), (d/dx)f(x)] = 0$. This is also used to obtain $\Psi_i(x)$ and to consider the Dirac equation.

[1]J. Schwinger, Phys. Rev. **128**, 2425 (1962).

[2]J. Schwinger, in *Quantum Electrodynamics (Collected Papers)*, edited by J. Schwinger (Dover, New York, 1958).

[3]N.N. Bogolyubov and V.D. Shirkov, *Introduction to the Theory of Quantized Field* (Interscience, New York, 1959).

[4]J. Bognar, *Indefinite Inner Product Space* (Springer, Berlin, 1973).

[5]M.G. Krein, Am. Math. Soc. Trans. **93**, 115 (1970).

[6]M.G. Krein, Am. Math. Soc. Trans. **85**, 93 (1969).

[7]M.G. Krein and J.L. Smuljan, Am. Math. Soc. Trans. **85**, 93 (1965).

[8]B. Klaiber, in *Boulder Lectures in Theor. Phys.* (Gordon and Breach, New York, 1967), Vol. X-A, p. 1.

[9]B. Uhlenbrock, Commun. Math. Phys. **4**, 64 (1967).

[10]V. Volovich and N. Sushko, Theor. Math. Phys. **9**, 1086 (1971).

[11]N. Nakanishi, Phys. Rev. D **5**, 1324 (1972).

[12]N. Nakanishi, Prog. Theor. Phys. Suppl. **51**, 1 (1972).

[13]J. Ginibre and G. Velo, Ann. Inst. Henri Poincaré **22**, 257 (1975).

[14]J. Glimm and A. Jaffe, in *Statistical Mechanics and Quantum Field Theory*, edited by C.D. Witt *et al.* (Gordon and Breach, New York, 1971).

[15]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).

[16]J. Glimm and A. Jaffe, Commun. Pure Appl. Math. **22**, 401 (1969).

[17]H. Araki, Publ. RIMS **A-4**, 387 (168).

[18]F.A. Berezin, *The Method of Second Quantization* (Academic, New York, 1967).

[19]H. Araki, Publ. RIMS **7**, 121 (1971).

[20]H. Araki and M. Shiraishi, Publ. RIMS **7**, 105 (1971).

[21]D. Shale, Am. Math. Soc. Trans. **103**, 149 (1962).

[22]B. Simon, "Bogolyubov Transformation," Princeton preprint, 1972, unpublished.

[23]K.R. Ito, "Estimation of the Functional Determinants in Quantum Field Theories" and "Schwinger Model via Matthew-Salam Formula," RIMS preprints, (1979).

[24]M. Veltman and G.'t Hooft, Nucl. Phys. B **44**, 189 (1972).

[25]J. Glimm and A. Jaffe, Commun. Math. Phys. **22**, 1 (1971).

[26]O.A. McBryan, "Convergence of the vacuum energy density, ϕ-bound and existence of Wightman functions foy Yukawa$_2$ model," Rockfeller preprint (1975)

[27]O.A. McBryan, Commun. Math. Phys. **45**, 179 (1975).

[28]L. Rosen and A. Cooper, "The weakly coupled Yukawa$_2$ field theory," University of Tronto preprint (1976).

[29]E. Seiler and B. Simon, Commun. Math. Phys. **45**, 99 ('975).

[30]F. Strocchi and A.S. Wightman, J. Math. Phys. **15**, 2198 (1974).

[31]F. Strocchi, Commun. Math. Phys. **56**, (1977).

[32]G.'t Hooft, Nucl. Phys. B **33**, 173 (1971).

[33]G.'t Hooft, Nucl. Phys. B **35**, 167 (1971).

[34]K.R. Ito, Publ. RIMS **14**, 503 (1978); Lett. Math. Phys. **2**, 349 (1978).

[35]F. Guerra, L. Rosen, and B. Simon, Ann. of Math. **101**, 111 (1975).

[36]B. Simon, *The $p(\phi)_2$ Euclidean Quantum Field Theory* (Princeton U.P., Princeton N.J., 1974).

[37]J. Glimm, Commun. Math. Phys. **6**, 120 (1967).

[38]J. Glimm and A. Jaffe, Ann. Phys. (N.Y.) **60**, 321 (1970).

[39]B. Simon and E. Seiler, Ann. Phys. (N.Y.) **97**, 470 (1976).

[40]F. Hansen, private communication.

[41]J.H. Lowenstein and Y.A. Swieca, Ann Phys. (N.Y.) **68**, 172 (1971).

[42]E. Seiler, Commun. Math. Phys. **42**, 163 (1975).

[43]K. Ostewalder and R. Schrader, Helv. Phys. Acta **46**, 277 (1973).

[44]K. Osterwalder and R. Schrader, Commun. Math. Phys. **42**, 281 (1975).

[45]J. Glimm and A. Jaffe, Commun. Math. Phys. **56**, 165 (1977).

[46]K. Osterwalder and E. Seiler, Ann. Phys. (N.Y.) **110**, 440 (1978).

[47]J. Fröhlich and E. Seiler, Helv. Phys. Acta **49**, 889 (1976).

[48]K.R. Ito, Nucl. Phys. B **102**, 309 (1976).

[49]K.R. Ito, Phys. Lett. B **65**, 450 (1976); Lett. Math. Phys. **2**, 357 (1978).

[50]M. Reed, in *Lecture on Elementary Particle and Quantum Field Theory* edited by S. Deser *et al.* (MIT, Cambridge, Mass., 1970).

[51]J. Fröhlich, Phys. Rev. Lett. **34**, 833 (1975).

[52]J. Fröhlich, Commun. Math. Phys. **47**, 233 (1976).

[53]S. Colemen, Phys. Rev. D **11**, 2088 (1975).

[54]S. Colemen, R. Jackiw, and R. Susskind, Ann. Phys. (N.Y.) **93**, 267 (1975).

[55]R. May, Phys.Lett. A **25**, 282 (1967).

[56]J. Fröhlich, Commun. Math. Phys. **47**, 269 (1976).

[57]J. Fröhlich and Y. Park, Helv. Phys. Acta **50**, 315 (1977).

[58]S. Albeverio and R. Hoegh-Krohn, J. Func. Anal. **16**, 39 (1974).

[59]E. Lieb and J. Lebowitz, Adv. Math. **9**, 316 (1972).

[60]C. Deutsch and M. Lavaud, Phys. Rev. A **9**, 2598 (1973).

Relativistic wave equations coupled to external fields: An algebraic study of the problem of constraints

P. M. Mathews, T. R. Govindarajan, and M. Seetharaman
Department of Theoretical Physics, University of Madras, Guindy Campus, Madras-600025 India

J. Prabhakar[a]
Department of Physics, Loyola College, Madras-600036, India

(Received 16 May 1979)

A general matrix algebraic study is made of higher spin wave equations with minimal electromagnetic interaction, in relation to one of the basic problems, namely the problem of possible change in the number of constraints implied in the equation on introducing the interaction. Considering equations of the general form $(\beta\pi - m)\psi = 0$, wherein the matrix β_0 is required to have a minimal equation $\beta_0^n = \beta_0^{n-2}$ to ensure uniqueness of mass, we show that when $n = 4$ extra constraints may be generated at critical external fields, while for $n = 5$ there may also be loss of constraints on introduction of external fields. We obtain general algebraic criteria which determine whether or not such pathologies would arise in any particular case, and verify the validity of these criteria by considering a variety of known equations.

1. INTRODUCTION

The classical problem of construction of relativistic wave equations has been studied from various angles—field theoretic, matrix algebraic, and group theoretic.[1] Of various types of equations which have resulted from these studies, the equation written in the manifestly covariant form

$$(-i\beta^{\mu}\partial_{\mu} + m)\,\psi(x) = 0, \tag{1.1}$$

which is linear in the derivative, has a special appeal because, unlike in higher order equations, the prescription for the introduction of the minimal electromagnetic interaction is unambiguous. The minimal prescription does not, however, ensure that the equation gives a consistent description of a particle with the desired properties. In fact, it has been discovered[2–8] over the years that several types of difficulties arise in theories of higher spin particles interacting with external fields. They are related to the fact that in every unique mass–spin equation of the form (1.1), excepting the Dirac equation for spin-$\frac{1}{2}$, the wavefunction ψ must necessarily have more components than are required for the particle. The wave equation must then incorporate the appropriate number of constraints to ensure that just the required number of components (no more, no less) are independent and the rest are determined by the essential components. While the free equation (1.1) is so formulated as to meet this requirement, the introduction of interaction with an electromagnetic or other field may, and often does, cause a failure in this respect by leading either to unacceptable restrictions on the external field (as noted by Fierz and Pauli[9]) or to a loss of or an increase in the number of constraints.

Loss of constraints on introduction of the minimal e.m. interaction was first found to occur in a theory of spin-2 particles described by a 50-component wavefunction[10]; and recently we have noted[11] that the same difficulty arises in the spin-$\frac{3}{2}$ equation formulated by Glass.[12] It has also been pointed out that in certain theories (e.g., Rarita–Schwinger spin-$\frac{3}{2}$),[13] extra constraints may appear at certain critical values of the external fields.[14] However, there has been no investigation so far on general classes of wave equations to determine whether any general criteria exist for the occurrence (or otherwise) of this type of pathology. Our main aim in this paper is to show that such criteria can indeed be established in terms of the structure of the β matrices.

It is already known[15] that diagonalizability of β_0 is a sufficient condition for a theory to be free from many of the ills that higher spin theories are subject to. For example, the noncausal modes of propagation in external fields, to which attention was drawn first by Velo and Zwanziger,[3] do not arise if β_0 is diagonalizable.[16] This virtue of diagonalizability property is vividly brought out by our studies[17] on the Bhabha–Gupta equation[18] which showed that, only for the particular choice of the free parameters in the equation which makes β_0 diagonalizable, does one have the correct number of constraints (for all external field strength), causality of propagation, and (anti) commutators free of the external field. However, for this choice of parameters, the charge density is of indefinite sign, even in the absence of the external field, leading to difficulties in quantization. In fact, for diagonalizable β_0, the requirement that the field equation describing particle of half-integral spin $\geqslant\frac{3}{2}$ should lead to secondary constraints (established in the pioneering work of Johnson and Sudarshan[2]) cannot be met. Also, if the spin were integer $\geqslant 1$ and β_0 diagonalizable, the requirement that the energy density be nonnegative cannot be satisfied.[19]

It becomes necessary therefore to consider especially various classes of theories wherein β_0 is not diagonalizable. Theories with nondiagonalizable β_0's such as Khalil's theory for spin-$\frac{1}{2}$ [20] and the spin-1 theory of Shamaly and Capri[21] demonstrate that nondiagonalizability of β_0 by itself need not make a theory prey to the various kinds of ills which some theories (e.g., Rarita–Schwinger theory for spin-$\frac{3}{2}$) suffer from. Thus encouraged, we take up for study the general classes of the equations of the form (1.1), characterized by

[a]Present address: The Ramakrishna Mission, Saradapitha, Belur Math, P.O., West Bengal-711202, India.

 0022-2488/80/061495-11$1.00

either of the minimal equations $\beta_0^4 - \beta_0^2 = 0$ or $\beta_0^5 - \beta_0^3 = 0$ for β_0. These are the simplest two cases (for nondiagonalizable β_0) of the Harish–Chandra condition for uniqueness of mass,[22] namely

$$\beta_0^{n-2}(\beta_0^2 - 1) = 0, \quad n \geqslant 2. \tag{1.2}$$

Assuming only the above algebraic property besides the conditions imposed by Lorentz invariance, we determine general criteria for the number of constraints contained in Eq. (1.1) to remain unchanged on introduction of the minimal electromagnetic interaction.

The plan of the paper is as follows: After a resumé of the conditions for relativistic invariance of equations of the form (1.1) in Sec. 2, we take up in Sec. 3 the study of equations with nondiagonalizable β_0 obeying $\beta_0^4 = \beta_0^2$. We enumerate the constraints in the presence of interaction and find that there is a possibility of *more* constraints than in the free case being generated for a particular strength of the external e.m. field. This would imply a breakdown of covariance, despite the manifestly covariant appearance of the equations. We derive a general condition that must be obeyed if such a breakdown is to be averted. The next section (Sec. 4) presents an analysis of equations coming under the algebra $\beta_0^5 = \beta_0^3$. In this case there is in general a possibility of *loss* of constraints. We obtain in Sec. 4 a set of conditions which would ensure that such a pathological situation does not arise. We then give examples of specific theories which suffer from, and others which are free of, these difficulties, and show how their behavior is traceable to violation of or conformity to our conditions. It may be mentioned here that the conditions deduced in Secs. 3 and 4 involve the Lorentz group generators appropriate to the wave function, and hence place restrictions on the structure of the wave equation in terms of the Lorentz group representations entering therein. Deferring a detailed analysis of these restrictions to later publication, we present in Sec. 5 a discussion of the results of this paper against the background of the earlier literature on relativistic wave equations with interaction.

2. MANIFEST COVARIANCE OF THE WAVE EQUATIONS

It is well known that if $\psi(x)$ is a finite component wave function, which under the Lorentz transformations $x \rightarrow x' = \Lambda x$ transforms according to a single or double valued representation $S(\Lambda)$ of the Lorentz group

$$\psi(x) \rightarrow \psi'(x') = S(\Lambda)\,\psi(x), \tag{2.1}$$

then the relativistic invariance of Eq. (1.1) demands that

$$S(\Lambda)^{-1}\beta^\mu S(\Lambda) = \Lambda^\mu{}_\nu \beta^\nu. \tag{2.2}$$

When expressed in terms of the infinitesimal generators **J** of rotations and **K** of boosts, Eq. (2.2) goes over into the following set of equations:

$$[J_i, \beta_0] = 0, \tag{2.3a}$$

$$[J_i, \beta_j] = i\epsilon_{ijk}\,\beta_k\,, \tag{2.3b}$$

$$[K_i, \beta_0] = i\beta_i\,, \tag{2.3c}$$

$$[K_i, \beta_j] = i\delta_{ij}\,\beta_0\,. \tag{2.3d}$$

By eliminating β_0 between the last two of these equations, one gets

$$[[K_i, \beta_0], K_j] = \delta_{ij}\,\beta_0\,. \tag{2.4}$$

Equations (2.3a) and (2.4) show how β_0 and the Lorentz generators are interrelated. When the latter are specified, the determination of the β_μ proceeds through the solution of this pair of equations for β_0 followed by the evaluation of the β_i using Eq. (2.3c). In carrying through such a process explicitly, it is conventional to employ a canonical basis which diagonalizes $\mathbf{J}^2$ and J_z. (See, for instance, Ref. 23). In the present work, however, it will be most convenient, for the purpose of the general analysis, to use a basis which reduces β_0 to a direct sum of minimal blocks [See Eqs. (3.3a) and (4.1) below.]

Apart from covariance, we demand also that Eq. (1.1) should describe particle of unique mass. The necessary and sufficient condition on the β matrices for this is that

$$(\beta\cdot p)^n = p^2(\beta\cdot p)^{n-2}, \quad n \geqslant 2 \tag{2.5a}$$

or

$$\sum_{\mathscr{P}}(\beta_{\mu_1}\beta_{\mu_2} - g_{\mu_1\mu_2})\beta_{\mu_3}\beta_{\mu_4}\cdots\beta_{\mu_n} = 0, \tag{2.5b}$$

where the sum is over permutations $\mathscr{P}$ of the vector indices. If all the μ_i in Eq. (2.5b) are chosen to be zero, the equation simplifies to

$$(\beta_0^2 - 1)\beta_0^{n-2} = 0, \quad n \geqslant 2. \tag{2.6}$$

Equations (2.4) and (2.6) are the only requirements that we shall make in our analysis.

When $n = 2$ in Eq. (2.5b), one has the familiar Dirac algebra. The case $n = 3$, of which the Kemmer algebra[24] is a special case,[25] has been investigated systematically by Hurley and Sudarshan.[23] There have been no systematic analyses of cases $n > 3$ (for which alone β_0 is nondiagonalizable). As already stated, our aim is to consider certain aspects of general classes of theories corresponding to $n = 4$ and $n = 5$.

3. THE ALGEBRA $\beta_0^4 - \beta_0^2 = 0$

In this section we consider equations involving nondiagonalizable β_0 characterized by the minimal equation

$$\beta_0^4 = \beta_0^2\,. \tag{3.1}$$

The corresponding relations among the β matrices are obtained by setting $n = 4$ in Eq. (2.5b):

$$\sum_{\mathscr{P}}(\beta_\mu\beta_\nu - g_{\mu\nu})\,\beta_\rho\beta_\sigma = 0. \tag{3.2}$$

Of the equations belonging to this category, the most familiar one is the Rarita–Schwinger equation[13] for spin-$\frac{3}{2}$ particles, which is known to suffer from the following troubles when minimal coupling to the electromagnetic field is introduced: (a) indefiniteness of anticommutators between field components,[2] (b) noncausal propagation,[3] (c) an energy spectrum ceases to be wholly real when the particle is placed in a sufficiently strong magnetic field,[6] and (d) the occurrence of an excessive number of constraints when the magnetic component of the external e.m. field takes a certain value. However, these troubles are not a necessary concomitant of the algebraic property (3.1). In fact, the spin-$\frac{1}{2}$ equation recently proposed by Khalil[20] [which does have Eq. (3.1) as the minimal equation for β_0] is not afflicted by any of the above problems. This observation raises the following question: Is it possible to understand the nice behavior of Khalil's equation in terms of any algebraic condition obeyed by β_0 in relation to the Lorentz group representations entering into the equation? Conversely, is the pathological behavior of the RS equation a consequence of violating some such condition? We seek to answer these questions in this section. We first obtain a condition which β_0 and the Lorentz generators need to obey in order that breakdown of covariance (in the form of a change in the number of constraints at particular values of the external fields) can be avoided. We then verify that this condition is indeed satisfied in Khalil's equation, while it is violated in the RS theory.

To carry through our calculations in a transparent fashion it is convenient to take the matrix β_0 in the block-diagonal form[26]

$$\beta_0 = \begin{pmatrix} A & & \\ & 0 & \\ & & B \end{pmatrix}, \tag{3.3a}$$

where 0 is a null matrix and the block A and B have the minimal equations

$$A^2 - 1 = 0 \quad \text{and} \quad B^2 = 0. \tag{3.3b}$$

The boost generators $\mathbf{K}$ (in the representation carried by ψ) will also be partitioned conformably to β_0:

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \mathbf{K}_{13} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \mathbf{K}_{23} \\ \mathbf{K}_{31} & \mathbf{K}_{32} & \mathbf{K}_{33} \end{pmatrix}. \tag{3.4}$$

By virtue of the relation $\beta_j = i[\beta_0, K_j]$, one can then obtain the matrix β_j in a similar form. Taking the column ψ also to be partitioned conformably into three parts ψ_1, ψ_2, ψ_3, we write the equation $(\beta_0\pi_0 - \boldsymbol{\beta}\cdot\boldsymbol{\pi} - m)\psi = 0$ as the following set of equations:

$$(A\pi_0 - m)\psi_1 + i[\mathbf{K}_{11}\cdot\boldsymbol{\pi}, A]\psi_1 - iA\,\mathbf{K}_{12}\cdot\boldsymbol{\pi}\psi_2 + i(\mathbf{K}_{13}\cdot\boldsymbol{\pi}B - A\,\mathbf{K}_{13}\cdot\boldsymbol{\pi})\psi_3 = 0, \tag{3.5a}$$

$$i\mathbf{K}_{21}\cdot\boldsymbol{\pi}A\psi_1 - m\psi_2 + i\mathbf{K}_{23}\cdot\boldsymbol{\pi}B\psi_3 = 0, \tag{3.5b}$$

$$(B\pi_0 - m)\psi_3 + i(\mathbf{K}_{31}\cdot\boldsymbol{\pi}A - B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi})\psi_1 - iB\,\mathbf{K}_{32}\cdot\boldsymbol{\pi}\,\psi_2 + i[\mathbf{K}_{33}\cdot\boldsymbol{\pi}, B]\psi_3 = 0. \tag{3.5c}$$

Equations (3.5) provide a convenient starting point for the analysis of the constraints on ψ. First of all we observe that Eq. (3.5a) provides an equation of motion for ψ_1 (A being nonsingular) and Eq. (3.5c) for $B\psi_3$. However, Eq. (3.5b) is a constraint equation as it does not involve any time derivatives. One more set of constraints may be generated by multiplying Eq. (3.5c) with B from the left. Since $B^2 = 0$, we then get

$$-mB\psi_3 + iB\,\mathbf{K}_{31}\cdot\boldsymbol{\pi}A\psi_1 + iB\,\mathbf{K}_{33}\cdot\boldsymbol{\pi}B\psi_3. \tag{3.6}$$

Equations (3.5b) and (3.6) are the "primary constraints" of the theory in the terminology of Johnson and Sudarshan.[2] To see whether further constraints exist, we have to differentiate the primary constraints with respect to time (or rather, apply the operator π_0). When this is done, Eq. (3.5b) does not lead to any new constraints. Instead it gives an equation of motion for ψ_2:

$$m\pi_0\psi_2 = \{im\mathbf{K}_{21}\cdot\boldsymbol{\pi} + \mathbf{K}_{21}\cdot\boldsymbol{\pi}[\mathbf{K}_{11}\cdot\boldsymbol{\pi}, A] - e\mathbf{K}_{21}\cdot\mathscr{E}A + \mathbf{K}_{23}\cdot\boldsymbol{\pi}(\mathbf{K}_{31}\cdot\boldsymbol{\pi}A - B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi})\}\psi_1 - \{\mathbf{K}_{21}\cdot\boldsymbol{\pi}A\,\mathbf{K}_{12}\cdot\boldsymbol{\pi} + \mathbf{K}_{23}\cdot\boldsymbol{\pi}B\,\mathbf{K}_{32}\cdot\boldsymbol{\pi}\}\psi_2 + \{\mathbf{K}_{21}\cdot\boldsymbol{\pi}(\mathbf{K}_{13}\cdot\boldsymbol{\pi}B - A\,\mathbf{K}_{13}\cdot\boldsymbol{\pi}) + \mathbf{K}_{23}\cdot\boldsymbol{\pi}[\mathbf{K}_{33}\cdot\boldsymbol{\pi}, B] - e\,\mathbf{K}_{23}\cdot\mathscr{E}B\}\psi_3. \tag{3.7}$$

On the other hand, Eq. (3.6) on operating with π_0 leads to

$$(iB\,\mathbf{K}_{33}\cdot\boldsymbol{\pi} - m)\pi_0 B\psi_3 + iB\,\mathbf{K}_{31}\cdot\boldsymbol{\pi}\pi_0 A\psi_1 - eB\,\mathbf{K}_{33}\cdot\mathscr{E}B\psi_3 - eB\,\mathbf{K}_{31}\cdot\mathscr{E}A\psi_1 = 0, \tag{3.8}$$

wherein $\pi_0 A\psi_1$ and $\pi_0 B\psi_3$ may be replaced by expressions free of time derivatives by using Eq. (3.5a) and (3.5b), respectively. Thereupon we get the *secondary constraint equations*

$$\{(im + B\,\mathbf{K}_{33}\cdot\boldsymbol{\pi})(\mathbf{K}_{31}\cdot\boldsymbol{\pi}A - B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi}) - eB\,\mathbf{K}_{31}\cdot\mathscr{E}A + imB\,\mathbf{K}_{31}\cdot\boldsymbol{\pi} + B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi}[\mathbf{K}_{11}\cdot\boldsymbol{\pi}, A]\}\psi_1 - \{(im + B\,\mathbf{K}_{33}\cdot\boldsymbol{\pi})B\,\mathbf{K}_{32}\cdot\boldsymbol{\pi} + B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi}A\,\mathbf{K}_{12}\cdot\boldsymbol{\pi}\}\psi_2 - \{m^2 + B(\mathbf{K}\cdot\boldsymbol{\pi}\beta_0\mathbf{K}\cdot\boldsymbol{\pi})_{33}\}\psi_3 + \{im\mathbf{K}_{33}\cdot\boldsymbol{\pi} + B(\mathbf{K}_{33}\cdot\boldsymbol{\pi})^2 + B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi}\mathbf{K}_{13}\cdot\boldsymbol{\pi} - eB\,\mathbf{K}_{33}\cdot\mathscr{E}\}B\psi_3 = 0. \tag{3.9}$$

The next step is to apply π_0 to this equation and eliminate $\pi_0\psi_1$, $\pi_0\psi_2$, and $\pi_0 B\psi_3$ using Eqs. (3.5a), (3.7), and (3.5c), respectively. The time derivative still persists through a term involving $\pi_0\psi_3$, namely

$$[B(\mathbf{K}\cdot\boldsymbol{\pi}\beta_0\mathbf{K}\cdot\boldsymbol{\pi})_{33} + m^2]\pi_0\psi_3.$$

This term can be rewritten using the equation $[[K_i, \beta_0], K_j] = \delta_{ij}\beta_0$ as

$$\{\tfrac{1}{2}B(\mathbf{K}\cdot\boldsymbol{\pi})^2_{33}B - \tfrac{1}{2}ieB(\mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K})_{33} + m^2\}\pi_0\psi_3, \tag{3.10}$$

where

$$\mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K} \equiv K_i\beta_0 F_{ij} K_j, \tag{3.10a}$$

with summation over the repeated indices. In the first term in Eq. (3.10), $B\pi_0\psi_3$ can once again be eliminated using Eq. (3.5c). The time derivative term which finally remains is, say,

$$\{m^2 - \tfrac{1}{2}ieB(\mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K})_{33}\}\pi_0\psi_3 \equiv M\pi_0\psi_3. \tag{3.11}$$

Observe that in the absence of electromagnetic interaction this term reduces to $m^2\pi_0\psi_3$, so that the equation obtained by differentiating Eq. (3.9) is an equation of motion for ψ_3 and not a constraint. We should require that the same state of affairs be maintained also in the presence of interaction. This means that M, the operator in curly brackets in Eq. (3.11), should be nonsingular for *all* values of the external fields. It is not difficult to verify that a sufficient condition for this to be true is that the second term in the square brackets in Eq. (3.11) be annihilated by B operating on the right, i.e.,

$$B(\mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K})_{33}B = 0 \tag{3.12a}$$

or equivalently

$$\beta_0(1 - \beta_0^2)\mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K}(1 - \beta_0^2)\beta_0 = 0. \tag{3.12b}$$

As this is to hold for all F_{ij}, we must have (in view of the antisymmetry of F_{ij})

$$\beta_0(1-\beta_0^2)(K_i\beta_0 K_j - K_j\beta_0 K_i)(1-\beta_0^2)\beta_0 = 0. \tag{3.13}$$

This condition, however, is somewhat stronger than necessary. It can be shown that if a basis in which β_0 is in the Jordan canonical form is employed, then the necessary and sufficient condition is that

$$B(K\beta_0{\cdot}F{\cdot}K)_{33}\, BB^\dagger \tag{3.14a}$$

be nilpotent or equivalently that

$$\beta_0(1-\beta_0^2)(K_i\beta_0K_j - K_j\beta_0K_i)(1-\beta_0^2)\beta_0\beta_0^\dagger \tag{3.14b}$$

be nilpotent. If this condition is not satisfied, M of Eq. (3.11) becomes singular from some value of the external magnetic field—note that the electric field components F_{0i} do not enter—and for such a field the equation obtained by differentiating Eq. (3.9) gives rise to extra constraints which have no counterparts in the noninteracting case. A change of Lorentz frame would change the value of the field, resulting in the disappearance of these new constraints, and this evidently means a breakdown of Lorentz covariance. To avoid this it is necessary that the operator (3.14) be nilpotent.

Having formulated this general condition, we now proceed to test whether it is obeyed in specific theories. In practice, we first test whether Eq. (3.13) is obeyed and only if it is not does it become necessary to test Eq. (3.14) for nilpotency. It is important to note that, though we have used the reducibility of β_0 to the form (3.3) in deriving Eq. (3.13), this condition itself is independent of whether β_0 is in such a reduced form or not. Therefore, in testing particular theories, β_0 and $\mathbf{K}$ as given with respect to any convenient basis can be employed in Eq. (3.13).

A. Rarita–Schwinger theory for spin-$\frac{3}{2}$

The Rarita-Schwinger equation for spin-$\frac{3}{2}$ particles employs a 16-component wave function $\psi(x)$ which transforms according to the representation $(1,\frac{1}{2})\oplus(\frac{1}{2},1)\oplus(0,\frac{1}{2})\oplus(\frac{1}{2},0)$ of the homogeneous Lorentz group (HLG). Written in the general linear form, the equation reads

$$(\beta_0\pi_0 - \boldsymbol{\beta}{\cdot}\boldsymbol{\pi} - m)\,\psi(x) = 0, \tag{3.15}$$

with[27]

$$\beta_0 = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}, \tag{3.16}$$

where A and B are 8×8 matrices given by[28]

$$A = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad B = \frac{1}{2}\begin{pmatrix} -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}. \tag{3.17}$$

In A, I stands for the unit matrix of dimensions four and in B, 1 stands for the unit matrix of dimension two. (Note that $B^2 = 0$.) The infinitesimal generators $\mathbf{K}$ of Lorentz transformations are given by

$$-i\mathbf{K} = \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{13} \\ \mathbf{K}_{31} & \mathbf{K}_{33} \end{pmatrix}, \tag{3.18}$$

where

$$\mathbf{K}_{11} = \begin{pmatrix} 0 & -\frac{1}{3}\boldsymbol{\Sigma} \\ -\frac{1}{3}\boldsymbol{\Sigma} & 0 \end{pmatrix}, \tag{3.19a}$$

$$\mathbf{K}_{13} = \begin{pmatrix} 0 & 0 & \mathbf{u} & 0 \\ 0 & \mathbf{u} & 0 & 0 \end{pmatrix}, \tag{3.19b}$$

$$\mathbf{K}_{31} = \mathbf{K}_{13}^\dagger, \tag{3.19c}$$

$$\mathbf{K}_{33} = \begin{pmatrix} 0 & 0 & 0 & -\frac{1}{2}\boldsymbol{\sigma} \\ 0 & 0 & -\frac{5}{6}\boldsymbol{\sigma} & 0 \\ 0 & -\frac{5}{6}\boldsymbol{\sigma} & 0 & 0 \\ -\frac{1}{2}\boldsymbol{\sigma} & 0 & 0 & 0 \end{pmatrix}. \tag{3.19d}$$

Here, $\boldsymbol{\Sigma}$ forms the spin-$\frac{3}{2}$ representation of angular momentum, while $\boldsymbol{\sigma}$ stands for the set of Pauli matrices. The components of $\mathbf{u}$ are rectangular matrices given by

$$u_1 = \frac{i\sqrt{2}}{3}\begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & -\sqrt{3} \end{pmatrix}; \quad u_2 = \frac{\sqrt{2}}{3}\begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & \sqrt{3} \end{pmatrix};$$

$$u_2 = -\frac{2i\sqrt{2}}{3}\begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{3.20}$$

The following relations among $\mathbf{u}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\Sigma}$ will be used in our calculations:

$$\begin{aligned}
&[\Sigma_i, \Sigma_j] = i\epsilon_{ijk}\Sigma_k\,,\\
&u_iu_j^\dagger - u_ju_i^\dagger = \tfrac{8}{9}\, i\epsilon_{ijk}\Sigma_k\,,\\
&\Sigma_iu_j - \Sigma_ju_i = \tfrac{5}{2}\, i\epsilon_{ijk}u_k\,,\\
&u_i\sigma_j - u_j\sigma_i = -\tfrac{1}{6}\, i\epsilon_{ijk}u_k\,,\\
&u_i^\dagger u_j - u_j^\dagger u_i = -\tfrac{8}{9}\, i\epsilon_{ijk}\sigma_k\,,\\
&[\sigma_i,\sigma_j] = 2i\epsilon_{ijk}\sigma_k\,.
\end{aligned} \tag{3.21}$$

Using these it is an easy matter to check that the condition (3.13) is indeed violated. In fact, the matrix of Eq. (3.13) may be seen to reduce to a nonnull matrix whose nonzero elements are proportional to $\boldsymbol{\sigma}{\cdot}\mathscr{H}$. (Here $\mathscr{H}$ is the magnetic field $\mathscr{H}_i = \frac{1}{2}\epsilon_{ijk}F_{jk}$.) Multiplying the nonnull matrix by $\beta_0^\dagger$ on the right, one readily finds that the less stringent condition of nilpotency of Eq. (3.14) is also not satisfied. Alternatively, one can verify directly that M of Eq. (3.11) becomes singular in the present case when $\frac{2}{3}e\mathscr{H} = m^2$. It may be recalled that it is at this critical value that other troubles of the RS theory begin to manifest themselves: The anticommutators between the field components cease to be of definite sign, the equation ceases to be hyperbolic[3] and the energy eigenvalue spectrum ceases to be wholly real.[8]

B. Khalil's equation for spin 1/2

Khalil's equation[20] for spin-$\frac{1}{2}$ particles employs a 20-component wave function transforming according to the representation

$$T(\Lambda) = T_1(\Lambda) + 2T_2(\Lambda) + \bar{T}_1(\Lambda) + 2\bar{T}_2(\Lambda)\,,$$

where

$$T_1(\Lambda) = (1,\tfrac{1}{2}) \quad\text{and}\quad T_2(\Lambda) = (\tfrac{1}{2},0). \tag{3.22}$$

In a basis diagonalizing $\mathbf{J}^2$ and J_z, the boost generators in $T_1(\Lambda)$ are given by

$$\mathbf{K}^{(1)} = -i\begin{pmatrix} \frac{1}{3}\boldsymbol{\Sigma} & \mathbf{u} \\ \mathbf{u}^\dagger & \frac{5}{6}\boldsymbol{\sigma} \end{pmatrix}. \tag{3.23a}$$

In $T_2(\Lambda)$,

$$\mathbf{K}^{(2)} = -\frac{i}{2}\boldsymbol{\sigma}. \tag{3.23b}$$

Here $\boldsymbol{\sigma}$, $\boldsymbol{\Sigma}$, and $\mathbf{u}$ are the matrices already defined in the previous subsection. The first row (column) of Eq. (3.23a) is associated with spin value $\frac{3}{2}$ and the second with $\frac{1}{2}$. In the representations $\bar{T}_1$ and $\bar{T}_2$, the boost generators are the negative of Eqs. (3.23a) and (3.23b), respectively. Using these we have, for the boost generator $\mathbf{K}$ in the reducible representation employed by Khalil,

$$i\mathbf{K} = \begin{pmatrix} \frac{1}{2}\boldsymbol{\sigma} & & & & & & & \\ & -\frac{1}{2}\boldsymbol{\sigma} & & & & & 0 & \\ & & \frac{1}{3}\boldsymbol{\Sigma} & \mathbf{u} & & & & \\ & & \mathbf{u}^\dagger & \frac{5}{6}\boldsymbol{\sigma} & & & & \\ & & & & -\frac{1}{3}\boldsymbol{\Sigma} & -\mathbf{u} & & \\ & 0 & & & -\mathbf{u}^\dagger & -\frac{5}{6}\boldsymbol{\sigma} & & \\ & & & & & & \frac{1}{2}\boldsymbol{\sigma} & \\ & & & & & & & -\frac{1}{2}\boldsymbol{\sigma} \end{pmatrix}. \tag{3.24}$$

In the same representation, the matrix β_0 is given by

$$\beta_0 = \begin{pmatrix} 0 & 1 & 0 & f & 0 & 0 & 0 & if \\ 1 & 0 & 0 & 0 & 0 & f & -if & 0 \\ 0 & 0 & & & & & & \\ f & 0 & & & & & & \\ 0 & 0 & & & 0 & & & \\ 0 & f & & & & & & \\ 0 & -if & & & & & & \\ if & 0 & & & & & & \end{pmatrix}, \tag{3.25}$$

wherein the nonvanishing blocks are multiples of the unit matrix of dimension 2. Using the above forms of β_0 and $\mathbf{K}$, one can easily verify that the condition (3.13) is satisfied in this case. The absence of any difficulty about the number of constraints in Khalil's equation is thus explained.

Before proceeding to the next higher algebra $\beta_0^5 = \beta_0^3$, we wish to draw attention to an interesting connection between our condition for covariance and a sufficient condition for causality which was recently obtained by Khalil.[29] He has shown that a field ψ obeying the general equation $(\beta\cdot\pi - m)\,\psi = 0$ (which incorporates minimal e.m. interaction) propagates causally if

$$\Big[\sum_{\mathscr{P}} (\beta_{\mu_1}\beta_{\mu_2}\beta_{\mu_3} - g_{\mu_1\mu_2}\beta_{\mu_3})\Big]\beta_{\mu_4}\cdots\beta_{\mu_n} = 0. \tag{3.26}$$

Note that this condition automatically ensures that the unique-mass condition (2.6) is satisfied, but it is more restrictive than Eq. (2.5b). For $n = 4$, Eq. (3.26) reduces to

$$\Big[\sum_{\mathscr{P}} (\beta_{\mu_1}\beta_{\mu_2}\beta_{\mu_3} - g_{\mu_1\mu_2}\beta_{\mu_3})\Big]\beta_{\mu_4} = 0. \tag{3.27}$$

Setting $\mu_1 = \mu_2 = \mu_3 = 0$ and $\mu_4 = i$ in this equation, one gets

$$(\beta_0^2 - 1)\,\beta_0\beta_i = 0. \tag{3.28}$$

On using $\beta_i = i[\beta_0, K_i]$ together with $\beta_0^4 = \beta_0^2$, Eq. (3.28) reduces to

$$(\beta_0^2 - 1)\,\beta_0 K_i \beta_0 = 0. \tag{3.29}$$

It is obvious that if Eq. (3.29) is satisfied, Eq. (3.13) is automatically satisfied. Thus, the condition (3.26) which is sufficient to ensure causality of propagation is (perhaps not surprisingly) sufficient also to prevent the appearance of constraints in a noncovariant fashion. It remains to be seen whether a weaker condition than Eq. (3.26) might be sufficient to keep the propagation causal.[30]

4. THE ALGEBRA $\beta_0^5 - \beta_0^3 = 0$

When the degree of the minimal equation (2.6) of β_0 goes up to five, a new type of trouble makes its appearance. It is the possibility of loss of constraints (or equivalently, an increase in the number of degrees of freedom) when interactions with external fields are switched on. As we pointed out in a recent paper,[11] Glass' equation[12] for spin-$\frac{3}{2}$ (which belongs to the category of equations with $\beta_0^5 = \beta_0^3$) is subject to this pathology; but there are other equations which are not. Examples are the spin-2 equations of Schwinger and Chang[31] involving a 30-component field (which was studied by Hagen[32]) and the Shamaly–Capri equation for spin-1.[21] Our aim in this section is to derive an algebraic criterion for the general linear wave equation $(\beta\cdot\pi - m)\,\psi = 0$ with $\beta_0^5 - \beta_0^3 = 0$ to lead to the same number of constraints in the presence of external fields as in the free case, and to show that the specific examples mentioned above are consistent with this criterion.

A. A General condition for preservation of constraints

The minimal equation $\beta_0^5 = \beta_0^3$ implies that β_0 can be brought to the block-diagonal form

$$\beta_0 = \begin{pmatrix} A & & \\ & B & \\ & & C \end{pmatrix}, \tag{4.1}$$

(with 0 in the off-diagonal positions) with the minimal equations

$$A^2 = 1, \quad B^2 = 0, \quad \text{and} \quad C^3 = 0 \tag{4.1a}$$

for the individual blocks. Partitioning $\mathbf{K}$ and ψ in a manner conformable to β_0 of Eq. (4.1):

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \mathbf{K}_{13} & \mathbf{K}_{14} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \mathbf{K}_{23} & \mathbf{K}_{24} \\ \mathbf{K}_{31} & \mathbf{K}_{32} & \mathbf{K}_{33} & \mathbf{K}_{34} \\ \mathbf{K}_{41} & \mathbf{K}_{42} & \mathbf{K}_{43} & \mathbf{K}_{44} \end{pmatrix}, \psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}, \tag{4.2}$$

we rewrite the equation of motion

$$(\beta_0\pi_0 - \boldsymbol{\beta}\cdot\boldsymbol{\pi} - m)\,\psi(x) = 0$$

as the following set of equations:

$$(A\pi_0 - m)\,\psi_1 + i[\mathbf{K}_{11}\cdot\boldsymbol{\pi},A\,]\,\psi_1 - iA\,\mathbf{K}_{12}\cdot\boldsymbol{\pi}\psi_2 + i(\mathbf{K}_{13}\cdot\boldsymbol{\pi}B - A\,\mathbf{K}_{13}\cdot\boldsymbol{\pi})\psi_3 + i(\mathbf{K}_{14}\cdot\boldsymbol{\pi}C - A\,\mathbf{K}_{14}\cdot\boldsymbol{\pi})\psi_4 = 0, \tag{4.3a}$$

$$i\mathbf{K}_{21}\cdot\boldsymbol{\pi}A\psi_1 - m\psi_2 + i\mathbf{K}_{23}\cdot\boldsymbol{\pi}B\psi_3 + i\mathbf{K}_{24}\cdot\boldsymbol{\pi}C\psi_4 = 0, \tag{4.3b}$$

$$(B\pi_0 - m)\,\psi_3 + i(\mathbf{K}_{31}\cdot\boldsymbol{\pi}A - B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi})\,\psi_1 - iB\,\mathbf{K}_{32}\cdot\boldsymbol{\pi}\,\psi_2 + i[\mathbf{K}_{33}\cdot\boldsymbol{\pi},B\,]\,\psi_3 + i(\mathbf{K}_{34}\cdot\boldsymbol{\pi}C - B\,\mathbf{K}_{34}\cdot\boldsymbol{\pi})\,\psi_4 = 0, \tag{4.3c}$$

$$(C\pi_0 - m)\,\psi_4 + i(\mathbf{K}_{41}\cdot\boldsymbol{\pi}A - C\,\mathbf{K}_{41}\cdot\boldsymbol{\pi})\,\psi_1 - iC\,\mathbf{K}_{42}\cdot\boldsymbol{\pi}\psi_2 + i(\mathbf{K}_{43}\cdot\boldsymbol{\pi}B - C\,\mathbf{K}_{43}\cdot\boldsymbol{\pi})\,\psi_3 + i[\mathbf{K}_{44}\cdot\boldsymbol{\pi},C\,]\,\psi_4 = 0. \tag{4.3d}$$

It is evident that Eq. (4.3b) is a constraint equation. Further constraints may be obtained by premultiplying Eq. (4.3c) by B and Eq. (4.3d) by C^2. We then get, in view of Eq. (4.1a),

$$-mB\psi_3 + iBK_{31}\cdot\pi A\psi_1 + iBK_{33}\cdot\pi B\psi_3 + iBK_{34}\cdot\pi C\psi_4 = 0, \tag{4.4}$$

$$-mC^2\,\psi_4 + iC^2\,K_{41}\cdot\pi A\psi_1 + iC^2K_{43}\cdot\pi B\psi_3 + iC^2K_{44}\cdot\pi C\psi_4 = 0. \tag{4.5}$$

Equations (4.3b), (4.4), and (4.5) are the primary constraints. Operating with π_0 on Eq. (4.3b), we get an equation of motion for ψ_2; but from Eq. (4.4) and (4.5) we get further constraints. They are

$$\begin{aligned}
&\{(im + B\,\mathbf{K}_{33}\cdot\boldsymbol{\pi})(\mathbf{K}_{31}\cdot\boldsymbol{\pi}A - B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi}) + imB\,\mathbf{K}_{31}\cdot\boldsymbol{\pi} + B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi}[\mathbf{K}_{11}\cdot\boldsymbol{\pi},A\,] + B\,\mathbf{K}_{34}\cdot\boldsymbol{\pi}(\mathbf{K}_{41}\cdot\boldsymbol{\pi}A - C\,\mathbf{K}_{41}\cdot\boldsymbol{\pi}) - eB\,\mathbf{K}_{31}\cdot\mathscr{E}\,A\,\}\,\psi_1\\
&\quad - \{(im + B\,\mathbf{K}_{33}\cdot\boldsymbol{\pi})B\,\mathbf{K}_{32}\cdot\boldsymbol{\pi} + B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi}A\,\mathbf{K}_{12}\cdot\boldsymbol{\pi} + B\,\mathbf{K}_{34}\cdot\boldsymbol{\pi}C\,\mathbf{K}_{42}\cdot\boldsymbol{\pi}\}\,\psi_2\\
&\quad + \{(im + B\,\mathbf{K}_{33}\cdot\boldsymbol{\pi})\,\mathbf{K}_{33}\cdot\boldsymbol{\pi} + B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi}\mathbf{K}_{13}\cdot\boldsymbol{\pi} + B\,\mathbf{K}_{34}\cdot\boldsymbol{\pi}\mathbf{K}_{43}\cdot\boldsymbol{\pi} - eB\,\mathbf{K}_{33}\cdot\mathscr{E}\}\,B\psi_3\\
&\quad + \{(im + B\,\mathbf{K}_{33}\cdot\boldsymbol{\pi})\,\mathbf{K}_{34}\cdot\boldsymbol{\pi} + B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi}\mathbf{K}_{14}\cdot\boldsymbol{\pi} + B\,\mathbf{K}_{34}\cdot\boldsymbol{\pi}\mathbf{K}_{44}\cdot\boldsymbol{\pi} - eB\,\mathbf{K}_{34}\cdot\mathscr{E}\}\,C\psi_4\\
&\quad - \{m^2 + B\,(\mathbf{K}\cdot\boldsymbol{\pi}\,\beta_0\,\mathbf{K}\cdot\boldsymbol{\pi})_{33}\}\,\psi_3 - B\,(\mathbf{K}\cdot\boldsymbol{\pi}\,\beta_0\,\mathbf{K}\cdot\boldsymbol{\pi})_{34}\,\psi_4 = 0,
\end{aligned} \tag{4.6}$$

$$\begin{aligned}
&\{C\,(im + C\,\mathbf{K}_{44}\cdot\boldsymbol{\pi})(\mathbf{K}_{41}\cdot\boldsymbol{\pi}A - C\,\mathbf{K}_{41}\cdot\boldsymbol{\pi}) + imC^2\,\mathbf{K}_{41}\cdot\boldsymbol{\pi} + C^2\mathbf{K}_{41}\cdot\boldsymbol{\pi}[\mathbf{K}_{11}\cdot\boldsymbol{\pi},A\,] + C^2\mathbf{K}_{43}\cdot\boldsymbol{\pi}(\mathbf{K}_{31}\cdot\boldsymbol{\pi}A - B\,\mathbf{K}_{31}\cdot\boldsymbol{\pi})\\
&\quad - eC^2\mathbf{K}_{41}\cdot\mathscr{E}\,A\,\}\,\psi_1 - \{C\,(im + C\,\mathbf{K}_{44}\cdot\boldsymbol{\pi})\,C\,\mathbf{K}_{42}\cdot\boldsymbol{\pi} + C^2\mathbf{K}_{41}\cdot\boldsymbol{\pi}A\,\mathbf{K}_{12}\cdot\boldsymbol{\pi} + C^2\mathbf{K}_{43}\cdot\boldsymbol{\pi}B\,\mathbf{K}_{32}\cdot\boldsymbol{\pi}\}\,\psi_2\\
&\quad + \{C\,(im + C\,\mathbf{K}_{44}\cdot\boldsymbol{\pi})\,\mathbf{K}_{43}\cdot\boldsymbol{\pi} + C^2\,\mathbf{K}_{41}\cdot\boldsymbol{\pi}\,\mathbf{K}_{13}\cdot\boldsymbol{\pi} + C^2\,\mathbf{K}_{43}\cdot\boldsymbol{\pi}\,\mathbf{K}_{33}\cdot\boldsymbol{\pi} - eC^2\mathbf{K}_{43}\cdot\mathscr{E}\}\,B\psi_3\\
&\quad + \{C\,(im + C\,\mathbf{K}_{44}\cdot\boldsymbol{\pi})\,\mathbf{K}_{44}\cdot\boldsymbol{\pi} + C^2\mathbf{K}_{41}\cdot\boldsymbol{\pi}\mathbf{K}_{14}\cdot\boldsymbol{\pi} + C^2\mathbf{K}_{43}\cdot\boldsymbol{\pi}\mathbf{K}_{34}\cdot\boldsymbol{\pi} - eC^2\mathbf{K}_{44}\cdot\mathscr{E}) - m^2\}\,C\psi_4\\
&\quad - C^2(\mathbf{K}\cdot\boldsymbol{\pi}\,\beta_0\,\mathbf{K}\cdot\boldsymbol{\pi})_{43}\,\psi_3 - C^2(\mathbf{K}\cdot\boldsymbol{\pi}\,\beta_0\,\mathbf{K}\cdot\boldsymbol{\pi})\,\psi_4 = 0.
\end{aligned} \tag{4.7}$$

Using Eq. (2.4), Eqs. (4.6) and (4.7) can be re-expressed as

$$\{m^2 - \tfrac{1}{2}ieB\,(\mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K})_{33}\}\,\psi_3 - \tfrac{1}{2}ieB\,(\mathbf{K}\,\beta_0\cdot\mathbf{F}\cdot\mathbf{K})_{34}\,\psi_4 = \cdots, \tag{4.6a}$$

$$\tfrac{1}{2}ieC^2(\mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K})_{43}\,\psi_3 + \tfrac{1}{2}ieC^2(\mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K})_{44}\,\psi_4 = \cdots. \tag{4.7a}$$

(Here the dots stand for terms involving only those parts of ψ for which equations of motion are available.)

In the absence of external fields, Eq. (4.6a) yields ψ_3 in terms of other quantities for which we already have equations of motion and hence time differentiation of Eq. (4.6a) simply yields an equation of motion for ψ_3. When $F_{ij} \neq 0$, a similar situation obtains, with the difference that the combination $[m^2 - \frac{1}{2}ieB\,(\mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K})_{33}\,]\,\psi_3 - \frac{1}{2}ieB\,(\mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K})_{34}\,\psi_4$ of the components of ψ_3 and ψ_4 is involved instead of just ψ_3. Leaving aside for the present the possibility that the number of such independent linear combinations may be less than the number of components of ψ_3 at some particular $F_{\mu\nu}$ (which would mean extra constraints at that $F_{\mu\nu}$), we observe that no further constraints follow from Eq. (4.6a).

Equation (4.7a), on the other hand, leads to tertiary stage constraints on differentiation, if there are no external fields. When $F_{\mu\nu}$ is nonvanishing, however, the presence of the last two terms on the left-hand side of Eq. (4.7)—which reduce to the terms exhibited in Eq. (4.7a)—would prevent the emergence of the corresponding constraints unless certain conditions are satisfied. Before examining what these conditions (for preservation of constraints) are in the most general situation, we consider a relatively simple special case corresponding to β_0 matrices wherein the block B in the form (4.1) is absent. In this case all terms involving B or quantities bearing the subscript 3 would be absent from the various equations. With the penultimate term in Eq. (4.7) thus dropping out, the troublesome term that still remains is the last term. The presence of this term causes an equation of motion (rather than the desired constraint) to result from operating on Eq. (4.7) with π_0, unless $C^2(\mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K})_{44}$ happens to have the form MC (M being some matrix operator). Only if it has this form, or equivalently only if

$$C^2(K\beta_0\cdot F\cdot K)_{44}\;C^2 = 0, \tag{4.8}$$

then the combinations $C\psi_4$ of the components of ψ_4 occur in Eq. (4.7), and Eqs. (4.3) are sufficient to reduce the time derivative of Eq. (4.7) to a constraint. Equation (4.8) is a necessary and sufficient condition for preventing a loss of constraints. It is easily verified that this equation may be transcribed as

$$(1 - \beta_0^2)\,\beta_0^2(\mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K})\,\beta_0^2(1 - \beta_0^2) = 0. \tag{4.9}$$

In the general case when the block B is also present in Eq. (4.1), the situation is much more complicated. It turns out (see Appendix for details) that in addition to the requirement (4.9), the following conditions also have to be obeyed for the preservation of correct number of constraints:

(i) $BX_{33}BB^+$ must be nilpontent:

$(BX_{33}BB^{\dagger})^k = 0$, with k a positive integer: (4.10a)

(ii) for any given k in Eq. (4.10a),

$$(C^2X_{43}BB^{\dagger r})(BX_{33}BB)^{\dagger r}(BX_{34}C^2C^{2\dagger}) = 0, \quad \text{for } r = 0,1,2,\cdots, \quad k-1, \tag{4.10b}$$

where

$$X = \mathbf{K}\beta_0\cdot\mathbf{F}\cdot\mathbf{K}.$$

Having thus obtained the general conditions[33] for the preservation of the number of constraints, we now proceed to examine the role of these conditions in the specific theories cited earlier. We first take up the 30-component theory for spin-2 formulated by Schwinger and Chang (and further studied by Hagen), and show that the freedom from loss of constraints in this case is a reflection of the fact that Eq. (4.9) is obeyed. Next, the Glass equation for spin-$\frac{3}{2}$ is analyzed and violation of the above condition is demonstrated, thus explaining the loss of constraints in this case. Finally, the spin-1 equation given by Shamaly and Capri is investigated and found to obey our conditions.

B. The Schwinger–Chang theory of spin-2

The Schwinger–Chang equation for spin-2 employs a 30-component wave function transforming according to the (reducible) representation

$$T = (1,1)\oplus(\tfrac{3}{2},\tfrac{1}{2})\oplus(\tfrac{1}{2},\tfrac{1}{2})\oplus(\tfrac{1}{2},\tfrac{3}{2})\oplus(0,0) \tag{4.11}$$

of HLG. In the canonical basis in which $\mathbf{J}^2$ and J_z are diagonal, the infinitesimal generators $\mathbf{K}$ of boosts may be partitioned into rows and columns labeled by the "spin" values associated with the eigenvalues of $\mathbf{J}^2$:

$$K = \begin{pmatrix} \mathbf{k}_{22} & \mathbf{k}_{21} & 0 \\ \mathbf{k}_{21}^{\dagger} & \mathbf{k}_{11} & \mathbf{k}_{10} \\ 0 & \mathbf{k}_{10}^{\dagger} & 0 \end{pmatrix} \begin{matrix} \text{spin-2} \\ \text{spin-1} \\ \text{spin-0} \end{matrix} \quad (\text{columns: spin-2, spin-1, spin-0}) \tag{4.12}$$

The explicit forms of $k - s$ are as follows:

$$\mathbf{k}_{22} = \frac{i}{2}\begin{pmatrix} \mathbf{s}_2 & 0 & 0 \\ 0 & \mathbf{s}_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4.13a}$$

$$\mathbf{k}_{11} = \frac{3i}{2}\begin{pmatrix} \mathbf{s}_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{s}_1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{4.13b}$$

$$\mathbf{k}_{21} = \frac{i}{2}\begin{pmatrix} \mathbf{k}_2^{\dagger} & 0 & 0 & 0 \\ 0 & 0 & -\mathbf{k}_2^{\dagger} & 0 \\ 0 & 0 & 0 & \frac{2}{\sqrt{3}}\mathbf{k}_2^{\dagger} \end{pmatrix}, \tag{4.13c}$$

$$\mathbf{k}_{10} = i\begin{pmatrix} 0 & 0 & 0 \\ \mathbf{k}_1^{\dagger} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{2\sqrt{2}}{\sqrt{3}}\mathbf{k}_1^{\dagger} & 0 \end{pmatrix}. \tag{4.13d}$$

$\mathbf{k}_2$ and $\mathbf{k}_1$ are rectangular matrices whose elements connect states of different angular momenta within the same irreducible representation of the Lorentz group.

The matrix β_0 may now be readily constructed. Since $[\beta_0, J] = 0$, β_0 does not link different spin values and is therefore block diagonal in the canonical representation used here. Further, only those elements of β_0 can be nonvanishing which link Lorentz group representations (m,n) and (m',n') such that $m' = m \pm \frac{1}{2}$ and $n' = n \pm \frac{1}{2}$. Consequently, we may write down β_0 as[34]

$$\beta_0 = \begin{pmatrix} X & & \\ & Y & \\ & & Z \end{pmatrix} \begin{matrix} \text{spin-2} \\ \text{spin-1}, \\ \text{spin-0} \end{matrix} \tag{4.14}$$

where

$$X = \begin{pmatrix} 0 & 0 & \beta_{13} \\ 0 & 0 & \beta_{23} \\ \beta_{31} & \beta_{32} & 0 \end{pmatrix}, \tag{4.15a}$$

$$Y = \begin{pmatrix} 0 & 0 & 0 & \beta_{47} \\ 0 & 0 & 0 & \beta_{57} \\ 0 & 0 & 0 & \beta_{67} \\ \beta_{74} & \beta_{75} & \beta_{76} & 0 \end{pmatrix}, \tag{4.15b}$$

$$Z = \begin{pmatrix} 0 & \beta_{89} & \beta_{8,10} \\ \beta_{98} & 0 & 0 \\ \beta_{10,8} & 0 & 0 \end{pmatrix}. \tag{4.15c}$$

Here the "elements" β_{rs} are block matrices which are multiples of identity. The dimensions are

$\beta_{13}, \beta_{31}, \beta_{23}, \beta_{32} \sim 5\times 5$,

$\beta_{47}, \beta_{14}, \beta_{57}, \beta_{67}, \beta_{75}, \beta_{76} \sim 3\times 3$,

$\beta_{89}, \beta_{8,10}, \beta_{98}, \beta_{10,8} \sim 1\times 1$ (numbers).

The condition (2.4) needed for relativistic invariance, namely, $[[K_i, \beta_0], K_j] = \delta_{ij}\beta_0$, imposes the following restrictions on the elements β_{rs}:

$$\begin{aligned} &\beta_{31} = \sqrt{3}\beta_{74}, \quad \beta_{13} = \sqrt{3}\beta_{47}, \\ &\beta_{75} = \frac{\sqrt{2}}{\sqrt{3}}\beta_{98}, \quad \beta_{57} = \frac{\sqrt{2}}{\sqrt{3}}\beta_{89}, \\ &\beta_{32} = \sqrt{3}\beta_{76}, \quad \beta_{23} = \sqrt{3}\beta_{67}. \end{aligned} \tag{4.16}$$

The algebra $\beta_0^5 - \beta_0^3 = 0$ leads to the additional requirements

$$\begin{aligned} &\beta_{31}\beta_{13} + \beta_{32}\beta_{23} = 1, \\ &\beta_{75}\beta_{57} = -\tfrac{1}{3}, \\ &\beta_{98}\beta_{89} = -\tfrac{1}{2}, \\ &\beta_{8,10}\beta_{10,8} = \tfrac{1}{2}. \end{aligned} \tag{4.17}$$

From Eqs. (4.14)–(4.17) we can easily verify the following: (a) X has the minimal equation $X^3 - X = 0$; (b) Y and Z have minimal equations $Y^3 = 0$, $Z^3 = 0$. It is clear from (a) that X is diagonalizable and would in a suitable basis go over into the A and 0 block of Eq. (4.1) while Y and Z together would constitute the C block. The B block is absent.

Having assembled all the materials we need, it is now a simple exercise to verify that

$$\beta_0^2(1-\beta_0^2)K_i\beta_0K_j\beta_0^2(1-\beta_0^2)=0, \tag{4.18}$$

which shows that the condition (4.9) is obeyed. Since, as we have just pointed out, β_0 of the present theory does not involve a B block, the conditions (4.10) drop out altogether. Equation (4.18) is therefore sufficient reason for nonloss of constraints in this theory.

C. Glass' equation for spin-3/2

Glass' equation employs a 20-component wave function ψ which is equivalent in its transformation properties to a vector-spinor together with a Dirac spinor. Thus, ψ transforms according to $T_1 + \bar{T}_1 + 2T_2 + 2\bar{T}_2$,

$$T_1 = (1,\tfrac{1}{2}),$$
$$T_2 = (\tfrac{1}{2},0).$$

The minimal equation obeyed by β_0 in the Glass equation is $\beta_0^5 = \beta_0^3$. The explicit form of β_0 given by Glass is

$$\beta_0 = \begin{pmatrix} P_0 & 0 \\ 0 & -P_0^\dagger \end{pmatrix}, \tag{4.19a}$$

where

$$P_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \\ 0 & -\frac{1}{2\sqrt{3}} & -\frac{1}{2} & 0 \\ 0 & -\frac{\sqrt{2}}{\sqrt{3}} & 0 & 1 \end{pmatrix}. \tag{4.19b}$$

Here P_0 is a 10×10 submatrix partitioned into 4 +2 +2 +2 subblocks. The infinitesimal generators of boost are

$$\mathbf{K} = -i\begin{pmatrix} 0 & N \\ N & 0 \end{pmatrix}, \tag{4.20a}$$

where

$$\mathbf{N} = \begin{pmatrix} -\frac{1}{3}\boldsymbol{\Sigma} & \mathbf{u} & 0 & 0 \\ \mathbf{u}^\dagger & -\frac{5}{6}\boldsymbol{\sigma} & 0 & 0 \\ 0 & 0 & -\frac{1}{2}\boldsymbol{\sigma} & 0 \\ 0 & 0 & 0 & -\frac{1}{2}\boldsymbol{\sigma} \end{pmatrix}. \tag{4.20b}$$

The matrices $\mathbf{u}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\sigma}$ are those defined earlier (Sec. 3). A straightforward calculation of $F_{ij}\,\beta_0^2(1-\beta_0^2)(K_i\beta_0K_j - K_j\beta_0K_i)\beta_0^2(1-\beta_0^2)$ shows that it is a nonnull matrix whose nonzero elements are proportional to $\boldsymbol{\sigma}\cdot\mathscr{H}$, thereby violating our condition (4.9). This is indeed the reason why the Glass equation suffers from a loss of constraints when there is a coupling to a nonzero magnetic field.

D. Spin-1 equation of Shamaly and Capri

The spin-1 equation given by Shamaly and Capri employs a 20-component wave function ψ which is equivalent in its transformation properties to a general second rank tensor together with a vector. Thus, ψ transforms according to the representation

$$(1,1)\oplus(1,0)\oplus(0,1)\oplus(\tfrac{1}{2},\tfrac{1}{2})\oplus(0,0)$$
$$\equiv T_1 + T_2 + \bar{T}_2 + T_3 + T_4\,.$$

In the canonical representation of the Lorentz generators, we partition $\mathbf{K}$ as follows:

$$\mathbf{K} = -i\begin{pmatrix} 0 & \mathbf{K}_{21} & 0 \\ \mathbf{K}_{21}^\dagger & \mathbf{K}_{11} & \mathbf{K}_{10} \\ 0 & \mathbf{K}_{10}^\dagger & 0 \end{pmatrix}, \tag{4.21}$$

where

$$\mathbf{K}_{21} = \overset{\begin{matrix} T_2 & \quad T_2 & \quad\;\; T_1 & \quad\; T_3 \end{matrix}}{\begin{pmatrix} 0 & 0 & \frac{1}{\sqrt{3}}\mathbf{k}_2^\dagger & 0 \end{pmatrix}} \, T_1\,, \tag{4.22a}$$

$$\mathbf{K}_{11} = \begin{pmatrix} \mathbf{s}_1 & 0 & 0 & 0 \\ 0 & -\mathbf{s}_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{matrix} T_2 \\ \bar{T}_2 \\ T_1 \\ T_3 \end{matrix}\,, \tag{4.22b}$$

$$\mathbf{K}_{10} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 2\sqrt{2}\,\mathbf{k}_1^\dagger & 0 \\ 0 & 0 & \mathbf{k}_1^\dagger \end{pmatrix}. \tag{4.22c}$$

$\mathbf{s}_1$, $\mathbf{k}_2$, and $\mathbf{k}_1$ are the same matrices which were introduced in Sec. 4.B. (The representations of HLG to which each block belongs is also indicated.)

Specification of β_0 now proceeds as in the last section. In the canonical representation it has the form

$$\beta_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & Y \end{pmatrix}, \tag{4.23}$$

where X is a 12×12 matrix

$$X = \begin{pmatrix} 0 & 0 & 0 & \alpha_1 \\ 0 & 0 & 0 & \alpha_2 \\ 0 & 0 & 0 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 & 0 \end{pmatrix}, \tag{4.24a}$$

with elements which are multiples of the unit matrix of dimension 3, and Y is a 3×3 matrix

$$Y = \begin{pmatrix} 0 & 0 & \delta_1 \\ 0 & 0 & \delta_2 \\ \delta_3 & \delta_4 & 0 \end{pmatrix}. \tag{4.24b}$$

Here, $\alpha_1\alpha_2\cdots\alpha_6$, δ_1, δ_2, δ_3, and δ_4 are arbitrary constants.

Equation (2.4) imposes the conditions

$$\sqrt{5}\delta_2 = \sqrt{6}\alpha_3\,, \quad \sqrt{5}\delta_4 = \sqrt{6}\alpha_6\,, \tag{4.25}$$

while the additional restrictions

$$\alpha_1\alpha_4 + \alpha_5\alpha_2 + \alpha_3\alpha_6 = 1,$$
$$\delta_3\delta_1 + \delta_4\delta_2 = 0, \tag{4.26}$$

are imposed by the algebra $\beta_0^5 = \beta_0^3$. Armed with this knowledge, one can readily verify that the minimal equations of X and Y are $(X^3 - X) = 0$ and $Y^3 = 0$, respectively, and hence that no part with the minimal equation $B^2 = 0$ is involved in β_0. Consequently, the conditions (4.10) become irrelevant. As for the condition (4.9), it is easily verified using Eqs. (4.21)–(4.26) that $(1-\beta_0^2)\beta_0^2 K_i\beta_0 K_j$ itself is a null matrix. This explains why the minimal electromagnetic interaction does not cause any change in the number of constraints in the case of the Shamaly–Capri equation.

The causality of propagation of the equations belonging to this algebra still remains an open question. As before, we can establish some connection between the causality of propagation on the one hand and the loss of constraints on the other. A sufficient condition for causality in equations obeying this algebra has been given recently by Khalil, as

$$\Big\{\sum_{\mathscr{P}} (\beta_{\mu_1}\beta_{\mu_2}\beta_{\mu_3} - g_{\mu_1\mu_2}\beta_{\mu_3})\Big\}\beta_{\mu_4}\beta_{\mu_5} = 0. \tag{4.27}$$

On taking $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$ and $\mu_5 = i$, Eq. (4.27) gives

$$(1-\beta_0^2)\beta_0^2\beta_i = 0, \tag{4.28}$$

which may be further reduced to

$$(1-\beta_0^2)\beta_0^2 K_i\beta_0 = 0, \tag{4.29}$$

by using the fact that $\beta_i = i[\beta_0, K_i]$ and $\beta_0^5 - \beta_0^3 = 0$. Comparison of Eq. (4.29) with our condition (4.9) shows that any equation obeying Khalil's condition for causality will automatically be free of the pathology of loss of constraints. Thus, Khalil's condition for causality is also a sufficient condition for nonloss of constraints. It may be noted that even the less stringent condition

$$\Big\{\sum_{\mathscr{P}} (\beta_{\mu_1}\beta_{\mu_2}\beta_{\mu_3}\beta_{\mu_4} - g_{\mu_1\mu_2}\beta_{\mu_3}\beta_{\mu_4})\Big\}\beta_{\mu_5} = 0 \tag{4.30}$$

is sufficient to prevent loss of constraints since it leads to Eq. (4.28).

5. RESULTS AND DISCUSSION

The work presented in the foregoing sections serves to emphasize and define more clearly than hitherto the link between the troubles of higher spin unique-mass wave equations and the algebra of the β matrices. The value of n in the minimal equation $\beta_0^n = \beta_0^{n-2}$ (the condition for unique mass) determines the number of levels of constraints in the wave equation which is strongly linked to the variety of consistency problems encountered in the theory. The types of difficulties which have come to light in the course of investigation of specific theories include (i) noncausal propagation in the presence of external fields, (ii) occurrence of modes of complex frequency when high magnetic fields are present, (iii) appearance of extra constraints at particular values of the external fields, (iv) loss of constraints, and (v) unacceptable changes in the commutation rules for field components on introduction of interactions, in addition to (vi) the problem of possible indefiniteness of charge/energy in the free theory itself.

When the degree n of the minimal equation goes up to three (i.e., $\beta_0^3 = \beta_0$), the wave equation involves just primary constraints, but none of the troubles (i) to (v) arises as for as in known[35]; in fact, it is known that whenever β_0 is diagonalizable (even in multimass theories), (i) and (ii)–(v) cannot occur.[7,17,36,37] However, as a special case of a result of general validity proved by Johnson and Sudarshan, the absence of secondary constraints in this case leads to the problem (vi), and thereby to difficulties in quantization.[38] With $n = 4$, there are both primary and secondary constraints—and the difficulties of the types (i)–(iii) and (v) appear for the first time. [Type (iv) loss of constraints does not occur until n goes up to 5 bringing in tertiary constraints.] However, theories characterized by $\beta_0^4 = \beta_0^2$ or $\beta_0^5 = \beta_0^3$ do exist in which the introduction of minimal electromagnetic interaction does not cause any change in the number of constraints. We have set up in this paper criteria by which such theories can be identified. Theories which fail by these criteria may still be made acceptable in the matter of constraints by introducing suitable nonminimal terms in the electromagnetic interaction. A general investigation of the possibilities in this respect is currently in progress. Those theories which satisfy the criteria for conservation of constraints may still be unsatisfactory in other respects, e.g., propagation of the field in question may be acausal in the presence of interactions. We have noted that a sufficient condition for causality, obtained by Khalil, does imply also conservation of constraints.[39] The interesting work of Cox on the causality question may be mentioned here,[30] but a necessary and sufficient general condition for causality is yet to be obtained. Until investigations on this and other question (such as the possibility of tachyonic modes) are completed, it will not be possible to rule out the existence of consistent higher spin theories.

ACKNOWLEDGMENT

One of us (TRG) would like to record his gratitude to University Grants Commission, India for financial support in the form of a fellowship.

APPENDIX

We present here the proof of the conditions (4.10) for general β_0. We first note that a basis can always be found wherein the matrices B and C of Eqs. (4.1) have the partitioned forms

$$B = \begin{pmatrix} 0 & I \\ 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & I & 0 \\ 0 & 0 & I \\ 0 & 0 & 0 \end{pmatrix}. \tag{A1}$$

Here the blocks are square (null and unit) submatrices. We use these forms in the following, and also break down the parts ψ_3 and ψ_4 of the wavefunction ψ further, so as to conform to the partitioning (A1) of B and C:

$$\psi_3 = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}, \quad \psi_4 = \begin{pmatrix} \chi_1 \\ \chi_2 \\ \chi_3 \end{pmatrix}. \tag{A2}$$

One sees readily that equations of motion follow for ϕ_2 from Eq. (4.3c) and for χ_2 and χ_3 from Eq. (4.3d). As for φ_1 and

χ_1, they do not figure in these equations, but do in Eqs. (4.6) and (4.7). Partioning the blocks X_{33}, X_{34}, X_{43}, and X_{44} of $X = \frac{1}{2} ie\mathbf{K}\beta_0 \cdot \mathbf{F} \cdot \mathbf{K}$ which appear in the latter equations into subblocks α_{ij}, β_{ij}, etc. as

$$X_{33} = (\alpha_{ij}), \quad i,j = 1,2; \; X_{34} = (\beta_{ij}), \quad i = 1,2, \; j = 1,2,3;$$
$$X_{43} = (\gamma_{ij}), \quad i = 1,2,3, \; j = 1,2; \; X_{44} = (\delta_{ij}), \quad i,j = 1,2,3; \qquad \text{(A3)}$$

we find the relevant equations to be

$$(\alpha_{21} + m^2)\varphi_1 + \beta_{21}\chi_1 + \cdots = 0, \qquad \text{(A4a)}$$

$$\gamma_{31}\varphi_1 + \delta_{31}\chi_1 + \cdots = 0. \qquad \text{(A4b)}$$

The dots stand for terms involving other parts of the wave function for which equations of motion have already been obtained.

In the absence of external fields, the equation of motion for φ_1 is obtained by solving Eq. (A4a) for φ_1 and differentiating with time; on the other hand, Eq. (A4b) would reduce to a constraint (not involving χ_1) from which an equation of motion for χ_1 follows, nevertheless, on differentiating twice. We require that the same scheme should prevail also in the presence of arbitrary external fields. Inspection of Eq. (A4a) shows that if the matrix operator $(\alpha_{21} + m^2)$ acting on φ_1 is to be nonsingular for all external fields, the term α_{21} which is linear in the field must be nilpotent, i.e.,

$$(\alpha_{21})^k = 0 \qquad \text{(A5)}$$

for some integer k, for all configurations of the field $F_{\mu\nu}$. Once this is satisfied, the explicitly written terms in Eq. (A4b) can be expressed as

$$[-\gamma_{31}(\alpha_{21} + m^2)^{-1}\beta_{21} + \delta_{31}]\chi_1 = \cdots. \qquad \text{(A6)}$$

If the χ_1 term is to be absent from Eq. (A4b) for all external fields (as when $F_{\mu\nu} = 0$), the square-bracketed operator in Eq. (A6) must vanish order by order in the external field strength. Since each of the quantities $\gamma_{31}, \alpha_{21}, \beta_{21}$, and δ_{31} is of first order in the field, this condition requires that

$$\delta_{31} = 0 \qquad \text{(A7a)}$$

and

$$\gamma_{31}(\alpha_{21})^r \beta_{21} = 0, \quad r = 0,1,\dots,k-1. \qquad \text{(A7b)}$$

[These, like Eq. (A5), must hold for all configurations of the external field.] Using the definitions (A1) and (A3), one can readily express the conditions (A5) and (A7) in terms of the larger blocks $X_{33}, X_{34}, \cdots$ of X as

$$(BX_{33}BB^\dagger)^k = 0,$$
$$(C^2X_{43}BB^\dagger)(BX_{33}BB^\dagger)^r(BX_{34}C^2C^{2\dagger}) = 0,$$
$$r = 0,1,\dots,k-1. \qquad \text{(A8)}$$

Finally, these in turn can be written directly in terms of $X \equiv \frac{1}{2}ie\mathbf{K}\beta_0 \cdot \mathbf{F} \cdot \mathbf{K}$ (rather than in terms of its partitioned blocks), with the help of projection operators P_B and P_C which isolate the B and C blocks, respectively, of β_0:

$$P_B\beta_0 = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & B & \\ & & & 0 \end{pmatrix},$$

$$P_C\beta_0 = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & 0 & \\ & & & C \end{pmatrix}. \qquad \text{(A9)}$$

It is readily verified that with B and C as in Eq. (A1), P_B and P_C are given by

$$P_C = P^2P^{\dagger 2} + P^{\dagger 2}P^2 + \tfrac{1}{2}(PP^{\dagger 2}P + P^\dagger P^2 P^\dagger),$$
$$P_B = (1 - P_C)(PP^\dagger + P^\dagger P), \qquad \text{(A10)}$$

respectively, where

$$P = (1 - \beta_0^2)\beta_0.$$

With the aid of these, Eqs. (A8) may be written as

$$(P_B\beta_0\, X\beta_0\beta_0^\dagger P_B)^k = 0,$$
$$(1 - \beta_0^2)\beta_0^2\, K\beta_0 \cdot F \cdot K\, \beta_0^2\, (1 - \beta_0^2) = 0,$$
$$(P_C\, \beta_0^2 X\beta_0\beta_0^\dagger P_B)(P_B\, \beta_0 X\beta_0\beta_0^\dagger P_B)^r(P_B X\beta_0\beta_0^\dagger\, P_C) = 0,$$
$$r = 0,1,2,\dots,k-1. \qquad \text{(A11)}$$

It may be noted that Eqs. (A8)–(A11) make use of the form (A1) of the B and C blocks of β_0. Actually, it is not difficult to see that the same results hold good even if the actual form of β_0 is related to the type of standard form considered here by a unitary transformation. The restriction to such representations of β_0 is not a serious one—it is simply the counterpart of the familiar assignment of hermiticity properties to the matrices in the Dirac theory.

[1] E.M. Corson, *Introduction to Tensors, Spinors and Relativistic Wave Equations* (Hafner, New York, 1953); H. Umezawa, *Quantum Field Theory* (North Holland, Amsterdam, 1956); Y. Takahashi, *An Introduction to Field Quantization* (Pergammon, Oxford, London, 1969); P.M. Mathews, "Wave Equations for Particles of Arbitrary Spin: Relativistic Invariance and Quantization," Lectures presented at the Institute for Theoretical Physics, University of Colorado, Summer 1969.

[2] K. Johnson and E.C.G. Sudarshan, Ann. Phys. (N.Y.) **13**, 126 (1961).

[3] G. Velo and D. Zwanziger, Phys. Rev. **186**, 1337 (1969).

[4] G. Velo and D. Zwanziger, Phys. Rev. **188**, 2218 (1969).

[5] A.S. Wightman, *Aspects of Quantum Theory*, edited by A. Salam and E.P. Wigner (Cambridge University, London, New York, 1971); see also review talks in A.S. Wightman, *Troubles in the External Field Problem for Invariant Wave Equations* (Gordon and Breach, New York, 1971).

[6] P.M. Mathews Phys. Rev. D **9**, 365 (1974); W.-y. Tsai and A. Yildiz, *ibid.* **4**, 3643 (1971); W.-y. Tsai, *ibid.* **7**, 1945 (1973).

[7] J. Prabhakaran, M. Seetharaman, and P.M. Mathews, J. Phys. A **8**, 560 (1975).

[8] P.M. Mathews, M. Seetharaman, and J. Prabhakaran, Phys. Rev. D **14**, 1021 (1976).

[9] N. Fierz and W. Pauli, Proc. R. Soc. (London) Ser. A **173**, 211 (1939).

[10] P. Federbush, Nuovo Cimento **19**, 572 (1961); see also Ref. 32.

[11] P.M. Mathews, M. Seetharaman, T.R. Govindarajan, and J. Prabhakaran, Phys. Rev. D (in press).

[12] A.S. Glass, Commun. Math. Phys. **23**, 176 (1971).

[13] W. Rarita and J. Schwinger, Phys. Rev. **60**, 61 (1941).

[14] J.D. Jenkins, J. Phys. A **5**, 1461 (1972); **6**, 1211 (1973); M. Seetharaman, J. Prabhakaran, and P.M. Mathews, Phys. Rev. D **12**, 458 (1975).

[15] V. Amar and U. Dozzio, Nuovo Cimento Lett. **5**, 355 (1972).

[16] V. Amar and U. Dozzio, Nuovo Cimento Lett. **12**, 659 (1975).

[17] J. Prabhakaran, T.R. Govindarajan, and M. Seetharaman, Nucl. Phys. B **127**, 537 (1977).

[18] H.J. Bhabha, Philos. Mag. **43**, 33 (1952); K.K. Gupta, Proc. R. Soc. (London) Ser. A **122**, 118 (1954).

[19] I.M. Gel'fand, R.A. Minlos, and Z.ya Schapirio, *Representations of the Rotation and Lorentz Groups and their Applications* (Pergamon, Oxford, 1963).

[20]M.A.K. Khalil, Phys. Rev. D **15**, 1532 (1977).
[21]A. Shamaly and A.Z. Capri, Can. J. Phys. **51**, 1467 (1973).
[22]Harish Chandra, Phys. Rev. **71**, 793 (1947).
[23]W.J. Hurley and E.C.G. Sudarshan, Ann. Phys. (Leipzig) **85**, 546 (1974).
[24]N. Kemmer, Proc. R. Soc. (London) Ser. A **173**, 91 (1939).
[25]In general, the β algebra (2.5b) for $n > 2$ leads to an infinite order algebra and hence the standard scheme applicable to semisimple finite algebras to enumerate representations cannot be carried out.
[26]Any matrix β_0 having the minimal equation $\beta_0^2(\beta_0^2 - 1) = 0$ may be brought by a similarity transformation to the form we have chosen. A is reducible further to a diagonal form with elements ± 1 and B can be reduced to a direct sum of Jordan blocks of dimension two with eigenvalue 0.
[27]Note that in this case, β_0 has no zero blocks on the diagonal. Consequently ψ_2, K_{2i}, and K_{i2} are not defined and terms involving any of those are to be dropped from Eqs. (3.4) and (3.5).
[28]A.S. Glass, Ph.D. thesis, Princeton University, 1971 (unpublished).
[29]M.A.K. Khalil, Prog. Theor. Phys. **58**, 1538 (1977).
[30]W. Cox, J. Phys. A **11** 1167 (1978).
[31]J. Schwinger, Phys. Rev. **130**, 800 (1963); S.J. Chang, Phys. Rev. **161**, 1308 (1967).
[32]C.R. Hagen, Phys. Rev. D **4**, 2204 (1971).
[33]Equations (4.10) presuppose that β_0 is in Jordan canonical form, or related to it by a unitary transformation. However, Eq. (4.9), like Eq. (3.13), is not subject to any such restriction.
[34]The reader is cautioned that the blocks in Eq. (4.14), which are labeled by the spin values, are not necessarily characterized by simple minimal equations. This representation of s_0 is quite different from Eq. (4.13), where the minimal equations were the distinguishing characteristics of the different blocks.
[35]However, even in the simplest of the theories of this category, namely Duffin–Kemmer theory for spin-1, there is no completely satisfactory way of coupling to the eletromagnetic field; see R.J. Duffin, Phys. Rev. **84**, 1114 (1938); B. Vijayalakshmi, M. Seetharaman, and P.M. Mathews, J. Phys. A **12**, 665 (1979); see also Ref. 24.
[36]Singh and Hagen have recently claimed that loss of constraints occurs in the case of Rarita–Schwinger equation if light front coordinates are used. However, the significance of constraints in such coordinates is unclear. In any case, the occurrence of the constraint loss can be understood in terms of the algebra of the β matrices which are relevant to the light front coordinates. See L.P.S. Singh and C.R. Hagen, Phys. Rev. D **16**, 347 (1977).
[37]A.K. Nagpal, Nucl. Phys. B **53**, 634 (1973); **64**, 544 (1975).
[38]See, however, D. Barua and S.N. Gupta, Phys. Rev. D **17**, 2028 (1978). They suggest a way of quantizing with indefinite metric.
[39]Note that the condition (4.30) we have obtained for nonloss of constraints in the case of $\beta_0^5 = \beta_0^3$ is weaker than Khalil's causuality condition. The former is satisfied if the wave equation has a "barnacled" structure built from basic equations which involve β_0 satisfying $\beta_0^n = \beta_0^{n-2}$, with $n = 4$ or 3. Thus, all such barnacled equations with minimal coupling are free of the problem of loss of constraints. On eliminating the barnacles from these equations, one gets equations (characterized by $n = 3$ or 4 as the case may be) which include nonminimal terms, and are nevertheless by construction free of constraints pathology.

Extended phase space. I. Classical fields

A. Das
Department of Mathematics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

(Received 30 October 1979; accepted for publication 17 January 1980)

Classical field theory is developed in the arena of extended phase space V_8, the space of position, time, momentum, and energy. This enables one to incorporate Born's reciprocity which demands equal status for the variables q and p. The present formulation is covariant under the extended Poincaré group P_8 acting in V_8. Variational methods for classical field theory are generalized. Besides the usual concept of the total 4-momentum, one encounters the notion of average position and time of the field distributions. The total charge emerges from a dynamical viewpoint. The Dirac and Duffin–Kemmer algebras are generalized in this setting. The corresponding wave equations would lead to a dynamical theory of the elementary particles. The symplectic structure is not considered because of the difficulties to represent spinors.

I. INTRODUCTION

In classical Hamiltonian mechanics and the subsequent quantum particle mechanics the basic equations remain intact under the substitutions $q' = -p, p' = q$. Born[1] and Landé[2] advocated this covariance (which was called the reciprocity principle) for quantum field theory. Yukawa[3] incorporated this idea of reciprocity in his formulation of nonlocal fields. In spite of these brilliant attempts, the simple and potent idea of reciprocity has not been pursued for the last decade.

The present author[4] formulated field theory in 4-dimensional complex space–time in order to obtain quantized or discrete space–time and at the same time to utilize the larger group of covariance for the interpretation of isotopic or internal groups.

In this series of papers the 4-dimensional complex space–time is reinterpreted as the extended phase space V_8. Also the idea of reciprocity is generalized to the group of transformations $\mathbb{P}: q' = q\cos\phi - p\sin\phi, p' = q\sin\phi + p\cos\phi$. The group $\mathbb{P}$ and the Poincaré group P_4 are subgroups of the 36-parameter extended Poincaré group P_8. That is why a field theory covariant under P_8 is developed in the present paper. It may be mentioned that space reflection, time reversal and Born reciprocity are special cases of the group $\mathbb{P}$.

In Sec. III the Lagrangian formalism is presented in detail. The various Noether's theorems are derived. The integral constants of a field are constructed. Among these is of course the usual total 4-momentum of the field. In addition, another 4-vector emerges which can be conveniently interpreted as the average position and time of the total field. The total charge, which is one of the integral constants, has a dynamical significance. In each of these four phase planes of V_8 a field quanta can have rotational motions, both orbital and spin. These motions viewed in space or time are nothing but the space or time oscillations or vibrations or tremors.[5]

The total charge is the sum of the total angular momentum of the field in each of these four phase planes or the total tremor of the field. The orbital part of the tremor corresponds to the isotopic contribution T to the total charge Q and the intrinsic part corresponds to the baryonic contribution $\frac{1}{2}B$ of the toal charge to make $Q = T + \frac{1}{2}B$. In Sec. IV the matrix-wave equations covariant under the extended Lorentz group L_8 are studied. The most general multiplication rules among eight matrices α^A are derived. The generalizations of Dirac and Duffin–Kemmer algebras are obtained.

In Sec. VI the Lagrangian formalism is applied to a matrix-wave field. The Green's functions are exhibited. The conserved integral constants of a matrix-wave field are determined.

The Appendix provides a brief review of the Lie algebra D_4, which is pertinent in the present theory.

II. NOTATIONS AND THE EXTENDED POINCARÉ GROUP

The extended phase space V_8 is a flat manifold of space–time–momentum–energy, where the points have real coordinates $(q^1,q^2,q^3,q^4,p^1,p^2,p^3,p^4)$ or, in short (q,p), which physically represnts the possible occupation of a particle at the event (q) with 4-momentum (p). Small italic indices take the values 1, 2, 3, 4, capital indices take the values 1, 2,...,8, and small Greek indices take the value 1, 2, 3. The summation convention is followed. The units are so chosen that $a = b = c = 1$, where a, b are the fundamental length and momentum, respectively, $(\hbar = ab)$, and c is the velocity of light. All other physical quantities are expressed as pure numbers.

The metric form of V_8 is assumed to be

$$\Phi \equiv \eta_{ij}(dq^i\, dq^j + dp^i\, dp^j), \tag{2.1}$$

where $\eta \equiv [\eta_{ij}] \equiv \mathrm{diag}(\epsilon(i)) = \mathrm{diag}(-1^3, +1)$ and thus signature of V_8 is -4.

The homogeneous linear transformations in V_8 which leave Φ invariant are the following:

$$q'^i = a^i_j q^j + b^{\,i}_{\;j}\, p^j, \qquad p'^i = c^i_j q^j + d^{\,i}_{\;j}\, p^j, \tag{2.2}$$

where

$$\eta_{ij}(a^i_k a^j_l + c^i_k c^j_l) = \eta_{ij}(b^{\,i}_{\;k} b^{\,j}_{\;l} + d^{\,i}_{\;k} d^{\,j}_{\;l}) = \eta_{kl}, \qquad \eta_{ij}(a^i_k b^{\,j}_{\;l} + c^i_k d^{\,j}_{\;l}) = 0. \tag{2.3}$$

 0022-2488/80/061506-07$1.00

Defining the matrices $A\equiv[a^i_j]$, $B\equiv[b^{\,i}_j]$, $C\equiv[c^i_j]$, $D\equiv[d^{\,i}_j]$ and denoting by A^{tr} the transposition of A, (2.3) can be written as

$$A^{\text{tr}}\eta A + C^{\text{tr}}\eta C = B^{\text{tr}}\eta B + D^{\text{tr}}\eta D = \eta,$$
$$A^{\text{tr}}\eta B + C^{\text{tr}}\eta D = 0. \tag{2.3'}$$

The above equations show that the transformations (2.2) generate a group involving 28 real parameters. This group contains the Lorentz group as a subgroup whenever $A = D$, $B = C = 0$ and thus will be called the extended Lorentz group L_8 (see the Appendix). This group also contains a 4-parameter Abelian subgroup of reciprocity [which is isomorphic to $(U_1\times U_1\times U_1\times U_1)$] and is given by

$$q'^a = q^a\cos\phi^{\,a} - p^a\sin\phi^{\,a},$$
$$\mathbb{P}\colon p'^a = q^a\sin\phi^{\,a} + p^a\cos\phi^{\,a} \quad \text{(no summation)}, \tag{2.4}$$
$$q'^a + i\,p'^a = (q^a + ip^a)e^{i\phi^a}.$$

The original Born reciprocity $\mathbb{P}_0$ can be recovered from (2.4) by putting $\phi^{\,a} = \pi/2$, i.e.,

$$\mathbb{P}_0\colon \begin{aligned} q'^a &= -p^a,\\ p'^a &= q^a. \end{aligned} \tag{2.4a}$$

One possible way of generating the usual reflection symmetries is via the continuous group $\mathbb{P}$. For example, the space reflection π_α can be generated by $\phi^{\,\alpha} = \pi$, $\phi^{\,\beta} = 0$ for $\beta\neq\alpha$, and $\phi^{\,4} = 0$ (see Fig. 1), i.e.,

$$\begin{aligned} \pi_\alpha\colon q'^\alpha &= -q^\alpha, \quad q'^\beta = q^\beta,\\ q'^4 &= q^4,\\ p'^\alpha &= -p^\alpha, \quad p'^\beta = p^\beta, \quad (\alpha\neq\beta),\\ p'^4 &= p^4. \end{aligned} \tag{2.4b}$$

The total space reflection is $P\equiv\pi_1\pi_2\pi_3$. The time reversal (the present definition differs from the usual time reversal T) T' can be generated by putting $\phi^{\,\alpha} = 0$, $\phi^{\,4} = \pi$, i.e.,

$$T'\colon q'^\alpha = q^\alpha, \quad q'^4 = -q^4, \quad p'^\alpha = p^\alpha, \quad p'^4 = -p^4. \tag{2.4c}$$

It can be noted that

$$PT' = \mathbb{P}_0^2\,. \tag{2.4d}$$

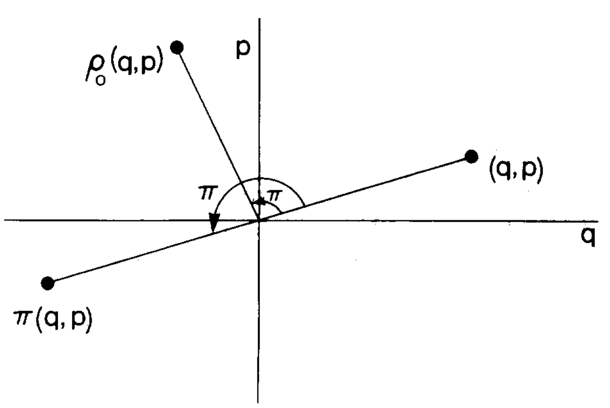

FIG. 1. (q-p)-phase plane.

The inhomogeneous transformations which leave Φ invariant are

$$P_8\colon \begin{aligned} q'^i &= b^{\,i} + a^i_j q^j + b^{\,i}_j p^j,\\ p'^i &= d^{\,i} + c^i_j q^j + d^{\,i}_j p^j, \end{aligned} \tag{2.5}$$

where $b^{\,i}$, $d^{\,i}$ are arbitrary real parameters and the remaining parameters satisfy (2.3). This group of transformations involve 36 real parameters and will be called the extended Poincaré group P_8.

Before the tensor representations of the group (2.2) or (2.5) are defined the following notations would be advantageous.

$$\begin{aligned} (\xi) &\equiv (\xi^1,\xi^2,\xi^3,\xi^4,\xi^5,\xi^6,\xi^7,\xi^8)\\ &\equiv (q^1,q^2,q^3,q^4,p^1,p^2,p^3,p^4),\\ \mathcal{N} &\equiv [\mathcal{N}_{AB}] \equiv \begin{bmatrix} \eta & 0\\ 0 & \eta \end{bmatrix},\\ \mathbb{L} &\equiv [l^A_B] \equiv \begin{bmatrix} A & B\\ C & D \end{bmatrix},\\ [l^A] &\equiv [b^{\,i}, d^{\,i}]. \end{aligned} \tag{2.6}$$

Now the Eqs. (2.1), (2.2), (2.3), (2.5) can be expressed neatly as

$$\Phi = \mathcal{N}_{AB}\,d\xi^A\,d\xi^B, \quad \xi'^A = l^A_B\xi^B, \quad \mathbb{L}^{\text{tr}}\mathcal{N}\mathbb{L} = \mathcal{N},$$
$$\det[\mathbb{L}] = +1, \quad \xi'^A = l^A + l^A_B\xi^B. \tag{2.7}$$

Now the tensor field representations under transformations (2.7) can be defined by the following:

$$T'^{A\cdots}_{B\cdots}(\xi') = l^A_C..l_B^{-D}..T^{C\cdots}_{D\cdots}(\xi),$$
$$[l_B^{-D}] \equiv [l^D_B]^{-1}. \tag{2.8}$$

Examples of the simple tensor fields are the following:

(i) Scalar field: $\phi'(\xi') = \phi(\xi)$, or, $\phi'(q',p') = \phi(q,p)$;
(ii) Vector field: $A'^B(\xi') = l^B_C A^C(\xi)$. (2.8′)

The above can also be expressed in terms of two 4-dimensional vector fields.

$$\begin{aligned} (A^{q^i}) &\equiv (A^1,A^2,A^3,A^4), \quad (A^{p^i}) \equiv (A^5,A^6,A^7,A^8),\\ A'^{q^i}(q',p') &= a^i_j A^{q^j}(q,p) + b^{\,i}_j A^{p^j}(q,p),\\ A'^{p^i}(q',p') &= c^i_j A^{q^j}(q,p) + d^{\,i}_j A^{p^j}(q,p). \end{aligned} \tag{2.8''}$$

(Summation convention still operating on index j above.) The raising and lowering of captical italic indices can be performed as follows:

$$[\mathcal{N}^{AB}] \equiv [\mathcal{N}_{AB}],$$
$$T_A \equiv \mathcal{N}_{AB}T^B, \quad T^A = \mathcal{N}^{AB}T_B\,. \tag{2.9}$$

An alternate way of raising and lowering indices is to introduce the totally antisymmetric Levi–Civita tensor $\epsilon_{ABCDEFGH}\equiv\pm 1{,}0$ in the usual manner. From an rth order tensor $T^{A_1\cdots A_r}$ a dual tensor of order $8-r$, $(0\leqslant r\leqslant 8)$ can be defined as

$$\check{T}_{A_{r+1}A_{r+2}\cdots A_8} \equiv \epsilon_{A_1\cdots A_r\cdots A_8}T^{A_1\cdots A_r}. \tag{2.10}$$

The expressions like

$$T_A T^A = \eta_{ij}(T^{q^i}T^{q^j} + T^{p^i}T^{p^j}).$$
$$U_{AB}V^{AB} = U_{q^iq^j}V^{q^iq^j} + U_{q^ip^j}V^{q^ip^j} + U_{p^iq^j}V^{p^iq^j} + U_{p^ip^j}V^{p^ip^j},$$

are invariant under the extended Poincaré group P_8.

III. THE LAGRANGIAN FORMALISM

The infinitesimal version of the extended Poincaré group P_8 can be summed up as

$$\xi'^A = \epsilon^A + (\delta^A_B + \epsilon^A_B)\xi^B, \quad \epsilon_{AB} = -\epsilon_{BA}, \qquad \delta\xi^A \equiv \xi'^A - \xi^A = \epsilon^A + \epsilon^A_B\,\xi^B, \tag{3.1}$$

where $|\epsilon^A|$, $|\epsilon^A_B|$ are small positive numbers.

Under (3.1) a tensor field $\phi^{A\cdot\cdot}(\xi)$ transform infinitesimally and the local variation is

$$\delta\phi^{A\cdot\cdot} \equiv \phi'^{A\cdot\cdot}(\xi') - \phi^{A\cdot\cdot}(\xi) = (\tfrac{1}{2})\epsilon^{CD}S_{CD,B\cdot\cdot}{}^{A\cdot\cdot}\phi^{B\cdot\cdot}(\xi), \tag{3.2}$$

neglecting $O(\epsilon^2)$ terms. Here $S_{CD,}{}^{A\cdot\cdot}{}_{B\cdot\cdot}$ is a numerical tensor representing generalized "spin." The substantial variation is defined as

$$\begin{aligned}\tilde{\delta}\phi^{A\cdot\cdot} &\equiv \phi'^{A\cdot\cdot}(\xi) - \phi^{A\cdot\cdot}(\xi)\\ &= -\epsilon^B\partial_B\phi^{A\cdot\cdot} + (\tfrac{1}{2})\epsilon^{CD}\,[S_{CD,B\cdot\cdot}{}^{A\cdot\cdot}\phi^{B\cdot\cdot}\\ &\quad + (\xi_C\partial_D - \xi^D\partial_D)\phi^{A\cdot\cdot}],\end{aligned} \tag{3.3}$$

where $\partial_A \equiv \partial/\partial\xi^A$.

The invariant action integral for a complex tensor field $\phi^{A\cdot\cdot}$, $\bar{\phi}^{A\cdot\cdot}$ (the bar stands for the complex conjugation) can be defined as

$$A \equiv \int_D d^8\xi\, L\,[\phi^{A\cdot\cdot}(\xi),\bar{\phi}^{A\cdot\cdot}(\xi),\partial_B\phi^{A\cdot\cdot},\partial_B\bar{\phi}^{A\cdot\cdot}], \tag{3.4}$$

where D is a simply connected domain in V_8, bounded by a piecewise smooth, orientable, closed, compact boundary hypersurface $\partial(D)$. The variational principle

$$\delta A = 0, \quad \delta\phi^{A\cdot\cdot}{}_{|\partial(D)} = 0, \tag{3.5}$$

yields the Euler–Lagrange field equations

$$\partial_A\left(\frac{\partial L}{\partial\partial_A\phi^{B\cdot\cdot}}\right) - \frac{\partial L}{\partial\phi^{B\cdot\cdot}} = 0. \tag{3.6}$$

The invariance of the action integral (3.5) under (3.1) implies that

$$\int_D d^8\xi\,\partial_A\left[\left(\frac{\partial L}{\partial\partial_A\phi^{B\cdot\cdot}}\right)\tilde{\delta}\phi^{B\cdot\cdot} + \left(\frac{\partial\lambda}{\partial\partial_A\bar{\phi}^{B\cdot\cdot}}\right)\tilde{\delta}\bar{\phi}^{B\cdot\cdot} + L\delta\xi^A\right] = 0. \tag{3.7}$$

By considering different possible cases for $\tilde{\delta}\phi^{B\cdot\cdot}$ one obtains various Noether's theorems. For example, in the case $\epsilon^A\neq 0$, $\epsilon_{AB} = 0$, using (3.1) and (3.3), Eq. (3.7) yields the following differential conservation law:

$$\partial_A T^A_B = 0, \qquad T^A_B \equiv \left(\frac{\partial L}{\partial\partial_A\phi^{\cdot\cdot}}\right)\partial_B\phi^{\cdot\cdot} + (\text{c.c.}) - \delta^A_B L. \tag{3.8}$$

The notation (c.c.) stands for the complex conjugate of the preceding terms and δ^A_B is the Kronecker delta.

In the case $\epsilon^A = 0$, $\epsilon_{AB}\neq 0$ Eq. (3.7) provides another conservation law, namely,

$$\begin{aligned}&\partial_A\, j^A_{BC} = 0, \quad j^A_{BC} \equiv M^A_{BC} + \mathscr{S}^A_{BC} = -j^A_{CB},\\ &M^A_{BC} \equiv \xi_B T^A_C - \xi_C T^A_B,\\ &\mathscr{S}^A_{BC} \equiv \left(\frac{\partial L}{\partial\partial_A\phi^{\cdot\cdot}}\right)S^{\cdot\cdot}{}_{BC\cdot\cdot}\phi^{\cdot\cdot} + (\text{c.c.}).\end{aligned} \tag{3.9}$$

The symmetrized conserved tensor corresponding to (3.8) is

$$\begin{aligned}\theta_{AB} = T_{AB} + (\tfrac{1}{2})\partial^C\Big[&\left(\frac{\partial L}{\partial\partial^C\phi^{\cdot\cdot}}\right)S^{\cdot\cdot}_{AB\cdot\cdot}\phi^{\cdot\cdot}\\ &+ \left(\frac{\partial L}{\partial\partial^A\phi^{\cdot\cdot}}\right)S^{\cdot\cdot}_{BC\cdot\cdot}\phi^{\cdot\cdot} + \left(\frac{\partial L}{\partial\partial^B\phi^{\cdot\cdot}}\right)S^{\cdot\cdot}_{AC\cdot\cdot}\phi^{\cdot\cdot}\\ &+ (\text{c.c.})\Big].\end{aligned} \tag{3.8$'$}$$

The infinitesimal version of the reciprocity transformation (2.4) is obtained by putting $\epsilon^A = 0$,

$$[\epsilon_{AB}] = \epsilon\begin{bmatrix} 0 & -\eta \\ \eta & 0\end{bmatrix}, \quad \epsilon\neq 0. \tag{3.10}$$

In this invariance the following conservation law which is a special case of (3.9) appears:

$$\begin{aligned}&\partial_A j^{(A)} = 0, \quad j^{(A)} \equiv T^{(A)} + (\tfrac{1}{2})A^{(A)},\\ &T^{(A)} \equiv q^j T^A_{p^j} - p^j T^A_{q^j},\\ &B^{(A)} \equiv \left(\frac{\partial L}{\partial\partial_A\phi^{\cdot\cdot}}\right)\eta^{ij}(S^{\cdot\cdot}_{q^ip^j\cdot\cdot} - S^{\cdot\cdot}_{p^iq^j\cdot\cdot})\phi^{\cdot\cdot} + (\text{c.c.}).\end{aligned} \tag{3.11}$$

The superscript in $j^{(A)}$ is bracketed for it is not necessarily a vector field under the extended Lorentz group L_8, although it contains two vector fields under the ordinary Lorentz group.

If the action integral (3.4) is invariant under the infinitesimal phase transformation (nongeometrical!)

$$\phi'^{\cdot\cdot} = \phi^{\cdot\cdot}e^{i\epsilon}, \qquad \bar{\phi}'^{\cdot\cdot} = \bar{\phi}^{\cdot\cdot}e^{-i\epsilon}, \tag{3.12}$$

then another conservation follows, namely,

$$\partial_A n^A = 0, \qquad n^A \equiv +i\left[\left(\frac{\partial L}{\partial\partial_A\bar{\phi}^{\cdot\cdot}}\right)\bar{\phi}^{\cdot\cdot} - \phi^{\cdot\cdot}\left(\frac{\partial L}{\partial\partial_A\phi^{\cdot\cdot}}\right)\right]. \tag{3.13}$$

To obtain the integral conservation laws, one would consider ξ^4, the time as the preferred coordinate and pick a cylindrical domain $D\subset V_8$ such that D lies between two $\xi^4 =$ constant hypersurfaces and surrounded by a closed wall σ generated by ξ^4-lines (see Fig. 2). It is also assumed that $T^A_B = j^A_{BC} = n^A = 0$ outside the tube having D as a part. Across σ the usual jump conditions

$$T^A_B\nu_A = j^A_{BC}\nu_A = n^A\nu_A = 0,$$

are imposed, where ν_A is the unit outer normal to σ. Then applying the 8-dimensional Gauss theorem to the differential conservation laws one finds the following integrals

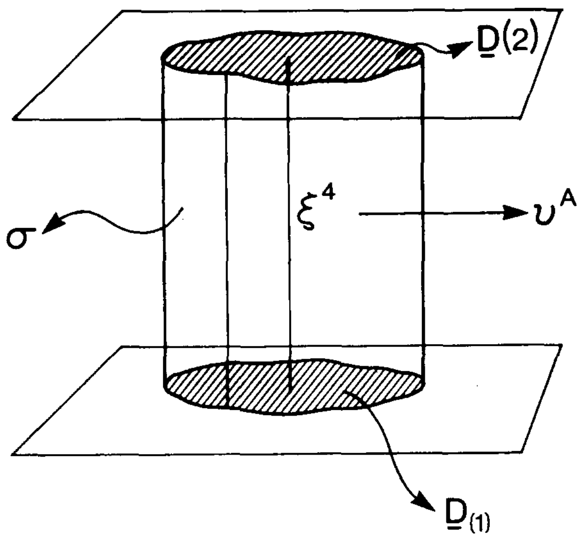

FIG. 2. The 7-dimensional bounding hypersurface is $\partial(D) = \sigma \cup \underline{D}_{(1)} \cup \underline{D}_{(2)}$.

$$
\begin{aligned}
K_A &\equiv \int_D d^7\xi \, T^4_A ,\\
\mathscr{M}_{BC} &\equiv \int_D d^7\xi \, M^4_{BC}, \quad \mathscr{S}_{BC} \equiv \int_D d^7\xi \, \mathscr{S}^4_{BC} ,\\
J_{BC} &\equiv \mathscr{M}_{BC} + \mathscr{S}_{BC} = -J_{BC} ,\\
T &\equiv \int_D d^7\xi \, T^{(4)}; \quad B \equiv \int_D d^7\xi \, B^{(4)} ,\\
N &\equiv \int_D d^7\xi \, n^4 ,
\end{aligned}
\tag{3.14}
$$

to be independent of ξ^4, and thus constants of motion. It is usual to extend $\underline{D}$ to the whole of the $\xi^4 =$ constant hypersurface V_7 and still assume the convergences of the above integrals. Keeping this in mind and Eq. (3.8), (3.9), (3.11), (3.14) one can also write the constants of motion as

$$
\begin{aligned}
K_\alpha &\equiv K_{q^\alpha} = \int_{V_7} d^3q \, d^4p \left[\left(\frac{\partial L}{\partial \partial_{q^4} \phi^{\cdot\cdot}} \right) \partial_{q^\alpha} \phi^{\cdot\cdot} + \text{(c.c.)} \right],\\
H &\equiv K_4 = \int_{V_7} d^3q \, d^4p \left[\left(\frac{\partial L}{\partial \partial_{q^4} \phi^{\cdot\cdot}} \right) \partial_{q^4} \phi^{\cdot\cdot} + \text{(c.c.)} - L \right],\\
X_i &\equiv K_{p^i} = \int_{V^7} d^3q \, d^4p \left[\left(\frac{\partial L}{\partial \partial_{q^4} \phi^{\cdot\cdot}} \right) \partial_{p^i} \phi^{\cdot\cdot} + \text{(c.c.)} \right],\\
\sigma_{ij} &\equiv \mathscr{S}_{q^i q^j} = \int_{V^7} d^3q \, d^4p \left[\left(\frac{\partial L}{\partial \partial_{q^4} \phi^{\cdot\cdot}} \right) S^{\cdot\cdot}_{q^i q^j \cdot\cdot} \phi^{\cdot\cdot} + \text{(c.c.)} \right],\\
J_{ij} &\equiv J_{q^i q^j} ,\\
L_{ij} &\equiv J_{p^i p^j} ,
\end{aligned}
$$

$$
\begin{aligned}
T &\equiv \int_{V_7} d^3q \, d^4p \left[\left(\frac{\partial L}{\partial \partial_{q^4} \phi^{\cdot\cdot}} \right) (q^i \partial_{p_i} - p^i \partial_{q^i}) \phi^{\cdot\cdot} + \text{(c.c)} \right.\\
&\quad \left. + p^4 L \right],\\
B &\equiv 2 \int_{V^7} d^3q \, d^4p \left[\left(\frac{\partial L}{\partial \partial_{q^4} \phi^{\cdot\cdot}} \right) \eta^{ij} (S^{\cdot\cdot}_{q^i p^j \cdot\cdot} - S^{\cdot\cdot}_{p^i q^j \cdot\cdot}) \phi^{\cdot\cdot} + \text{(c.c.)} \right],
\end{aligned}
$$

$$
\begin{aligned}
Q &\equiv e_0(T + \tfrac{1}{2}B) ,\\
N &\equiv i \int d^3q d^4p \left[\left(\frac{\partial L}{\partial \partial_{q^4} \bar{\phi}^{\cdot\cdot}} \right) \bar{\phi}^{\cdot\cdot} - \phi^{\cdot\cdot} \left(\frac{\partial L}{\partial \partial_{q^4}} \phi^{\cdot\cdot} \right) \right],
\end{aligned}
\tag{3.15}
$$

where e_0 is a charge parameter and $\partial_{q^i} \equiv \partial / \partial_{q^i}$, $\partial_{p^i} \equiv \partial / \partial p^i$, and Greek indices take the values 1, 2, 3. Some physical interpretation is in order now. For the $\phi^{\cdot\cdot}$-field distribution,

(i) K_α is the total momentum; (ii) H is the total energy; (iii) X_α is the center of mass; (iv) X_4 is the center of time; (v) J_{ij} is the total angular momentum, part of which is the total spin σ_{ij}; (vi) L_{ij} is the total angular momentum in momentum space (which is a generalization of the Newtonian concept $\mathbf{p} \times \mathbf{F}$ for a particle); (vii) $T + \frac{1}{2}B$ is the total oscillation of tremor of the field which is proportional to the total charge; (viii) B is the total intrinsic (spinlike) tremor to be identified with the baryon property; (ix) T is the total external (orbital-like) tremor to be identified with the isotopic plus strangeness property.

IV. THE COVARIANT MATRIX-WAVE EQUATIONS

The matrix-wave equations which are covariant under the extended Lorentz group L_8 are taken to be the linear, first-order set of partial differential equations

$$
[\alpha^A \partial_A - imI] \psi(\xi) = 0 , \tag{4.1}
$$

where the eight matrices α^A are required to be an irreducible representation of the generators of an abstract algebra and I is the unit matrix.

The general commutation rules or multiplication rules of the algebra are derived from the invariance of the wave equation (4.1) under the extended group L_8:

$$
\xi'^A = l^A_B \xi^B, \quad \psi' = T(L)\psi, \quad T(L)\alpha^A T(L)^{-1} = l_B^{-A} \alpha^B, \tag{4.2}
$$

where $T(L)$ is the representation of the extended group in the vector space of representation of the algebra. An element of this algebra $K \equiv k_A \alpha^A$, $((k_A) \in V_8)$ undergoes the following transformation

$$
K' = k'_A \alpha^A = T(L) K T(L)^{-1}. \tag{4.3}
$$

As a matrix of finite degree, K satisfies a minimal polynomial equation with coefficients being polynomial functions of the numbers k_A. But (4.3) shows that the minimal equation is invariant under the action of the extended group L_8 so that the coefficients must be functions of the invariant combination $k^2 \equiv k_A k^A$. Therefore the minimal polynomial of the matrix K of the highest degree $2n$ or $2n+1$ can be factorized into the forms.[6]

$$
(K^2 - \lambda_1 k^2 I)(K^2 - \lambda_2 k^2 I) \cdots (K^2 - \lambda_{2n} k^2 I) = 0 , \tag{4.4a}
$$

$$
K(K^2 - \lambda_2 k^2 I)(K^2 - \lambda_3 k^2 I) \cdots (K^2 - \lambda_{2n+1} k^2 I) = 0 , \tag{4.4b}
$$

where λ_i are complex constants in general.

Equations (4.4a) and (4.4b) can be written explicitly in terms of α-matrices as

$$k_{A_1}k_{B_1}\cdots k_{A_{2n}}k_{B_{2n}}(\alpha^{A_1}\alpha^{B_1}-\lambda_1\mathcal{N}^{A_1B_1}I)\cdots$$
$$(\alpha^{A_{2n}B_{2n}}-\lambda_{2n}\mathcal{N}^{A_{2n}B_{2n}}I)=0\,, \tag{4.5a}$$
$$k_{A_1}k_{A_2}k_{B_2}\cdots k_{A_{2n+1}}k_{B_{2n+1}}\alpha^{A_1}$$
$$\times(\alpha^{A_2}\alpha^{B_2}-\lambda_2\mathcal{N}^{A_2B_2}I)\cdots(\alpha^{A_{2n+1}}\alpha^{B_{2n+1}}$$
$$-\lambda_{2n+1}\mathcal{N}^{A_{2n+1}B_{2n+1}}I)=0\,. \tag{4.5b}$$

Since above equations should hold for every $(k_c)\in V_8$, the α-matrices should satisfy the generalized commutation rules

$$\sum_p(\alpha^{A_1}\alpha^{B_1}-\lambda_1\mathcal{N}^{A_1B_1}I)\cdots(\alpha^{A_{2n}B_{2n}}-\lambda_{2n}\mathcal{N}^{A_{2n}B_{2n}}I)=0\,, \tag{4.6a}$$

$$\sum_p\alpha^{A_1}(\alpha^{A_2}\alpha^{B_2}-\lambda_1\mathcal{N}^{A_2B_2}I)\cdots(\alpha^{A_{2n+1}B_{2n+1}}$$
$$-\lambda_{2n+1}\mathcal{N}^{A_{2n+1}B_{2n+1}}I)=0\,, \tag{4.6b}$$

where $_p$ stands for the summation over all possible permutations of $4n$ or $4n+2$ indices respectively.

For the infinitesimal version $T(L)=I+(1/2)\epsilon^{AB}S_{AB}$ of the transformations (4.2) of the α^A-matrices one has the following:

$$[\alpha_C,S_{AB}]=\mathcal{N}_{CA}\alpha_B-\mathcal{N}_{CB}\alpha_A\,, \tag{4.7}$$

where $[A,B]\equiv AB-BA$.

Together with the algebraic multiplication of the corresponding Lie algebra D_4 (see Appendix), namely,

$$[S_{KL},S_{MN}]$$
$$=\mathcal{N}_{KN}S_{LM}-\mathcal{N}_{KM}S_{LM}+\mathcal{N}_{LM}S_{KN}-\mathcal{N}_{LN}S_{KM}\,, \tag{4.8}$$

the problem of finding all covariant equations (4.1) boils down to finding the general solutions of (4.6a) or (4.6b), (4.7), (4.8) for α^A and S_{AB}.

For the variational derivation of the wave equation and the construction of the tensors T_{AB}, j^A_{BC} in Eqs. (3.8), (3.9), one needs a nonsingular matrix Λ such that

$$(\alpha^A)^\dagger=\Lambda\alpha^A\Lambda^{-1}\,, \tag{4.9}$$

where the dagger denotes the Hermitian conjugation. If such a matrix exists, it can be chosen to be Hermitian because $(\Lambda^\dagger)^{-1}\Lambda$ commutes with all the irreducible α^A's and thus must be a scalar matrix. In case the constants λ_i's in Eqs. (4.6a), (4.6b) are real one can infer by taking the Hermitian conjugates of these equations that $(\alpha^A)^\dagger$-matrices satisfy the same multiplication rules as α^A-matrices. Therefore in such a case Λ must exist.

As a simple example, consider the minimal polynomial of degree 2. In this case λ_1 can be set to 1 by suitable normalization. According to (4.6a) the multiplication rules are

$$\alpha^A\alpha^B+\alpha^B\alpha^A=2\mathcal{N}^{AB}I\,. \tag{4.10}$$

An irreducible representation of the α^A-matrices satisfying (4.10) are 16×16 matrices (this is the only possible size) exhibited below[6]:

$$\alpha^1=i\sigma^1\times I\times I\times I,\quad \alpha^2=i\sigma^3\times\sigma^1\times I\times I,$$
$$\alpha^3=i\sigma^3\times\sigma^3\times\sigma^1\times I,\quad \alpha^4=\sigma^3\times\sigma^3\times\sigma^3\times\sigma^1,$$
$$\alpha^5=i\sigma^2\times I\times I\times I,\quad \alpha^6=i\sigma^3\times\sigma^2\times I\times I,$$
$$\alpha^7=i\sigma^3\times\sigma^3\times\sigma^2\times I,\quad \alpha^8=\sigma^3\times\sigma^3\times\sigma^3\times\sigma^2,$$
$$\Lambda=\sigma^3\times\sigma^3\times\sigma^3\times I,$$
$$\Lambda^\dagger=\Lambda=\Lambda^{-1},$$
$$S_{AB}=(\tfrac{1}{4})[\alpha_A,\alpha_B]=(\tfrac{1}{2})[\alpha_A\alpha_B-\mathcal{N}_{AB}I]\,, \tag{4.11}$$

where $M\times N$ denotes the Kronecker product of the matrices M, N; σ^α are the Pauli matrices, and I is the 2×2 identity matrix.

These α^A-matrices generate an associative Clifford algebra of order 256. One basis is $e^{n_1n_2\cdots n_8}\equiv(\alpha^1)^{n_1}(\alpha^2)^{n_2}\cdots(\alpha^8)^{n_8}$, where the n_A are integers mod (2). This is a generalization of the Dirac algebra.

V. GENERALIZATION OF DUFFIN–KEMMER ALGEBRA

For the next example the minimal polynomial of degree 3 is considered. Using the symbols β^A instead of α^A in this case (4.6) yields

$$\beta_A\beta_B\beta_C+\beta_C\beta_B\beta_A=\beta_A\mathcal{N}_{BC}+\beta_C\mathcal{N}_{AB}\,. \tag{5.1}$$

These β_A's generate a generalization of the Duffin—Kemmer algebra.

In this case $S_{AB}=[\beta_A,\beta_B]$. Furthermore, one can show that the matrix-wave equation

$$[\beta^A\partial_A-imI]\psi(\xi)=0\,, \tag{5.2}$$

contains the scalar and vector field equations.[7] This can be proved using the following consequences of (5.1):

$$\begin{aligned}&\beta_A^3=\epsilon(A)\beta_A\,,\\&\beta_A\beta_B^2=[\epsilon(B)I-\beta_B^2]\beta_A\quad(A\neq B)\,,\\&\beta_A\beta_B\beta_A=\mathcal{N}_{AB}\beta_A\quad(A\text{ not summed})\,,\\&\beta_A^2\beta_B^2=\beta_B^2\beta_A^2\,,\end{aligned} \tag{5.3}$$

where $\epsilon(A)\equiv\mathcal{N}_{AA}$, A not summed. Some additional matrices would be required and they are

$$\Gamma\equiv\prod_{A=1}^{8}\beta_A^2\,,\quad \pi_A\equiv\Gamma\beta_A\,,$$
$$\pi'_A\equiv\begin{cases}-\Big(\prod_{C=1}^{7}\beta_C^2\Big)\beta_A\beta_8, & \text{for } A=1,2,\cdots,7\,,\\ \Big(\prod_{C=1}^{7}\beta_C^2\Big)(I-\beta_8^2), & \text{for } A=8\,,\end{cases}$$
$$\pi'_{AB}\equiv\pi'_A\beta_B\,,\quad \pi_A\beta_B=\mathcal{N}_{AB}\Gamma\,,$$
$$\pi'_{AB}=-\pi'_{BA}\,,\quad \pi'_{AB}\beta_C=\mathcal{N}_{BC}\pi'_{AB}-\mathcal{N}_{AC}\pi'_B\,. \tag{5.4}$$

Defining four wavefunctions

$$\phi\equiv i\Gamma\psi,\quad \phi_A\equiv\pi_A\psi,\quad V_A\equiv(im)^{-1}\pi'_A\psi,\quad U_{AB}\equiv\pi'_{AB}\psi\,, \tag{5.5}$$

and multiplying (5.2) from the left by π_A, Γ, π'_{AB}, π'^A, one can obtain the following four wave equations, respectively,

$$\partial_A\phi=-m\phi_A\,,\quad \partial_A\phi^A=m\phi\,, \tag{5.6a}$$
$$\partial_BV_A-\partial_AV_B=U_{AB},\quad \partial_BU^{AB}=-m^2V^A\,. \tag{5.6b}$$

Equations (5.6a) and (5.6b) are scalar and vector field equations. By (5.2), (5.5), (5.6a), and (5.6b) one can construct the irreducible representations of β^A-algebra in dimensions 9 and 36, respectively. These together with the 1-dimensional trivial representation $\beta_A=0$ are believed to be the only possible irreducible representations. The wave equation (4.1) with the choice of (4.10), (5.6a), and (5.6b) would be the most relevant one for the applications to elementary particles.

VI. THE MATRIX-WAVE FIELD THEORY

We return to the general case of the wave equation (4.1). This equation is derivable from the variational principle with the Lagrangian

$$L = (2i)^{-1}[(\partial_A \tilde{\psi})\alpha^A \psi] + (\text{h.c.}) + m\tilde{\psi}\psi ,$$
$$\tilde{\psi} \equiv \psi^\dagger \Lambda , \tag{6.1}$$

and (h.c.) stands for the Hermitian conjugate of the previous terms.

The various matrix-Green's functions of the field equation (4.1) are furnished by

$$G_{(a)}(\xi - \xi') \equiv -(2\pi)^{-8} i \int_{-\infty}^{\infty} d^7\xi \int_{C_{(a)}} d\xi^4 \,[\alpha^A \xi_A - mI]^{-1} \times e^{i\xi_A(\xi^A - \xi'^A)} , \tag{6.2}$$

where $C_{(a)}$ are the various contours in the complex ξ^4-plane. The integrand in the last contour integral will have simple poles at the finite points of the ξ^4-plane. The homogeneous Green's functions would involve contours $C_{(a)}$, which are closed Jordan curves enclosing some of these poles. On the other hand, for the inhomogeneous Green's functions the contours $C_{(a)}$ should be along the real line $(-\infty, \infty)$ and in the neighborhood of each real pole either the principal value is considered, or the semicircular detours (in upper or lower half-planes) are taken.

The various tensor fields obeying the differential conservation laws are computed from (3.8), (3.9), (3.13), (6.1), (noticing $L = 0$ by the field equations) to be

$$T_{AB} = (2i)^{-1}[(\partial_B \tilde{\psi})\alpha_A \psi] + (\text{h.c.}) ,$$
$$M^A_{\ BC} = \xi_B T^A_{\ C} - \xi_C T^A_{\ B} ,$$
$$\mathscr{S}^A_{\ BC} = -(2i)^{-1}\tilde{\psi}\alpha^A S_{BC}\psi + (\text{h.c.}) , \tag{6.3}$$
$$j^A_{BC} = M^A_{\ BC} + \mathscr{S}^A_{\ BC} ,$$
$$n^A = \tilde{\psi}\alpha^A \psi .$$

The integral constants (3.15) of the ψ-field are,

$$K_a = (2i)^{-1} \int_{V_7} d^3q\, d^4p\, [(\partial_{qa}\tilde{\psi})\alpha^4\psi] + (\text{h.c.}) ,$$

$$X_a = (2i)^{-1} \int_{V_7} d^3q\, d^4p\, [(\partial_{pa}\tilde{\psi})\alpha^4\psi] + (\text{h.c.}) ,$$

$$J_{ab} = (2i)^{-1} \int_{V_7} d^3q\, d^4p\, [q_a(\partial_{qb}\tilde{\psi})\alpha^4\psi - q_b(\partial_{qa}\tilde{\psi})\alpha^4\psi] ,$$
$$-(2i)^{-1} \int_{V_7} d^3q\, d^4p\, [\tilde{\psi}\alpha^4 S_{q^a q^b}\psi] + (\text{h.c.}) ,$$

$$L_{ab} = (2i)^{-1} \int_{V_7} d^3q\, d^4p\, [p_a(\partial_{pb}\tilde{\psi})\alpha^4\psi - p_b(\partial_{pa}\tilde{\psi})\alpha^4\psi] ,$$
$$-(2i)^{-1} \int_{V_7} d^3q\, d^4p\, [\tilde{\psi}\alpha^4 S_{p^a p^b}\psi] + (\text{h.c.}) ,$$

$$T = \frac{i}{2} \int_{V_7} d^3q\, d^4p\, [\tilde{\psi}\alpha^4(q^a\partial_{pa} - p^a\partial_{qa})\psi] + (\text{h.c.}) ,$$
$$B = \frac{i}{2} \int_{V_7} d^3q\, d^4p\, [\tilde{\psi}\alpha^4(S^{q^a}_{\ p^a} - S^{p^a}_{\ q^a})\psi] + (\text{h.c.}) . \tag{6.4}$$

ACKNOWLEDGMENT

The author is indebted to NRC Grant No. 3993.

APPENDIX

The homogeneous extended Lorentz group L_8 has 28 infinitesimal generators.

$$S_{AB} = -S_{BA} \equiv \xi_B \partial_A - \xi_A \partial_B . \tag{A1}$$

These generate a split simple Lie algebra

$$[S_{AB}, S_{CD}] = C^{EF}_{AB,CD} S_{EF} , \tag{A2}$$

where the structure constants are

$$C^{EF}_{AB,CD} \equiv \mathscr{N}_{AD}\delta^E_C\delta^F_B + \mathscr{N}_{AC}\delta^E_B\delta^F_D + \mathscr{N}_{BD}\delta^E_A\delta^F_C + \mathscr{N}_{BC}\delta^E_D\delta^F_A . \tag{A3}$$

For the canonical form of (A.2) one has to make the following complex linear transformation

$$\xi'^a = 2^{-1/2}[\xi^a + i\xi^{a+4}] ,$$
$$\xi'^{a+4} = 2^{-1/2}[\xi^a - i\xi^{a+4}] ,$$
$$S'_{ab} = \xi'_a\partial'_b - \xi'_b\partial'_a . \tag{A4}$$

The canonical forms of the generators are

$$H_a \equiv S'_{a,a+4} , \quad E_{(e_a+e_b)} \equiv S'_{a,b} , \quad E_{(e_a-e_b)} \equiv S'_{a,b+4} , \tag{A5}$$
$$E_{(-e_a-e_b)} \equiv S'_{a+4,b} , \quad E_{(-e_a--e_b)} \equiv S'_{a+4,b+4} .$$

The root diagram in an inner-product vector space V_4 consists of the 24 vectors $\pm e_a \pm e_b$, $a \neq b$. The corresponding Lie algebra is D_4.

The Cartan matrix is

$$\begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 2 & -1 & -1 \\ 0 & -1 & 2 & 0 \\ 0 & -1 & 0 & 2 \end{pmatrix} . \tag{A6}$$

The corresponding Dynkine diagram (Fig. 3) is shown where $\alpha_1 = e_1 - e_2$, $\alpha_2 = e_2 - e_3$, $\alpha_3 = e_3 - e_4$, $\alpha_4 = e_3 + e_4$ are a set of fundamental roots.

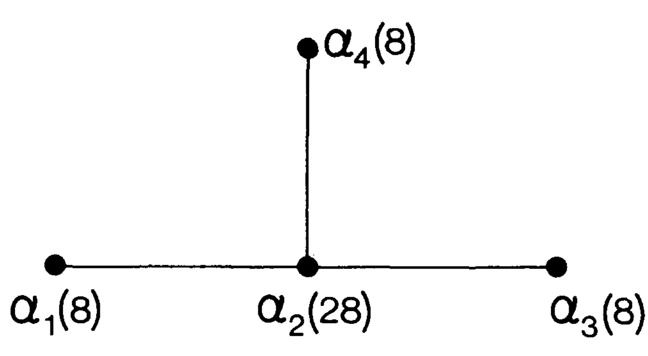

FIG. 3. Dynkine diagram.

Any representation of the algebra has a weight vector (m_1,m_2,m_3,m_4) where m_i's are all integers or all half-integers. The Weyl reflection group W is the group of permutations of m_i's with an even number of changes of sign. Any irreducible representation can be characterized by the highest weight (m_1,m_2,m_3,m_4). Four fundamental representations of D_4 can be chosen to be α_1, α_2, α_3, α_4 in Fig. 3 where the numbers inside the parentheses indicate dimensionality of these representations. The *triality* of 8-dimensional representations α_1, α_3, α_4 are to be noted.

Weyl's dimension formula for the irreducible representation (m_1,m_2,m_3,m_4) of D_4 yields the dimensionality

$$d = \frac{(m_1+3)^2-(m_2+2)^2}{5}\cdot\frac{(m_1+3)^2-(m_3+1)^2}{8}\cdot\frac{(m_1+3)^2-m_4^2}{9}\cdot\frac{(m_2+2)^2-(m_3+1)^2}{3}\cdot\frac{(m_2+2)^2-m_4^2}{4}\cdot\frac{(m_3+1)^2-m_4^2}{1}. \tag{A7}$$

[1]M. Born, Proc. R. Soc. London Ser. A **165**, 291 (1938); **116**, 552 (1938); Nature **163**, 207 (1949); Rev. Mod. Phys. **21**, 463 (1949). Also see H.S. Green, Nature **163**, 208 (1949).

[2]A. Landé, Phys. Rev. **56**, 482, 486 (1939).

[3]H. Yukawa, Phys. Rev. **76**, 300 (1949).

[4]A. Das, J. Math. Phys. **7**, 45 (1966).

[5]C. Lanczos, Rev. Mod. Phys. **29**, 337 (1957).

[6]E.M. Corson, *Introduction to Tensors, Spinors and Relativistic Wave Equations* (Blackie, London, 1955), pp. 143, 164.

[7]H. Umezawa, *Quantum Field Theory* (North-Holland, Amsterdam, 1956), p. 85.

Extended phase space. II. Unified meson fields

A. Das
Department of Mathematics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

(Received 30 October 1979; accepted for publication 17 January 1980)

The classical scalar equation in the extended phase space V_8 is studied. It is the generalization of the usual Klein–Gordon equation and is covariant under the extended Poincaré group P_8. In this equation there is obvious symmetry of the variables q and p and thus the principle of reciprocity is automatically incorporated. The scalar field is expressed as a Fourier integral $\phi(q,p) \sim \int dk\, dx\, \alpha(k,x) e^{i(kq+px)}$ to compute the integral constants like the total energy, total momentum, etc. Then the integral constants turn out to be meaningful quantities by interpreting $f(k,x) \sim |\alpha|^2$ as the statistical distribution function for the scalar field particles. Next the scalar field is expressed as the Fourier–Bessel integral

$$\phi(\rho,\theta) \sim \sum_{t=-\infty}^{\infty} \int_0^{\infty} dk\, k\alpha(k,t) J_t(k\rho) e^{it\theta} = \sum_{t=-\infty}^{\infty} \phi^{(t)}, \quad \rho = \sqrt{q^2+p^2},\ \theta = \arctan(p/q).$$

The integral constants are computed from a single (t)-mode $\phi^{(t)}$. These are accessible to physical interpretations. Especially, the total charge can be linked up with the Gell-Man–Nishijima's formula, provided one of the quantum numbers (t), say t_3, is identified with the isotopic quantum number and $2t_4$ is identified with strangeness. With each of the (t)-mode $\phi^{(t)}$ a meson field is associated so that the ϕ-field itself is the unified meson field.

I. INTRODUCTION

Extended phase space V_8 of space–time–momentum–energy is advocated instead of the space–time, for the formulation of the meson field theory.

The generalized Klein–Gordon equation for the scalar field in V_8 is introduced. The group of covariance for this field equation is the extended Poincaré group P_8. As an element of this group one has the transformation $q' = -p, p' = q$ which is the reciprocity transformation of Born.[1] Yukawa[2] pursued Born's reciprocity in his theory of the nonlocal scalar field. He had two equations for the scalar field, one for the space–time and another for the momentum–energy space. In the present paper the generalized Klein–Gordon equation is obtained by gluing his two equations together so to say. The first advantage of this procedure is that the group of covariance is enlarged and secondly the scalar field thus obtained can describe various mesons in a unified fashion. The generalized Klein–Gordon equation studied here is mathematically equivalent to one written for the complex space–time.[3]

In Sec. II, the Lagrangian methods for the scalar field are pursued. Applying the variational formalisms developed[4] in Paper I, the various integral constants such as the total energy, the total momentum etc., are expressed.

In Sec. III, the plane wave decomposition or the Fourier transform of the scalar field is obtained. Next the integral constants are computed by using the plane wave decomposition. These are all expressed as 7-dimensional integrals in the dual extended phase space. The integrands are proportional to $f(k,x) \equiv |\alpha(k,x)|^2$, where $\alpha(k,x)$ is the complex valued amplitude, k is the dual momentum x is the dual position–time. These integral constants become tantalizingly meaningful if $f(k,x)$ is interpreted as the statistical distribution function for the mesons in the dual phase space. This view point strongly supports the statistical interpretation of the quantum mechanics.[5] These integral constants fit in well with the total energy, the total momentum, the average position, and the average time of the scalar field.

In Sec. IV, instead of the coordinates q,p the variables $\rho = (p^2+q^2)^{1/2}, \theta = \arctan(p/q)$ are introduced. These are analogous to the action-angle variables. The first purpose is to facilitate by the variables θ, the description of the rotational motions in the p-q planes (which are oscillations or tremors in space–time) and to obtain the corresponding tremor quantum numbers $(t) = (t_1,t_2,t_3,t_4)$. Secondly, in the subsequent development of the quantized V_8, the 4 eigenvalues of the quantized ρ will bring lattice structure of V_8 while the θ variables will be completely uncertain. This change of variables brings in, naturally, the Hankel transform $\phi(\rho,\theta) \sim \Sigma_{t=-\infty}^{\infty} \int_0^{\infty} dk\, k\, \alpha(k,t) J_t(k,\rho) e^{it\theta} = \Sigma_{t=-\infty}^{\infty} \phi^{(t)}$ instead of the usual Fourier transform. The integral constants are evaluated from the Lagrangian formalism for a single (t)-mode $\phi^{(t)}(\rho,\theta)$. These are expressed as the 3-dimensional integrals in the dual momentum space and are accessible to physical interpretations. The most interesting one among these constants is the total charge Q which is proportional to $(t_1+t_2+t_3+t_4)$. For the special case $t_1 = t_2 = 0$ writing the strangeness $s = 2t_4$ one obtains the Gell-Man–Nishijima's[6] expression for the meson charge. Furthermore, if one accepts $-\eta^{ij}[\partial^2/\partial p^i \partial p^j + (1/p^i)\partial/\partial p^j]$ as the (mass)^2-operator instead of the commonly used[7] $-\eta^{ij}\partial^2/\partial q^i \partial q^j$, then the field $\phi^{(t)}$ will have (mass)$^2 = m^2 - \eta^{ij}(t_i t_j/\rho^i \rho^j)$. The last formula for the special case $t_1 = t_2 = 0$, $s = 2t_4$ goes over to (mass)$^2 = m^2 + t_3^2/\rho_3^2 - s^2/4\rho_4^2$ which is the analog of the Okubo[8] mass formula. From the preceding discussions it appears that each of the (t)-mode $\phi^{(t)}(\rho,\theta)$ can describe a particular meson so that the scalar field $\phi(\rho,\theta) = \Sigma_{t=-\infty}^{\infty} \phi^{(t)}(\rho,\theta)$ stands for the unified meson field.

In Appendices I and II the rigorous computations of the integral constants are provided.

 0022-2488/80/061513-08$1.00

II. NOTATIONS AND THE FREE SCALAR FIELD EQUATION

The extended phase space V_8 of space–time–momentum–energy is coordinatized by $\xi \equiv (\xi^1,\xi^2,\xi^3,\xi^4,\xi^5,\xi^6,\xi^7,\xi^8) \equiv (q^1,q^2,q^3,q^4,p^1,p^2,p^3,p^4) \equiv (q,p)$. The capital italic indices take 1,2,...,8; the small italic indices take 1,...,4; the Greek indices take 1, 2, 3. Summation convention is followed over the indices, which occur more than once. The units are so chosen that $a = b = c = 1$, where these are, respectively, fundamental length, momentum, velocity, and all physical quantities are pure numbers. The metric tensor of V_8 is given by $[\mathcal{N}_{AB}] \equiv \text{diag}[-1^3,1,-1^3,1]$ and the Minkowskian metric tensor is $[\eta_{ij}] \equiv [\eta_{q^i q^j}] \equiv \text{diag}[-1^3,1]$. The raising and lowering of the tensor indices in V_8 are accomplished by $\mathcal{N}_{AB}$ and $\mathcal{N}^{AB}$. The partial derivatives are denoted by $\partial_A \equiv \partial/\partial\xi^A$, $\partial_{q^i} \equiv \partial/\partial q^i, \equiv \partial/\partial_{p^i}$. The equation, say (3.8) of Paper I is indicated by (I-3.8).

The Lagrangian for the complex scalar field is assumed to be

$$L(\phi,\bar{\phi},\partial_A\phi,\partial_A\bar{\phi}) \equiv \mathcal{N}^{AB}(\partial_A\bar{\phi})(\partial_B\phi) - m^2\bar{\phi}\phi, \quad (2.1)$$

where m is the mass parameter, and the bar stands for the complex conjugation.

The Euler–Lagrange field equations in this case go over to

$$[\partial^A\partial_A + m^2]\,\phi(\xi) = 0, \quad (2.2)$$

which is the generalization of the Klein–Gordon equations.

The various Green's functions for the partial differential Eq. (2.2) are

$$\underset{(a)}{G}(\xi - \xi') \equiv (2\pi)^{-8}\int_{-\infty}^{\infty} d^7\zeta \times \int_{C_{(a)}} d\zeta^4\,(m^2 - \zeta^B\zeta_B)^{-1} e^{i\zeta_A(\xi^A - \xi'^A)}, \quad (2.3)$$

where $C_{(a)}$ are the different contours in the complex ζ^4-plane, which are exhibited in Fig. 1. In the complex ζ^4-plane,

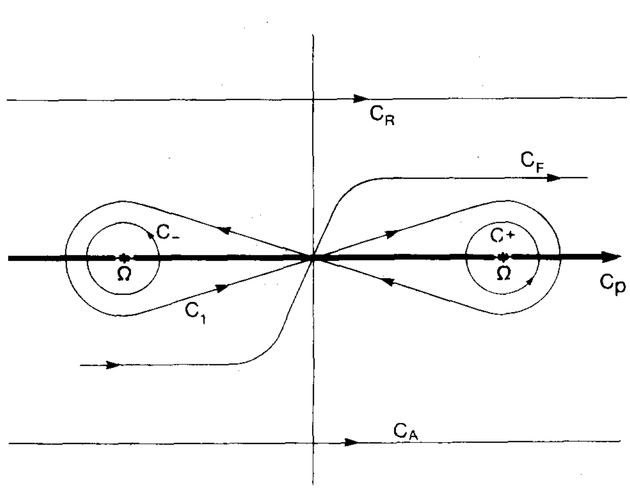

FIG. 1. The complex ζ^4-plane.

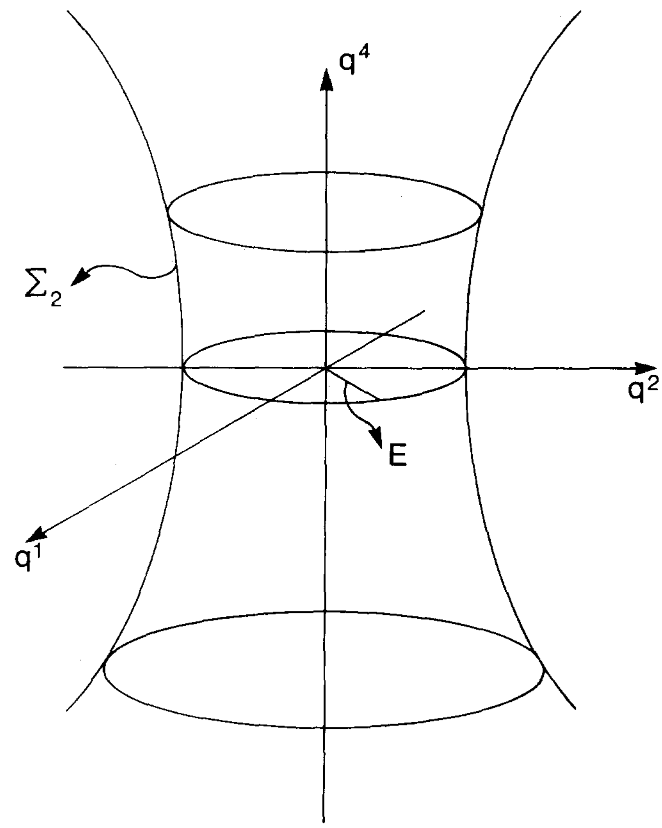

FIG. 2. A singular surface Σ_2 of the Green's functions.

the two poles of the integrand in (2.3) occurs at points

$$\zeta^4 = \pm\,[(\zeta^1)^2 + (\zeta^2)^2 + (\zeta^3)^2 + (\zeta^5)^2 + (\zeta^6)^2 + (\zeta^7)^2 - (\zeta^8)^2 + m^2]^{1/2} \equiv \pm\,\Omega.$$

These points are situated either on the real axis or else on the imaginary axis and are symmetric about the origin.

The Green's functions (2.3) are singular [one can easily see from (2.3) for the point $\xi^A = \xi'^A$ on N_7 that $\lim_{\xi\to\xi'} |G_a(\xi - \xi')| = \infty$] on the null hypersurfaces in V_8 given by

$$N_7: (\xi^A - \xi'^A)(\xi_A - \xi'_A) = 0. \quad (2.4)$$

To get a visual picture of these null hypersurfaces, consider $N_7^{(0)}: \xi^A\xi_A = 0$ and five hyperplanes given by $\xi^3 = \xi^5 = \xi^6 = \xi^7 = 0$, $\xi^8 = E > 0$, a constant. The intersection Σ_2 of all these hypersurfaces is given by the 3-dimensional equation $(\xi^1)^2 + (\xi^2)^2 = (\xi^4)^2 + E^2$ and plotted in Fig. 2. The various tensor quantities obeying the differential conservation laws for the complex scalar field can be computed from (2.1) and [cf. (I-3.8), (I-3.9), (I-3.13)]

$$T^A{}_B = (\partial^A\bar{\phi})(\partial_B\phi) + (\text{c.c.}) - \delta^A{}_B L,$$

$$\mathcal{M}^A{}_{BC} = \xi_B T^A{}_C - \xi_C T^A{}_B, \quad \mathcal{S}^A{}_{BC} = 0,$$

$$j^A{}_{BC} = \mathcal{M}^A{}_{BC} + \mathcal{S}^A{}_{BC}, \quad n_A = i\bar{\phi}\partial_A\phi + (\text{c.c.}), \quad (2.5)$$

where (c.c.) stands for the complex conjugation of the previous terms.

Now using (2.1) and (2.5) some of the integral constants [cf. (I-3.15)] become

$$K_\alpha \equiv \int_{V_7} d^3q\, d^4p\, T^4{}_\alpha = \int_{V_7} d^3q\, d^4p \times [(\partial_{q^4}\bar{\phi})(\partial_{q^\alpha}\phi) + (\text{c.c.})],$$

$$H \equiv K_4 \equiv \int_{V_7} d^3q\, d^4p\, T^4{}_4 = \int_{V_7} d^3q\, d^4p\, [(\partial_{q^4}\bar{\phi})(\partial_{q^4}\phi)] + (\partial_{q^\alpha}\bar{\phi})(\partial_{q^\alpha}\phi) + (\partial_{p^\alpha}\bar{\phi})(\partial_{p^\alpha}\phi) - (\partial_{p^4}\bar{\phi})(\partial_{p^4}\phi) + m^2\bar{\phi}\phi,$$

$$X_i \equiv \int_{V_7} d^3q\, d^4p\, T^4{}_{p^i} = \int_{V_7} d^3q\, d^4p \times [(\partial_{q^4}\bar{\phi})(\partial_{p^i}\phi) + (\text{c.c.})],$$

$$(\sigma)^2 \equiv (\sigma_{12})^2 + (\sigma_{23})^2 + (\sigma_{31})^2 = 0,$$

$$T = \int_{V_7} d^3q\, d^4p\, \{(\partial_{q^4}\bar{\phi})(q^i\partial_{p^i} - p^i\partial_{q^i})\phi + (\text{c.c.}) + p^4\,[\eta^{ij}(\partial_{q^i}\bar{\phi}\cdot\partial_{q^j}\cdot\phi + \partial_{p^i}\bar{\phi}\cdot\partial_{p^j}\phi) - m^2\bar{\phi}\phi\,]\},$$

$$B = 0, \quad Q = e_0 T,$$

$$N = \int_{V_7} d^3q\, d^4p\, n^4 = i\int_{V_7} d^3q\, d^4p\, [\bar{\phi}\partial_{q^4}\phi\,] + (\text{c.c.}), \qquad (2.6)$$

where e_0 is a charge parameter.

III. THE PLANE WAVE SOLUTIONS AND THE STATISTICAL INTERPRETATION

Now assuming the appropriate restrictions on ϕ satisfying (2.2), the Fourier integral of this function can be stated as

$$\phi(\xi) = (2\pi)^{-4}\int_{-\infty}^{\infty} d^8\zeta\, A(\zeta)\delta(\zeta^B\zeta_B - m^2)e^{-i\xi_C\zeta^C}, \qquad (3.1)$$

where $\delta(x)$ is the Dirac delta function for the variable x. Integrating out the variable ζ^4 (3.1) yields

$$\phi(q,p) = (2\pi)^{-7/2}\int_{-\infty}^{\infty} d^3k\, d^4x\, (2\Omega(k,x))^{-1/2} \times [\alpha(k,x)e^{-i(k_a q^a + x_a p^a)} + \bar{\beta}(k,x)e^{i(k_a q^a + x_a p^a)}],$$

$$\bar{\phi}(q,p) = (2\pi)^{-7/2}\int_{-\infty}^{\infty} d^3k\, d^4x\, (2\Omega(k,x))^{-1/2} \times [\bar{\alpha}(k,x)e^{i(k_a q^a + x_a p^a)} + \beta(k,x)e^{-i(k_a q^a + x_a p^a)}], \qquad (3.2)$$

where

$$(\zeta^A) \equiv (k_i, x_i), \quad (k) \equiv (k_1, k_2, k_3),$$

$$(x) \equiv (x_1, x_2, x_3, x_4),$$

$$\Omega(k,x) \equiv +\sqrt{k_\alpha k_\alpha + x_\alpha x_\alpha - x_4^2 + m^2}, \qquad (3.3)$$

$$k_a q^a \equiv k_\alpha q^\alpha + \Omega q^4, \quad \alpha(k,x) \equiv (4\pi\Omega)^{-1/2}A(k,\Omega,x),$$

$$\bar{\beta}(k,x) \equiv (4\pi\Omega)^{-1/2}A(-k, -\Omega, x).$$

It can be noticed from (3.3) that Ω can take imaginary values. To exclude that possibility for the sake of physical credibility the 7-dimensional integration in (3.2) should be restricted to the domain

$$D_7 \equiv \{(k,x)\,|\,k_\alpha k_\alpha + x_\alpha x_\alpha - x_4^2 + m^2 > 0\}.$$

This procedure might imply some additional restrictions on $\phi(q,p)$.

Computing some of the integral constants in (2.6) (calculated at $q^4 = 0$) with (3.2) one obtains (cf. Appendix I)

$$K_\alpha = \int_{D_7} d^3k\, d^4x\, [f_+(k,x) + f_-(k,x)]k_\alpha,$$

$$H \equiv K_4 = \int_{D_7} d^3k\, d^4x\, [f_+(k,x) + f_-(k,x)]\,\Omega(k,x),$$

$$X_a = \int_{D_7} d^3k\, d^4x\, [f_+(k,x) + f_-(k,x)]x_a, \qquad (3.4)$$

$$N = \int_{D_7} d^3k\, d^4x\, [f_+(k,x) - f_-(k,x)],$$

$$f_+(k,x) \equiv |\beta|^2 \geqslant 0, \quad f_-(k,x) \equiv |\alpha|^2 \geqslant 0.$$

Some physical interpretation can be reached now. From the above equations, $f_+(k,x), f_-(k,x)$ can be quite naturally intepreted as Boltzmann type distribution functions for the similarly created scalar particles and antiparticles in the (Fourier) dual phase space. This viewpoint definitely supports the statistical interpretation of the quantum mechanics.[5] The integral constants:

(i) K_α represents the total momentum; (ii) H is the total energy; (iii) X_a if the first moment for position and time; (iv) N is the total number.

All these integral constants are for the scalar field representing the scalar particle ensemble.

IV. APPLICATIONS TO THE UNIFIED THEORY OF MESONS

To relate the complex scalar fields to the meson theory, it is more suitable to treat the Hankel transform rather than the Fourier transform of the field $\phi(\xi)$. For that purpose in each of the four phase planes (see Fig. 3) the polar coordinates are introduced as

$$\rho^a \equiv \sqrt{(q^a)^2 + (p^a)^2},$$
$$\theta^a \equiv \arctan(p^a/q^a) \quad \text{(no summation)}, \qquad (4.1)$$

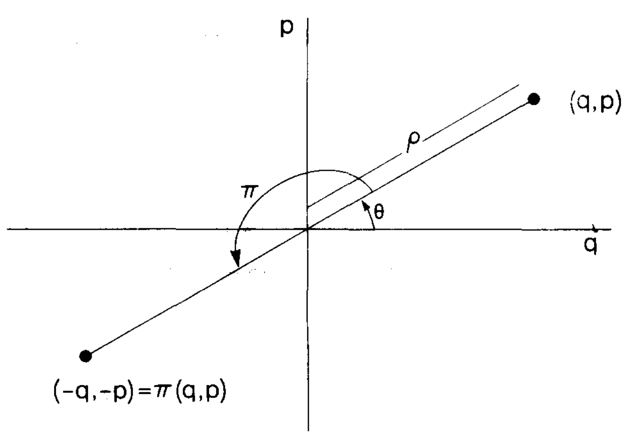

FIG. 3. The $(q-p)$-phase plane.

where $0<\rho^a$, $0\leqslant\theta^a<2\pi$. In these coordinates the generalized Klein–Gordon equation (2.2) goes over to

$$\{\eta^{ij}[\partial_{\rho^i}\partial_{\rho^j}+(1/\rho^i)\partial_{\rho^j} + (1/\rho^i\rho_j)\partial_{\theta^i}\partial_{\theta^j}]+m^2\}\,\phi(\rho,\theta)=0, \quad (4.2)$$

where $(\rho)\equiv(\rho^1,\rho^2,\rho^3,\rho^4)$, $(\theta)\equiv(\theta^1,\theta^2,\theta^3,\theta^4)$ and summation has been carried out on the indices which occur more than once.

Separating the variables and demanding that the angular functions are single-valued, one gets the basic solution functions as

$$\phi^{(t)}(\rho,\theta)=\phi^{(t)}(\rho)e^{it_a\theta^a}, \quad (4.3)$$

where $(t)\equiv(t_1,t_2,t_3,t_4)$ and each of $t_i=0,\pm 1,\pm 2,\cdots$. The functions $\phi^{(t)}(\rho)$ satisfy the partial differential equation

$$\{\eta^{ij}[\partial_{\rho^i}\partial_{\rho^j}+(1/\rho^i)\partial_{\rho^j}-(t_it_j/\rho^i\rho^j)]+m^2\}\,\phi^{(t)}(\rho)=0. \quad (4.4)$$

The solutions of the above equation can be expressed as Bessel functions and some of the Green's function[9] can be cited as

$$\begin{aligned}\underset{(a)}{G^{(t)}}(\rho,\rho') &= \int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3\int_{C(a)} d\kappa_4\,\kappa_4(m^2-\eta^{ij}\kappa_i\kappa_j)^{-1}J_{t_1}(\kappa_1\rho^1)\\ &\times J_{t_1}(\kappa_1\rho'^1)J_{t_2}(\kappa_2\rho^2)J_{t_2}(\kappa_2\rho'^2)J_{t_3}(\kappa_3\rho^3)J_{t_3}(\kappa_3\rho'^3)\\ &\times J_{t_4}(\kappa_4\rho^4)J_{t_4}(\kappa_4\rho'^4),\end{aligned} \quad (4.5)$$

where for $C_{(a)}$ see Fig. 4. In the complex κ_4-plane the two poles of the integrand in (4.5) occur at points

$$\kappa_4=\pm\omega(\underline{\kappa})\equiv\pm\sqrt{\kappa_\alpha\kappa_\alpha+m^2}\,. \quad (4.6)$$

The most general solutions of Eq. (4.2) can be written as

$$\begin{aligned}\phi(\rho,\theta) &= (2\pi)^{-3/2}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3\\ &\times\sum_{(t)=-\infty}^{\infty}[\alpha(\underline{\kappa},t)J_{t_1}(\kappa_1\rho^1)J_{t_2}(\kappa_2\rho^2)J_{t_3}(\kappa_3\rho^3)J_{t_4}(\omega\rho^4)\\ &+\gamma(\underline{\kappa},t)Y_{t_1}(\kappa_1\rho^1)Y_{t_2}(\kappa_2\rho^2)Y_{t_3}(\kappa_3\rho^3)Y_{t_4}(\omega\kappa_4)]\,e^{it_a\theta^a},\end{aligned} \quad (4.7)$$

where $J_t(\kappa\rho)$, $Y_t(\kappa\rho)$ are the Bessel functions of the first and second kind and $\Sigma_{(t)}\equiv\Sigma_{t_1}\Sigma_{t_2}\Sigma_{t_3}\Sigma_{t_4}$. But the above expression is singular at $(\rho)=(0)$. To avoid such a singularity one should exclude the Bessel functions of the second kind in general. Therefore the solutions of the following form would be used:

$$\begin{aligned}\phi(\rho,\theta) &= \sum_{(t)=-\infty}^{\infty}\phi^{(t)},\\ \phi^{(t)} &\equiv (2\pi)^{-3/2}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3[\alpha(\underline{\kappa},t)J_{t_1}(\kappa_1\rho^1)\\ &\times J_{t_2}(\kappa_2\rho^2)J_{t_3}(\kappa_3\rho^3)J_{t_4}(\omega\rho^4)e^{it_a\theta^a}].\end{aligned} \quad (4.8)$$

One can also derive (4.8) from (3.2) using the Jacobi–Anger formulas $e^{iu\sin\varphi}=\Sigma_{t=-\infty}^{\infty}J_t(u)e^{it\varphi}$, $e^{iu\cos\varphi}=\Sigma_{t=-\infty}^{\infty}(i)^tJ_t(u)e^{it\varphi}$.

To investigate the physical properties of the ϕ field in the above series expansion the various consequences of a single (t)-mode $\phi^{(t)}$, $(t)\neq(0)$ would be looked into. For that purpose the integral constants (2.6) would be computed using $\phi^{(t)}$ field. At this stage a difficulty is encountered. If one allows p^4 to take negative as well as positive values in the process of integrations, physically unacceptable results emerge; for example, $N^{(t)}=0$. Therefore the range of p^4-integration would be restricted to $0<p^4<\infty$. Mathematically, it means that some of the derivatives of $\phi^{(t)}$ would be singular at $p^4=0$ and physically it means that there would be a barrier or sources along the hypersurface $p^4=0$, separating positive and negative energy parts of V_8. There are also usual sources at $(\rho)\to(\infty)$. With this understanding some of the computed integral constants (see Appendix II) are the following:

$$\begin{aligned}&K_\alpha^{(t)}=0,\quad H^{(t)}=\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3 f(\underline{\kappa},t)\omega(\underline{\kappa}),\\ &X_\alpha^{(t)}=0,\quad \sigma^{(t)}=0,\quad B^{(t)}=0,\\ &Q^{(t)}=-\frac{\pi e_0}{2}(t_1+t_2+t_3+t_4)\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3 f(\underline{\kappa},t),\\ &f(\underline{\kappa},t)=(2/\pi)|\alpha(\underline{\kappa},t)|^2\geqslant 0.\end{aligned} \quad (4.9)$$

In the above $f(\underline{\kappa},t)$ is the distribution function of the particles at (t)-mode in the wave-number space. $K_\alpha{}^{(t)}=X_\alpha{}^{(t)}=0$ can be physically understood by noticing that in the basis of Bessel functions the incoming and the outgoing waves are equally mixed. The charge constant $Q^{(t)}$ has a peculiar multiplier $\pi/2$ as well as an interesting factor $(t_1+t_2+t_3+t_4)$. If the antiparticles of the (t)-mode should have the opposite charge they must be associated with the $(-t)$-mode, i.e., with the field $\phi^{(-t)}$. The charge-conjugation operation could be defined as

$$\mathbb{C}\phi^{(t)}\equiv\phi^{(-t)}. \quad (4.10)$$

The various reflection properties (see Fig. 3) of the $\phi^{(t)}$-field can be brought under investigation by Eq. (4.8) [cf. (I-2.4a)–(I-2.4c)]. These consequences are straightforward and can be summarized in the following:

$$\begin{aligned}&\pi_\alpha\phi^{(t)}(\rho,\theta)\equiv\phi^{(t)}(\rho,\ldots,\theta^\alpha+\pi,\ldots)=\xi_\alpha\phi^{(t)}(\rho,\theta),\\ &\xi_\alpha=e^{it_\alpha\pi},\\ &\xi_p\equiv\xi_1\xi_2\xi_3=e^{i(t_1+t_2+t_3)\pi},\\ &T'\phi^{(t)}(\rho,\theta)\equiv\phi^{(t)}(\rho,\ldots,\theta^4+\pi)=\xi_{T'}\phi^{(t)}(\rho,\theta),\\ &\xi_{T'}=e^{it_4\pi},\end{aligned}$$

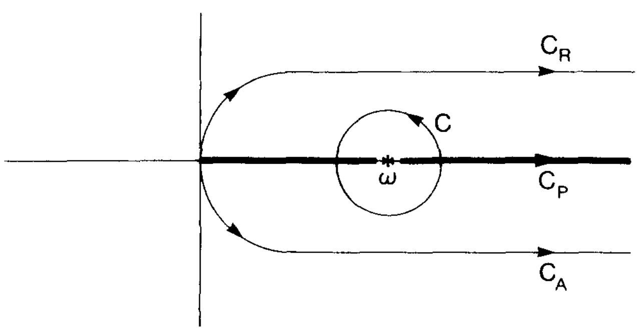

FIG. 4. The complex κ_4-plane.

$$
\begin{aligned}
&P_0\phi^{(t)}(\rho,\theta)\equiv\phi^{(t)}(\rho,\theta^i+\pi/2)=\xi_{P_0}\phi^{(t)}(\rho,\theta),\\
&\xi_{P_0}=e^{i(t_1+t_2+t_3+t_4)\pi/2},\\
&\xi_P\xi_{T'}=(\xi_{P_0})^2.
\end{aligned}
\tag{4.11}
$$

Furthermore, a plausible mass operator will be introduced below:

$$
M^2(\rho,\partial\rho)\phi^{(t)}(\rho,\theta)\equiv-\{\eta^{ij}[\partial_{\rho^i}\partial_{\rho^j}+(1/\rho^i)\partial_{\rho^j}]\}\times\phi^{(t)}(\rho,\theta).
\tag{4.12}
$$

By (4.4) one obtains

$$
(\text{mass})^2=m^2-\eta^{ij}(t_it_j/\rho^i\rho^j).
\tag{4.13}
$$

Excitation of all four of (t)-modes or any of the larger t_i-modes would require very high energy and thus be less probable. Most likely is the excitation of one or two t_i-modes to some lower quantum values.

For an interesting special case $t_1=t_2=0$, t_3 and $s/2\equiv t_4$ being not necessarily zero, and with the definition of the charge parameter $e\equiv-(\pi/2)e_0$, one can collect from (4.9), (4.11), (4.13), the following formulas:

$$
H^{(t_3,s)}=\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3 f(\kappa,t_3,s)\omega(\kappa),
\tag{4.14a}
$$

$$
Q^{(t_3,s)}=e(t_3+s/2)\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3 f(\kappa,t_3,s),
\tag{4.14b}
$$

$$
\xi_p=e^{it_3\pi},\quad \xi_{T'}=e^{is\pi/2},\quad \xi_{P_0}=e^{i(t_3+s/2)(\pi/2)},
\tag{4.14c}
$$

$$
(\text{mass})^2=(m^2)+(t_3^2/\rho_3^2)-(s^2/4\rho_4^2).
\tag{4.14d}
$$

Equation (4.14b) can be identified with the Gell-Man–Nishijima's formula[6] for the meson charge provided the space tremor quantum number t_3 is identified with the isotropic quantum number and twice the time tremor quantum number t_4 is identified with the strangeness s. Equation (4.14d) is the analog of the Okubo mass formula[8] for the mesons. One can generalize in a straightforward manner the above formulas for the case when $\phi(\rho,\theta)$ is in general double valued function around the origin, or in other words when $t_i=0,\pm1/2,\pm1,\pm3/2,\cdots$. With this understanding, applications of the above results could be made to the mesons in Table I.

By Eq. (4.8) infinite numbers of mesons exist, so that Table I can be continued as long as one wishes, predicting new mesons on the way. But it has been confined to the discussions of the most familiar mesons. Moreover, the old assignments of the quantum numbers have been retained for t_3 and s, although such adoption is not absolutely imperative from the theory. Comparing with the usual SU_3-model mainly three differences have arisen here.

(i) π^0 in the present scheme has parity $\xi_P=1$ instead of -1.

(ii) The parity of the κ mesons satisfy $\xi_P^2=-1$, which is a reminder of the parity of the spinor fields. This situation has been forced into by the choice of the half-integral values for the quantum numbers t_i.

(iii) Each of the mesons (except for π^0) is one of the three kinds which are practically indistinguishable. They can only be separated out by measuring the individual parities ξ_1, ξ_2, ξ_3.

TABLE I.

Mesons		Tremor Quantum Numbers				Charge	Parity				Time-Parity	Born Number
$(J=B=0)$		t_2	t_2	t_3	$s=2t_4$	$e(t_1+t_2+t_3+s/2)$	$\xi_1=e^{it_1\pi}$	$\xi_2=e^{it_2\pi}$	$\xi_3=e^{it_3\pi}$	$\xi_P=\xi_1\xi_2\xi_3$	$\xi_{T'}=e^{is\mu/2}$ $\xi_{T'_0}$	
π^0	0	0	0	0	0	1	1	1	1	1	1	
	$\pi_1^\pm$	±1	0	0	0	$\pm e$	-1	1	1	-1	1	$\pm i$
$\pi^\pm$	$\pi_2^\pm$	0	±1	0	0	$\pm e$	1	-1	1	±1	1	$\pm i$
	$\pi_3^\pm$	0	0	±1	0	$\pm e$	1	1	-1	-1	1	$\pm i$
	$\kappa_1^\pm$	$\pm\frac{1}{2}$	0	0	±1	$\pm e$	$\pm i$	1	1	$\pm i$	$\pm i$	$\pm i$
$\kappa^\pm$	$\kappa_2^\pm$	0	$\pm\frac{1}{2}$	0	±1	$\pm e$	1	$\pm i$	1	$\pm i$	$\pm i$	$\pm i$
	$\kappa_3^\pm$	0	0	$\pm\frac{1}{2}$	±1	$\pm e$	1	1	$\pm i$	$\pm i$	$\pm i$	$\pm i$
	$\kappa_1^0\bar{\kappa}_1^0$	$\mp\frac{1}{2}$	0	0	±1	0	$\mp i$	1	1	$\mp i$	$\pm i$	1
$\kappa^0\bar{\kappa}^0$	$\kappa_2^0\bar{\kappa}_2^0$	0	$\mp\frac{1}{2}$	0	±1	0	1	$\mp i$	1	$\mp i$	$\pm i$	1
	$\kappa_3^0\bar{\kappa}_3^0$	0	0	$\mp\frac{1}{2}$	±1	0	1	1	$\mp i$	$\mp i$	$\pm i$	$\pm i$
	1											

For a convincing assignment of the quantum numbers t_i, the experimental informations on ξ_1, ξ_2, ξ_3, $\xi_{T'}$ are definitely required.

ACKNOWLEDGMENTS

The author is indebted to NRC Grant No. 3993. He also wishes to thank Prof. D.L. Sharma for some information discussions on Bessel functions.

APPENDIX I

The derivation of

$$X_a = \int_{D_7} d^3\kappa\, d^4x\, [f_+(k,x) + f_-(k,x)]\, x_a, \quad \text{(AI1)}$$

would be performed here. Recall the representation of the delta function

$$\delta(y) = (2\pi)^{-1} \int_{-\infty}^{\infty} e^{iyu}\, du. \quad \text{(AI2)}$$

From (3.2) one gets

$$[\partial_{q^\mu}\phi]_{q^4=0} = (2\pi)^{-7/2} \int_{D_7} d^3k\, d^4x\, (2\Omega)^{-1/2} (ik_\mu) \times [-\alpha e^{-i(k_\nu q^\nu + x_a p^a)} + \bar{\beta} e^{i(k_\nu q^\nu + x_a p^a)}],$$

$$[\partial_{q^4}\phi]_{q^4=0} = (2\pi)^{-7/2} \int_{D_7} d^3k\, d^4x\, (2\Omega)^{-1/2} (i\Omega) \times [-\alpha e^{-i(k_\nu q^\nu + x_a p^a)} + \bar{\beta} e^{i(k_\nu q^\nu + x_a p^a)}],$$

$$[\partial_{q^a}\phi]_{q^4=0} = (2\pi)^{-7/2} \int_{D_7} d^3k\, d^4x\, (2\Omega)^{-1/2} (ix_a) \times [-\alpha e^{-i(k_\nu q^\nu + x_a p^a)} + \bar{\beta} e^{i(k_\nu q^\nu + x_a p^a)}], \quad \text{(AI3)}$$

From (2.6), (AI.2), (AI.3), it follows that

$$\begin{aligned} X_c = [X_c]_{q^4=0} &= (2\pi)^{-7} \int_{V_7} d^3q\, d^4p \int_{D_7} d^3k\, d^4x\, (i\Omega)^{-1/2} \\ &\times \int_{D_7} d^3k'\, d^4x'\, (2\Omega')^{-1/2} \{(i\Omega')(ix_c) \\ &\times [\bar{\alpha}' e^{i(k'_\mu q^\mu + x'_a p^a)} - \beta' e^{-i(k'_\mu q^\mu + x'_a p^a)}] \\ &\times [-\alpha e^{-i(k_\nu q^\nu + x_b p^b)} + \bar{\beta} e^{i(k_\nu q^\nu + x_b p^b)}] \\ &+ (\text{c.c.})\} \\ &= \int_{D_7} d^3k\, d^4x\, (2\Omega)^{-1/2} \int_{D_7} d^3k'\, d^4x'\, (2\Omega')^{-1/2} \\ &\times \{[\Omega' x_c (\bar{\alpha}'\alpha + \beta'\bar{\beta}) + \Omega x'_c (\alpha'\bar{\alpha} + \bar{\beta}'\beta)] \\ &\times \delta^3(k - k')\delta^4(x - x') - [\Omega' x_c (\bar{\alpha}'\bar{\beta} + \beta'\alpha) \\ &+ \Omega x'_c (\alpha'\beta - \bar{\beta}'\bar{\alpha})]\delta^3(k + k')\delta^4(x + x')\} \\ &= \int_{D_7} d^3k\, d^4x\, (|\alpha|^2 + |\beta|^2) x_c - (\tfrac{1}{2}) \int_{D_7} d^3k\, d^4x \\ &\times \{[\bar{\alpha}(-k, -x)\bar{\beta}(k,x) - \bar{\alpha}(k,x)\bar{\beta}(-k, -x)] \\ &+ [\alpha(k,x)\beta(-k, -x) - \alpha(-k, -x)\beta(k,x)]\}\, x_c. \end{aligned}$$

The last integral vanishes because whenever $(k,x) \in D_7$ also the point $(-k, -x) \in D_7$. Thus (AI1) is obtained.

APPENDIX II

Some of the computations of the integral constants (4.6) would be demonstrated here. For that purpose some basic equations regarding the Bessel function of the first kind would be listed below.[10]

$$(d/d\rho) J_t(\rho) = (1/2)[J_{t-1}(\rho) - J_{t+1}(\rho)], \quad \text{(AII1a)}$$

$$J_{t-1}(\rho) + J_{t+1}(\rho) = (2t/\rho) J_t(\rho), \quad \text{(AII1b)}$$

$$\begin{aligned} \partial_q\, [J_t(\kappa\rho) e^{\pm it\theta}] &= [\cos\theta\partial_\rho - (1/\rho)\sin\theta\partial_\theta]\, [J_t(\kappa\rho) e^{\pm it\theta}] \\ &= (\kappa/2)\, [J_{t-1} e^{\mp i\theta} - J_{t+1} e^{\pm i\theta}]\, e^{\pm it\theta} \\ &= [(1/2)\kappa\cos\theta(J_{t-1} - J_{t+1}) \mp (it/\rho)\sin\theta J_t]\, e^{\mp it\theta}, \end{aligned} \quad \text{(AII1c)}$$

$$\begin{aligned} \partial_p\, [J_t(\kappa\rho) e^{\pm it\theta}] &= [\sin\theta\partial_\rho + (1/\rho)\cos\theta\partial_\theta]\, [J_t(\kappa\rho) e^{\pm it\theta}] \\ &= \pm (i\kappa/2)(J_{t-1} e^{\mp i\theta} + J_{t+1} e^{\pm i\theta})\, e^{\pm i\theta} \\ &= [(1/2)\kappa\sin\theta(J_{t-1} - J_{t+1}) \pm (it/\rho)\cos\theta J_t]\, e^{\pm it\theta}, \end{aligned} \quad \text{(AII1d)}$$

where the upper and lower signs should be read separately. Further equations are

$$\int_0^\infty d\rho\, \rho J_t(\kappa\rho) J_t(\kappa'\rho) = (\kappa)^{-1}\delta(\kappa - \kappa'), \quad \text{(AII1e)}$$

$$\int_0^\infty d\rho(\rho)^{-1}\, [J_t(\Omega\rho)]^2 = (2t)^{-1}, \quad \text{(AII1f)}$$

$$\begin{aligned} \int_0^\infty d\rho(\rho)^{-u} J_t(\kappa\rho) J_{t'}(\kappa\rho) &= (2)^{-1}(\kappa/2)^{u-1}\Gamma(u)\Gamma[(t + t' - u + 1)/2] \\ &\times \{\Gamma[(u + t + t' + 1)/2]\Gamma[(u + t - t' + 1)/2] \\ &\times \Gamma[(u - t + t' + 1)/2]\}^{-1}, \end{aligned} \quad \text{(AII1g)}$$

where the above integral is valid for $0 < u < t + t' + 1$ and is the critical case of the Weber–Schafheitlin integral. Some combinations of such integrals do exist by analytic continuations of the original restrictions on u,t,t'. To prove that the following formulas on the gamma functions would be needed. $\Gamma(z)$ has poles at $z = 0, -1, -2, \cdots$, and furthermore,

$$\Gamma(z + 1) = z\Gamma(z), \quad \text{(AII1h)}$$

$$\Gamma(z)\Gamma(1 - z) = -z\Gamma(z)\Gamma(-z) = \pi\csc(\pi z), \quad \text{(AII1i)}$$

$$\Gamma(2z) = 2^{2z-1}\pi^{-1/2}\Gamma(z + 1/2). \quad \text{(AII1j)}$$

Then from (AII1g)–(AII1j) it follows that

$$\begin{aligned} \int_0^\infty d\rho\, \rho(J_{t-1}^2 + J_{t+1}^2 - 2J_t^2) &= -\lim_{u\to -1} \Big[\pi^{-1/2}(\kappa/2)^{u-1} 2^{2u-1}\Gamma[(u + 2)/2] \\ &\times [4t^2 - (u + 1)^2]^{-1} \Big] \left(\frac{\Gamma[(2t - u + 1)/2]}{\Gamma[2t + u + 1)/2]} \right) \\ &\times \left(\frac{1}{\Gamma[(u + 1)/2]} \right) \Big] \\ &= 0, \end{aligned} \quad \text{(AII1k)}$$

$$\begin{aligned} \int_0^\infty d\rho\, (J_t^2 + J_{t-1} J_{t+1}) &= \lim_{u\to 0} \left[\frac{(2)^{-1}(\kappa/2)^{u-1}\Gamma(u)\Gamma[(2t - u + 1)/2]}{\Gamma[(2t + u + 1)/2]\Gamma[(u + 1)/2]\Gamma[(u - 1)/2]} \right. \\ &\times \left. \left(\frac{2}{(u - 1)} + \frac{2}{(u + 1)} \right) \right] \\ &= -4\kappa^{-1}[\Gamma(1/2)\Gamma(-1/2)]^{-1} = 2(\pi\kappa)^{-1}. \end{aligned} \quad \text{(AII1l)}$$

To calculate the integral constants $N^{(t)}$, $Q^{(t)}$, $H^{(t)}$, respectively, at $q^4=0$ with the restriction $0<p^4<\infty$, one has to put $\theta^4=\pi/2$. Therefore from (AII1c) it follows that

$$\phi^{(t)}=(2\pi)^{-3/2}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3\alpha(\kappa,t)e^{it_a\theta^a}J_{t_1}(\kappa_1\rho^1)\cdots J_{t_4}(\omega\rho^4),$$

$$[\partial_{q^4}\bar\phi^{(t)}]_{\theta^4=\pi/2}=it_4(2\pi)^{-3/2}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3(\rho^4)^{-1}\times(-1)^{t_4}\bar\alpha e^{-it_\mu\theta^\mu}J_{t_1}(\kappa_1\rho^1)\cdots J_{t_4}(\omega\rho^4). \quad \text{(AII2)}$$

Now from (2.6), (AII2) and the restriction $0<p^4<\infty$ the expression for $N^{(t)}$is

$$N^{(t)}=[N^{(t)}]_{\theta^4=\pi/2}$$
$$=\left(-i\int_{-\infty}^\infty d^3q\,d^3p\int_0^\infty dp^4\,[(\partial_{q^4}\bar\phi^{(t)}\phi^{(t)}]_{\theta^4=\pi/2}\right)+(\text{c.c.})$$
$$=\Big((t_4(2\pi)^{-3}\int_0^{2\pi}d^3\theta\int_0^\infty d^3\rho\,\rho^1\rho^2\rho^3\int_0^\infty\frac{d\rho^4}{\rho^4}\times\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3$$
$$\times\int_0^\infty d^3\kappa'\kappa_1'\kappa_2'\kappa_3'\;[(-i)]^{t_4}\,\bar\alpha e^{-it_\mu\theta^\mu}\,[(i)^{t_4}\alpha' e^{it_\nu\theta^\nu}]$$
$$\times J_{t_1}(\kappa_1\rho^1)J_{t_1}(\kappa_1'\rho^1)\cdots J_{t_4}(\omega\rho^4)J_{t_4}(\omega'\rho^4)\Big)+(\text{c.c.}). \quad \text{(AII3)}$$

Now the θ-integrations are performed. Then working out $\rho^\alpha,\rho^4,\kappa_\alpha'$ integrals using (AII1e), (AII1f) one finally arrives at

$$N^{(t)}=\left((\tfrac{1}{2})\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3|\alpha|^2\right)+(\text{c.c.})$$
$$=\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3|\alpha(\kappa,t)|^2. \quad \text{(AII4)}$$

Next from (2.6) consider the integral constant

$$Q^{(t)}=e_0(I_1+\bar I_1+I_2),$$
$$I_1\equiv\int_{-\infty}^\infty d^3q\,d^3p\int_0^\infty dp^4\,\{\partial_{q^4}\bar\phi^{(t)}\times(q^i\partial_{p^i}-p^i\partial_{q^i})\phi^{(t)}\}_{\theta^4=\pi/2},$$
$$I_2\equiv\int_{-\infty}^\infty d^3q\,d^3p\int_0^\infty dp^4\,(p^4\{\eta^{ij}[\partial_{q^i}\bar\phi^{(t)}\partial_{q^j}\phi^{(t)}+\partial_{p^i}\bar\phi^{(t)}+\partial_{p^j}\phi^{(t)}-m^2\bar\phi^{(t)}\phi^{(t)}\})_{\theta^4=\pi/2}. \quad \text{(AII5)}$$

To evaluate I_1 it should be noted that

$$[(q^i\partial_{p^i}-p^i\partial_{q^i})\phi^{(t)}]_{\theta^4=\pi/2}$$
$$=[(\partial_{\theta^1}+\partial_{\theta^2}+\partial_{\theta^3}+\partial_{\theta^4})\,\phi^{(t)}]_{\theta^4=\pi/2}$$
$$=i(t_1+t_2+t_3+t_4)[\phi^{(t)}]_{\theta^4=\pi/2}. \quad \text{(AII6)}$$

From (AII1e), (AII1f), (AII2), (AII5), (AII6) it follows that

$$I_1=-(t_1+t_2+t_3+t_4)t_4(2\pi)^{-3}\times\int_0^{2\pi}d^3\theta\int_0^\infty d^3\rho\,\rho^1\rho^2\rho^3\int_0^\infty\frac{d\rho^4}{\rho^4}$$
$$\times\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_2\int_0^\infty d^3\kappa'\kappa_1'\kappa_2'\kappa_3'\bar\alpha\alpha'$$
$$\times J_{t_1}(\kappa_1\rho^1)J_{t_1}(\kappa_1'\rho^1)\cdots J_{t_4}(\omega\rho^4)J_{t_4}(\omega'\rho^4)$$
$$=-(1/2)(t_1+t_2+t_3+t_4)\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3\,|\alpha(\kappa,t)|^2. \quad \text{(AII7)}$$

Now in the second integral I_2 the individual terms diverge but the combinations exist. To prove that the usual definition of an improper integral can be employed, and that is

$$I_2=\lim_{L\to\infty}\,[I_2^{(i)}(L)+I_2^{(ii)}(L)+I_2^{(iii)}(L)+I_2^{(iv)}(L)],$$
$$I_2^{(i)}(L)\equiv-\int_0^{2\pi}d^3\theta\int_0^\infty d^3\rho\,\rho^1\rho^2\rho^3\int_0^L d\rho^4\rho^4\times[\partial_{q^\mu}\bar\phi^{(t)}\cdot\partial_{q^\mu}\phi^{(t)}]_{\theta^4=\pi/2},$$
$$I_2^{(ii)}(L)\equiv-\int_0^{2\pi}d^3\theta\int_0^\infty d^3\rho\,\rho^1\rho^2\rho^3\int_0^L d\rho^4\rho^4\times[\partial_{q^\mu}\phi^{(t)}\cdot\partial_{q^\mu}\phi^{(t)}]_{\theta^4=\pi/2},$$
$$I_2^{(iii)}(L)=\int_0^{2\pi}d^3\theta\int_0^\infty d^3\rho\,\rho^1\rho^2\rho^3\int_0^L d\rho^4\rho^4\times[\partial_{q^4}\bar\phi^{(t)}\cdot\partial_{q^4}\phi^{(t)}]_{\theta^4=\pi/2},$$
$$I_2^{(iv)}(L)=\int_0^{2\pi}d^3\theta\int_0^\infty d^3\rho\,\rho^1\rho^2\rho^3\int_0^L d\rho^4\rho^4\times[\partial_{q^4}\bar\phi^{(t)}\cdot\partial_{q^4}\phi^{(t)}]_{\theta^4=\pi/2},$$
$$I_2^{(v)}(L)=-m^2\int_0^{2\pi}d^3\theta\int_0^\infty d^3\rho\,\rho^1\rho^2\rho^3\int_0^L d\rho^4\rho^4\times[\bar\phi^{(t)}\phi^{(t)}]_{\theta^4=\pi/2}. \quad \text{(AII8)}$$

Using (AII1c), (AII1e), (AII2) one has for a part of

$$I_2^{(i)}(L)$$
$$-\int_0^{2\pi}d^3\theta\int_0^\infty d^3\rho\,\rho^1\rho^2\rho^3\int_0^L d\rho^4\rho^4\times[\partial_{q^1}\bar\phi^{(t)}\cdot\partial_{q^1}\phi^{(t)}]_{\theta^4=\pi/2}$$
$$=-(2\pi)^{-3}\int_0^{2\pi}d^3\theta\int_0^\infty d^3\rho\,\rho^1\rho^2\rho^3\int_0^L d\rho^4\rho^4$$
$$\times\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3\int_0^\infty d^3\kappa'\,\kappa_1'\kappa_2'\kappa_3'$$
$$\times(\kappa_1\kappa_1'/4)\bar\alpha\alpha'\,(J_{t_1-1}e^{i\theta^1}-J_{t_1+1}e^{-i\theta^1})$$
$$\times(J'_{t_1-1}e^{-i\theta^1}-J'_{t_1+1}e^{i\theta^1})$$
$$\times J_{t_2}(\kappa_2\rho^2)J_{t_2}(\kappa_2'\rho^2)\cdots J_{t_4}(\omega\rho^4)J_{t_4}(\omega'\rho^4)$$
$$=-(1/4)\int_0^\infty d^3\rho\,\rho^1\rho^2\rho^3\int_0^L d\rho^4\rho^4$$
$$\times\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3\int_0^\infty d^3\kappa'\,\kappa_1'\kappa_2'\kappa_3'\kappa_1\kappa_1'\bar\alpha\alpha'$$
$$\times(J_{t_1-1}J'_{t_1-1}+J_{t_1+1}J'_{t_1+1})\,J_{t_2}J'_{t_2}\cdots J_{t_4}J'_4$$
$$=-\tfrac{1}{2}\int_0^L d\rho^4\rho^4\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3\,|\alpha|^2\kappa_1^2(J_{t_4})^2. \quad \text{(AII9)}$$

Likewise using (AII1a)–(AII1f) all other integrals can be expressed as

$$I_2^{(i)}(L)=-\tfrac{1}{2}\int_0^L d\rho^4\rho^4\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3|\alpha|^2\kappa_\mu\kappa_\mu(J_{t_4})^2,$$
$$I_2^{(ii)}(L)=I_2^{(i)}(L),$$

$$I_2^{(iii)}(L) = \tfrac{1}{4}\int_0^L d\rho^4 \rho^4 \int_0^\infty d^3\kappa\, \kappa_1\kappa_2\kappa_3 |\alpha|^2\omega^2(J_{t_4-1} + J_{t_4+1})^2,$$

$$I_2^{(iv)}(L) = \tfrac{1}{4}\int_0^L d\rho^4 \rho^4 \int_0^\infty d^3\kappa\, \kappa_1\kappa_2\kappa_3 |\alpha|^2\omega^2(J_{t_4-1} + J_{t_4+1})^2,$$

$$I_2^{(v)}(L) = -m^2 \int_0^L d\rho^4 \rho^4 \int_0^\infty d^3\kappa\, \kappa_1\kappa_2\kappa_3\, |\alpha|^2 (J_{t_4})^2. \tag{AII10}$$

Therefore

$$I_2^{(i)}(L) + I_2^{(ii)}(L) + I_2^{(iii)}(L) + I_2^{(iv)}(L) + I_2^{(v)}(L) = \int_0^L d\rho^4 \rho^4 \int_0^\infty d^3\kappa\, \kappa_1\kappa_2\kappa_3\, |\alpha|^2 \{ -(\kappa_\mu\kappa_\mu + m^2)(J_{t_4})^2 + (\omega^2/4)[(J_{t_4-1} + J_{t_4+1})^2 + (J_{t_4-1} - J_{t_4+1})^2]\}. \tag{AII11}$$

Using (4.6), (AII1k), (AII8), (AII11) one obtains

$$I_2 = \tfrac{1}{2}\int_0^\infty d^3\kappa\, \kappa_1\kappa_2\kappa_3\, |\alpha|^2\omega^2 \int_0^\infty d\rho^4 \rho^4 \times (J^2_{t_4-1} + J^2_{t_4+1} - 2J^2_{t_4}) = 0. \tag{AII12}$$

Therefore from (AII5), (AII7), (AII12) it follows that

$$Q^{(t)} = -e_0(t_1 + t_2 + t_3 + t_4) \int_0^\infty d^3\kappa\, \kappa_1\kappa_2\kappa_3\, |\alpha(\kappa,t)|^2. \tag{AII13}$$

Next the integral constant $H^{(t)}$ will be calculated. From (2.6)

$$H^{(t)} = \lim_{L\to\infty} [I_1(L) + I_2(L) + I_3(L) + I_4(L) + I_5(L)],$$

$$I_1(L) \equiv \int_0^{2\pi} d^3\theta \int_0^\infty d^3\rho\, \rho^1\rho^2\rho^3 \times \int_0^L d\rho^4 \{[\bar\partial_{q^\mu}\bar\phi^{(t)}]\cdot[\partial_{q^\mu}\phi^{(t)}]\}_{\theta^4=\pi/2},$$

$$I_2(L) \equiv \int_0^{2\pi} d^3\theta \int_0^\infty d^3\rho\, \rho^1\rho^2\rho^3 \times \int_0^L d\rho^4 \{[\bar\partial_{p^\mu}\bar\phi^{(t)}]\cdot[\partial_{p^\mu}\phi^{(t)}]\}_{\theta^4=\pi/2},$$

$$I_3(L) \equiv \int_0^{2\pi} d^3\theta \int_0^\infty d^3\rho\, \rho^1\rho^2\rho^3 \times \int_0^L d\rho^4 \{[\bar\partial_{q^4}\bar\phi^{(t)}]\cdot[\partial_{q^4}\phi^{(t)}]\}_{\theta^4=\pi/2},$$

$$I_4(L) \equiv -\int_0^{2\pi} d^3\theta \int_0^\infty d^3\rho\, \rho^1\rho^2\rho^3 \times \int_0^L d\rho^4 \{[\bar\partial_{p^4}\bar\phi^{(t)}]\cdot[\partial_{p^4}\phi^{(t)}]\}_{\theta^4=\pi/2},$$

$$I_5(L) \equiv m^2 \int_0^{2\pi} d^3\theta \int_0^\infty d^3\rho\, \rho^1\rho^2\rho^3 \times \int_0^L d\rho^4\, [\bar\phi^{(t)}\phi^{(t)}]_{\theta^4=\pi/2}. \tag{AII14}$$

Evaluating as before one can obtain

$$I_1(L) = \tfrac{1}{2}\int_0^L d\rho^4 \int_0^\infty d^3\kappa\, \kappa_1\kappa_2\kappa_3\, |\alpha|^2 \kappa_\mu\kappa_\mu J^2_{t_4},$$

$$I_2(L) = I_1(L),$$

$$I_3(L) = \tfrac{1}{4}\int_0^L d\rho^4 \int_0^\infty d^3\kappa\, \kappa_1\kappa_2\kappa_3\, |\alpha|^2\omega^2(J_{t_4-1} + J_{t_4+1})^2,$$

$$I_4(L) = -\tfrac{1}{4}\int_0^L d\rho^4 \int_0^\infty d^3\kappa\, \kappa_1\kappa_2\kappa_3\, |\alpha|^2\omega^2(J_{t_4-1} + J_{t_4+1})^2,$$

$$I_5(L) = m^2 \int_0^L d\rho^4 \int_0^\infty d^3\kappa\, \kappa_1\kappa_2\kappa_3\, |\alpha|^2 J^2_{t_4}. \tag{AII15}$$

Using (4.6), (AII1l), (AII14), (AII15) it follows that

$$H^{(t)} = \int_0^\infty d^3\kappa\, \kappa_1\kappa_2\kappa_3\, |\alpha|^2\omega^2 \int_0^\infty d\rho^4 (J^2_{t_4} + J_{t_4-1}J_{t_4+1}) = \frac{2}{\pi}\int_0^\infty d^3\kappa\, \kappa_1\kappa_2\kappa_3\, |\alpha(\kappa,t)|^2\omega. \tag{AII16}$$

[1]M. Born, Proc. R. Soc. London Ser. A **165**, 291 (1938); Nature **163**, 207 (1949); Also, see H.S. Green, Nature **163**, 208 (1949).
[2]H. Yukawa, Phys. Rev. **76**, 300 (1949).
[3]A. Das, J. Math. Phys. **7**, 45 (1966).
[4]A. Das, J. Math. Phys. **21**, 1506 (1980).
[5]L. Ballentine, Rev. Mod. Phys. **42**, 358 (1970).
[6]M. Gell-Man, California Inst. of Tech. Lab Report CTSL-20 (1961).
[7]Equation (6.1) in the statement of one of the NO-GO theorems as stated in L. O'Raifeartaigh, Phys. Rev. **139**, 258 (1965).
[8]S. Okubo, Prog. Theor. Phys. (Kyoto) **27**, 949 (1962).
[9]P.M. Morse and H. Feshbach, *Methods of Theoretical Physics, Volume* I (McGraw-Hill, New York, 1953), p. 825.
[10]G.N. Watson, *A Treatise of the Theorey of Bessel Functions*, 2nd. ed. (Cambridge U.P. 1966), p. 403.

Extended phase space. III. Unified spin-1/2 fields

A. Das
Department of Mathematics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

(Received 30 October 1979; accepted for publication 17 January 1980)

In extended phase space V_8, the classical field equation for spin-1/2 elementary particles is written as $[v^k\partial/\partial q^k + a^k\partial/\partial p^k - imI]\psi(q,p) = 0$. The 16×16 matrices v^k, a^k stand for the instantaneous 4-velocity and 4-acceleration. The equation is called the Boltzmann–Dirac–Yukawa (in short BDY) equation. This equation treats q,p variables on equal footings and is covariant under the extended Poincaré group P_8. The spin-1/2 field ψ is decomposed into plane wave solutions, and integral constants such as the total energy, the total momentum, the average position, etc., are computed. These integral constants become meaningful provided the modulus squared of each amplitude is interpreted as the statistical distribution function. Next the ψ field is subject to the Hankel transform $\psi(\rho,\theta)\sim\Sigma_{t=-\infty}^{\infty}\Sigma_{R=1}^{8}\Sigma_{\alpha=1}^{2}\int_0^\infty d\kappa\,\kappa[\beta_{(R\alpha)}u^{(R\alpha)} + \bar{\gamma}_{(R\alpha)}v^{(R\alpha)}]J_t(\kappa\rho)e^{it\theta} = \Sigma_{t=-\infty}^{\infty}\psi^{(t)}$; $\rho = (q^2+p^2)^{1/2}$, $\theta = \arctan(p/q)$. The integral constants are constructed from a single (t)-mode $\psi^{(t)}$. These turn out to be physically meaningful for eight spin-1/2 particles. Specially the total charge can be identified with Gell-Mann–Nishijima's expression for the baryon provided the quantum number t_3 corresponds to the isotopic spin, $2t_4$ is identified with the strangeness, and the baryon number b is allowed to take 0, $\pm$ 1,2. With each of the (t)-mode $\psi^{(t)}$, eight baryon fields can be associated so that ψ stands for the unified baryon fields. Finally some brief comments are made on the possibility of treating lepton fields within the framework of the BDY equation.

I. INTRODUCTION

The usual Dirac equation is generalized in extended phase space V_8 to the equation $[v^k\partial/\partial q^k + a^k\partial/\partial p^k]\psi(q,p) = im\psi(q,p)$. Here the 16×16 matrices v^k, a^k satisfy $v^kv^l + v^lv^k = a^ka^l + a^la^k = 2\eta^{kl}I$, $v^ka^l + a^lv^k = 0$.

The v^k, a^k matrices correspond to the 4-velocity and 4-acceleration and the right-hand side of the equation corresponds to the contribution from the self-collision so that the equation has a striking resemblance to Boltzmann's transport equation. This equation is covariant under the extended Poincaré group P_8 acting in V_8 and thus obviously invariant under the reciprocity transformation $q' = -p$, $p' = q$. Yukawa[1] has already treated the reciprocity-invariant spin-1/2 particle equations. He had three matrix equations, one for the space–time variables, another for the momentum–energy space, and the third for a consistency condition. In the present treatment the first two of Yukawa's equations are "fused together." All these considerations have led to the christening of the present spin-1/2 field equation as the Boltzmann–Dirac–Yukawa (or, in short BDY) equation. The BDY equation has two advantages over Yukawa's original three equations. First, the group of covariance is broadened and second the associated ψ-field can describe infinitely many spin-1/2 particles in a unified fashion. The BDY equation is formally identical with the spin-1/2 field equation written for the complex space–time[2] model.

In Sec. II the Lagrangian formalism developed[3] in Paper I is applied to the spin-1/2 field ψ. The integral constants such as the total energy, the total momentum, the average position, the total charge, etc., are expressed as 7-dimensional integrals in the extended phase space.

In Sec. III the Fourier transform method or the plane wave decomposition for the ψ-fields is performed. Again the integral constants are computed from the plane wave solutions. They are expressible as 7-dimensional integrals in the dual extended phase space. The corresponding integrands are proportional to $f_R(k,x)\equiv|\alpha_R(k,x)|^2$, $R = 1,\ldots,8$, where $\alpha_R(k,x)$ are amplitude functions of k,x, the dual momentum and position–time. It is only natural to interpret $f_R(k,x)$ as a statistical distribution function for the spin-1/2 particles.

In Sec. IV the variables $\rho = (p^2+q^2)^{1/2}$, $\theta = \arctan\times(p/q)$ are introduced in place of q,p. This step leads in an obvious fashion to the Fourier–Bessel integral $\psi(\rho,\theta)\sim\Sigma_{t=-\infty}^{\infty}\Sigma_{R=1}^{8}\Sigma_{\alpha=1}^{2}\int_0^\infty d\kappa\,\kappa[\beta_{(R\alpha)}u^{(R\alpha)} + \bar{\gamma}_{(R\alpha)}v^{(R\alpha)}]\times J_t(\kappa\rho)e^{it\theta} = \Sigma_{t=-\infty}^{\infty}\psi^{(t)}$.

The quantum numbers $(t) = (t_1,t_2,t_3,t_4)$ represent oscillations or tremors in space–time. The integral constants like the total energy, the total charge, etc. are computed via Lagrangian formalism for a single (t)-mode $\psi^{(t)}$. These 3-dimensional integrals turn out to be physically meaningful. They are contributed by eight baryon particles. Particularly, the total charge Q is proportional to $(t_1+t_2+t_3+t_4+b)$, $b = 0, \pm 1,2$. For the special case $t_1 = t_2 = 0$, $s = 2t_4$ the charge formula corresponds to Gell-Mann–Nishijima's[4] expression for the baryons, provided baryon number b takes values 0, $\pm$ 1,2. In addition if one accepts the (mass)^2-operator as $-\eta^{ij}[\partial^2/\partial\rho^i\partial\rho^j + (1/\rho^i)\partial/\partial\rho^j]$ instead of the usual[5] one $-\eta^{ij}(\partial^2/\partial q^i\partial q^j)$, then the $\psi^{(t)}$ field will have (mass)$^2 = m^2I - \mathrm{diag}[\eta^{ij}t_i^Lt_j^L/\rho^i\rho^j]$, $L = 1,\ldots,16$. Specially for the case $t_1 = t_2 = 0$, $s = 2t_4$, one has (mass)$^2 = m^2 + (t_3+c_1)^2/\rho_3^2 - (s+c_2)/4\rho_4^2 + \theta_1(\rho^1,\rho^2)$. This mass formula has some resemblance to Okubo's[6] expression. It may be mentioned that Nicolson's asymptotic formula[7] for large order is

$$J_t(\kappa\rho)\sim 3^{-2/3}(\xi/\kappa\rho)^{1/3}[J_{1/3}(\xi) + J_{-1/3}(\xi)],$$
$$\xi = (2/3)(\kappa\rho/2)^{-1/2}|\kappa\rho - t|^{-3/2},\quad \kappa\rho > t.$$

 0022-2488/80/061521-12$1.00

In the present context this formula might suggest that particles with very large quantum numbers (t) behave as quarks and that is why quarks would be very difficult to produce.

In Sec. V the possibility of dealing with lepton fields as the special solutions $\psi(q,p) = \psi_1(q)\times\psi_2(p)$ of the BDY equation is briefly mentioned.

In Appendices I, II, and III, the rigorous derivations of the integral constants from the Fourier as well as Hankel transforms of the ψ-field are furnished.

II. NOTATIONS AND THE FREE SPIN-1/2 FIELD EQUATION

The coordinates of the extended phase space V_8 stand for the space–time and momentum–energy and are denoted in various ways, namely $(\xi)\equiv(\xi^1,\xi^2,\xi^3,\xi^4,\xi^5,\xi^6,\xi^7,\xi^8) \equiv(q^1,q^2,q^3,q^4,p^1,p^2,p^3,p^4)\equiv(q,p)$. Capital italic indices take the values 1,2,...,8; lower italic the values 1,2,...,4; and Greek the values 1,2,3. The capital italic L,M are bispinor indices with values 1,2,...,16.

The summation convention is followed on repeated indices. The units are chosen to be $a = b = c = 1$, where a,b,c are, respectively, the fundamental length, momentum, velocity. All the physical quantities are expressed as pure numbers. The Minkowskian metric tensor is $[\eta_{ij}]\equiv[\eta_{q^iq^j}] \equiv\mathrm{diag}[-1^3,1]$ and the metric tensor of V_8 is $[\mathscr{N}_{AB}] \equiv\mathrm{diag}[-1^3,1,-1^3,1]$. The derivatives are denoted by $\partial_A\equiv\partial/\partial\xi^A$, $\partial_{q^i}\equiv\partial/\partial q^i$, $\partial_{p^i}\equiv\partial/\partial p^i$.

For the spin-1/2 field the relevant matrix algebra is the Clifford Algebra of eight generators α^A satisfying[8]

$$\alpha^A\alpha^B + \alpha^B\alpha^A = 2\mathscr{N}^{AB}I, \tag{2.1}$$

where I stands for the unit element. A convenient basis set for this 256-dimensional algebra is given by

$$e^{n_1n_2\cdots n_8}\equiv(\alpha^1)^{n_1}(\alpha^2)^{n_2}\dots(\alpha^8)^{n_8}, \tag{2.2}$$

where $n_A = 0,1$. It can be noticed that

$$(e^{n_1\cdots n_8})^2 = \pm I. \tag{2.3}$$

In the representation of this algebra one can conclude that

$$\mathrm{Tr}[e^{n_1\cdots n_8}] = 0. \tag{2.4}$$

where not all of the n_A's are zero. There exists a nonsingular, Hermitian matrix Λ such that

$$(\alpha^A)^\dagger = \Lambda\alpha^A\Lambda^{-1}, \tag{2.5}$$

where the dagger stands for the Hermitian conjugation. The only irreducible representation of this algebra is 16-dimensional.

The Lagrangian for the 16-component spin-1/2 field $\psi(\xi)$ is taken to be

$$L = (2i)^{-1}(\partial_A\tilde{\psi})\alpha^A\psi + (\mathrm{h.c.}) + m\tilde{\psi}\psi, \tag{2.6}$$

where α^A are given by (2.1), $\tilde{\psi}\equiv\psi^\dagger\Lambda$, (h.c.) stands for the Hermitian conjugation of the previous terms, and m is the mass parameter. The Euler–Lagrange field equations are

$$[\alpha^A\partial_A - imI]\,\psi(\xi) = 0. \tag{2.7}$$

The second-order equation that follows from (2.7) is

$$[\alpha^A\partial_A + m^2I]\,\psi(\xi) = 0. \tag{2.8}$$

The various matrix-Green's functions for (2.7) are given by

$$S_{(a)}(\xi-\xi') \equiv -i(2\pi)^{-8}\int_{-\infty}^{\infty} d^7\zeta \times\int_{C_{(a)}} d\zeta^4(\zeta^A\zeta_A - m^2)^{-1} \times(\alpha^A\zeta_A + mI)e^{i\zeta_A(\xi^A-\xi'^A)}, \tag{2.9}$$

where $C_{(a)}$ are the different contours in the complex ζ^4-plane shown in Fig. 1.

The two simple poles of the integrand in (2.9) are situated at the points

$$\zeta_4 = \pm\mathscr{E} \equiv\pm\sqrt{\zeta_1^2+\zeta_2^2+\zeta_3^2+\zeta_5^2+\zeta_6^2+\zeta_7^2-\zeta_8^2+m^2}. \tag{2.10}$$

The various tensor fields obeying differential conservation laws similar to $\partial_A T^A_B = 0$ can be constructed from (2.6). They are

$$\begin{aligned} T_{AB} &= (2i)^{-1}[(\partial_B\tilde{\psi})\alpha_A\psi] + (\mathrm{h.c.}),\\ \mathscr{M}^A_{BC} &= \xi_B T^A_C - \xi_C T^A_B,\\ \mathscr{S}^A_{BC} &= (i/2)\,\tilde{\psi}\alpha^A[\alpha_B\alpha_C - \alpha_C\alpha_B]\,\psi + (\mathrm{h.c.}),\\ \mathscr{J}^A_{BC} &= \mathscr{M}^A_{BC} + \mathscr{S}^A_{BC},\\ n^A &= \tilde{\psi}\alpha^A\psi. \end{aligned} \tag{2.11}$$

Some of the integral constants which can be constructed out of (2.11) are [cf. Eq. (I-5.18)]

$$\begin{aligned} K_\alpha &\equiv \int_{v_7} d^3q\,d^4p\,T^4_\alpha\\ &= (2i)^{-1}\,d^3q\,d^4p\,[(\partial_{q^\alpha}\tilde{\psi})\alpha^4\psi] + (\mathrm{h.c.}),\\ H\equiv K_4 &\equiv \int_{v_7} d^3q\,d^4p\,T^4_4 = (2i)^{-1}\\ &\quad\times\int_{v_7} d^3q\,d^4p\,[(\partial_{q^4}\tilde{\psi})\alpha^4\psi] + (\mathrm{h.c.}), \end{aligned}$$

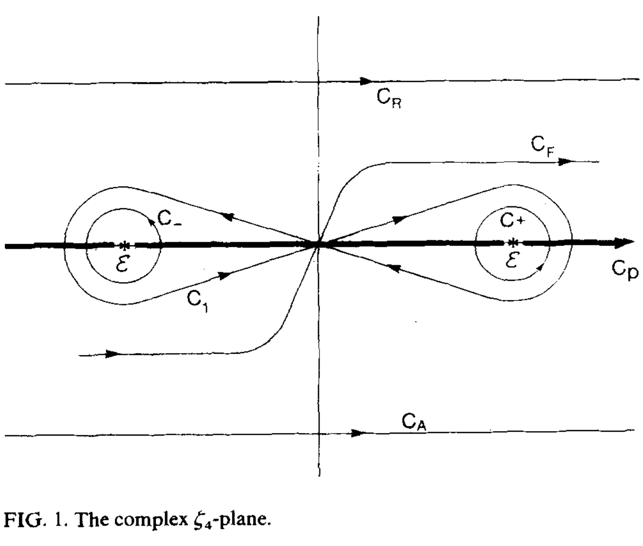

FIG. 1. The complex ζ_4-plane.

$$X_a \equiv \int_{v_7} d^3q\, d^4p\, T^4_{p^a} = (2i)^{-1} \times \int_{v_7} d^3q\, d^4p\, [(\partial_{p^a}\tilde{\psi})\alpha^4\psi] + (\text{h.c.}),$$

$$\sigma_{ab} \equiv \int_{v_7} d^3q\, d^4p\, \mathscr{S}^4_{ab} = \frac{i}{2}\int_{v_7} d^3q\, d^4p\, [\,\tilde{\psi}\alpha^4(\alpha_a\alpha_b - \alpha_b\alpha_a)\,\psi] + (\text{h.c.}),$$

$$T \equiv \int_{v_7} d^3q\, d^4p\, [q^j T^4_{p^j} - p^j T^4_{q^j}] = \frac{i}{2}\int_{v_7} d^3q\, d^4p\, [\,\tilde{\psi}\alpha^4(q^a\partial_{p^a} - p^a\partial_{q^a})\,\psi] + (\text{h.c.}),$$

$$B \equiv \frac{i}{4}\int_{v_7} d^3q\, d^4p\, [\,\tilde{\psi}\alpha^4(\alpha^a\alpha_{p^a} - \alpha^{p^a}\alpha_a)\,\psi] + (\text{h.c.}),$$

$$Q \equiv e_0\left[T + \frac{1}{2}(B)\right],$$

$$N \equiv \int_{v_7} d^3q\, d^4p\, n^4 = \int_{v_7} d^3q\, d^4p(\,\tilde{\psi}\alpha^4\psi),$$

where e_0 is a charge parameter, $\alpha_{p^a} \equiv \alpha_{a+4}$.

III. THE PLANE WAVE SOLUTIONS AND THE STATISTICAL INTERPRETATION

From the generators α^A in (2.1) the following can be defined:

$$v^b \equiv \alpha^b, \quad a^b \equiv \alpha^{b+4}, \tag{3.1}$$

$$v^b v^c + v^c v^b = 2\eta^{bc} I, \quad v^b a^c + a^c v^b = 0.$$

A possible 16×16 irreducible representation is furnished by the matrices

$$\begin{aligned}
&v^1 = i\sigma^1 \times I \times I \times I, \quad v^2 = i\sigma^3 \times \sigma^1 \times I \times I,\\
&v^3 = i\sigma^3 \times \sigma^3 \times \sigma^1 \times I, \quad v^4 = \sigma^3 \times \sigma^3 \times \sigma^3 \times \sigma^1,\\
&a^1 = i\sigma^2 \times I \times I \times I, \quad a^2 = i\sigma^3 \times \sigma^2 \times I \times I,\\
&a^3 = i\sigma^3 \times \sigma^3 \times \sigma^2 \times I, \quad a^4 = \sigma^3 \times \sigma^3 \times \sigma^3 \times \sigma^2,\\
&\Lambda = \sigma^3 \times \sigma^3 \times \sigma^3 \times I,
\end{aligned} \tag{3.2}$$

where the σ^{α}'s are the Pauli matrices, and I is the 2×2 unit matrix and the cross $\times$ denotes the Kronecker product. The wave equation (2.7) goes over to

$$[v^b\partial_{q^b} + a^b\partial_{p^b}]\,\psi(q,p) = im\,\psi(q,p). \tag{3.3}$$

The v^b-matrices have the significance of the 4-velocity, a^b are the components of the 4-acceleration, and $f(q,p) \equiv \tilde{\psi}\psi$ is analogous to the statistical distribution function. The right-hand side is the self-collision term. Equation (3.3) obviously bears a striking analogy to Boltzmann's transport equation for the statistical distribution function. From the history behind the matrix-wave equation (3.3) it will be called the Boltzmann–Dirac–Yukawa or in short BDY equation.

For the plane wave solutions of (3.3) the 16-component wave function ψ is taken as

$$\psi(q,p) = U(k,x)e^{-i(k_a q^a + x_a p^a)}. \tag{3.4}$$

Putting (3.4) into (3.3) one obtains the homogeneous linear algebraic system of equations

$$[v^b k_b + a^b x_b + mI\,]U(k,x) = 0. \tag{3.5}$$

Multiplying above from the left by $[v^b k_b + a^b x_b - mI\,]$ one can obtain

$$k_4 = \pm\mathscr{E}(k,x) \equiv \pm\sqrt{k_\alpha k_\alpha + x_\alpha x_\alpha - x_4^2 + m^2}, \tag{3.6}$$

for any nonzero solution $U(k,x)$. The necessary condition $\det[v^b k_b + a^b x_b - mI\,] = 0$ must be satisfied by (3.6).

There exist suitable basic solutions $U^R, V^S, (R,S = 1,2,...,8)$ of Eq. (3.5), which satisfy the normalization condition (see Appendix I)

$$\begin{aligned}
&\tilde{U}^{(R)}\tilde{U}^{(S)} = -\tilde{V}^{(R)}V^{(S)} = -(\sigma^3\times\sigma^3\times\sigma^3)_{(RS)},\\
&\tilde{U}^{(R)}V^{(S)} = \tilde{V}^{(R)}U^{(S)} = 0,\\
&\tilde{U}^{(R)}v^4U^{(S)} = \tilde{V}^{(R)}v^4V^{(S)} = (\mathscr{E}/m)(\sigma^3\times\sigma^3\times I)_{(RS)}.
\end{aligned} \tag{3.7}$$

The Fourier integral of the ψ-field satisfying the BDY Eq. (3.3) can be expressed now by virtue of (3.4), (3.7) as

$$\begin{aligned}
\psi(q,p) = (2\pi)^{-7/2}\int_{D_7} d^3k\, d^4x\, [m/\mathscr{E}(k,x)]^{1/2}
&\times \sum_{R=1}^{8} \big[\alpha_R(k,x)U^{(R)}(k,x)e^{-i(k_a q^a + x_a p^a)}\\
&+ \bar{\beta}_R(k,x)V^{(R)}(k,x)e^{i(k_a q^a + x_a p^a)}\big],
\end{aligned}$$

$$k_a q^a \equiv k_\alpha q^\alpha + \mathscr{E} q^4,$$

$$D_7 \equiv \{(k,x) \mid k_\alpha k_\alpha + x_\alpha x_\alpha - x_4^2 + m^2 > 0\}. \tag{3.8}$$

The domain D_7 has been chosen to avoid to avoid the imaginary values for $\mathscr{E}$; this choice would imply some restrictions upon the inversion of the Fourier integral.

Some of the integral constants (2.12) are evaluated using (3.8) and they are (see Appendix I)

$$\begin{aligned}
K_\alpha &= \int_{D_7} d^3k\, d^4x \sum_{R=1}^{8} [\epsilon_R f^-_R(k,x) - \epsilon_R f^+_R(k,x)]k_\alpha,\\
H &= \int_{D_7} d^3k\, d^4x \sum_{R=1}^{8} [\epsilon_R f^-_R(k,x) - \epsilon_R f^+_R(k,x)]\mathscr{E}(k,x),\\
X_a &= \int_{D_7} d^3k\, d^4x \sum_{R=1}^{8} [\epsilon_R f^-_R(k,x) - \epsilon_R f^+_R(k,x)]x_a,
\end{aligned} \tag{3.9}$$

$$N = \int_{D_7} d^3k\, d^4x \sum_{R=1}^{8} [\epsilon_R f^-_R(k,x) - \epsilon_R f^+_R(k,x)],$$

$$f^-_R \equiv |\alpha_R(k,x)|^2 \geqslant 0, \quad f^+_R \equiv |\beta_R(k,x)|^2 \geqslant 0, \quad \epsilon_R = \pm 1.$$

Some physical interpretations are now necessary. The nonnegative function $f^-_R(k,x)$ is the statistical distribution function in the dual phase space for some spin-1/2 particle ensemble and f^+_R stands for the antiparticle distribution function. The difficulty of indefiniteness of $\pm f^+_R$ is encountered also in the classical Dirac theory, and can be remedied in the second quantization. The integral constants, K_α, H, X_α, X_4, N can be interpreted as the total momentum, total energy, center of mass, time center, total number, respectively, of the particle–antiparticle ensemble.

IV. APPLICATION TO THE BARYON THEORY

For application of the BDY Eq. (3.3) to the baryon theory, polar coordinates will be introduced as

$$\rho^a \equiv \sqrt{(q^a)^2 + (p^a)^2}, \tag{4.1}$$

$$\theta^a \equiv \arctan(p^q/q^a) \quad \text{(no summation)},$$

where $0 < \rho^a$, $0 \leqslant \theta^a < 2\pi$.

In these coordinates the BDY equation goes over to[2]

$$\{\alpha^{b+} e^{-i\theta^b}[\partial_{\rho^b} - (i/\rho^b)\partial_{\theta^b}] + \alpha^{b-} e^{i\theta^b}[\partial_{\rho^b} + (i/\rho^b)\partial_{\theta^b}] - imI\}\, \psi(\rho,\theta) = 0,$$

$$\alpha^{b\,\pm} \equiv \tfrac{1}{2}(v^b \pm i a^b). \tag{4.2}$$

To solve (4.2) by the method of separation of variables each of the 16 components of the ψ-function is written in the form

$$\psi^L(\rho,\theta) = \chi^L(\rho) e^{it_b^L\theta^b} \quad (L \text{ not summed}), \tag{4.3}$$

where the bispinor indices L and M take the values 1,2,...,16. Summation is carried out on repeated indices, except when these are all subscripts or superscripts. Demanding that the ψ-function is either a single-valued or a double-valued function around the origin, the constants t_b^L must take either integer or half-interger values. Putting (4.3) into (4.2) and writing $(\alpha^{b\pm})_L^M$ for the entries of $\alpha^{b\pm}$-matrices, the 16 explicit partial differential equations can be exhibited as

$$(\alpha^{b\pm})_L^M[\partial_{\rho^b}\chi^L + (t_b^L/\rho^b)\chi^L]e^{i(t_c^L\theta^c - \theta^b)} + (\alpha^{b-})_L^M[\partial_{\rho^b}\chi^L - (t_b^L/\rho^b)\chi^L]e^{i(t_c^L\theta^c + \theta^b)} - im\chi^M e^{it_b^M\theta^b} = 0. \tag{4.4}$$

In order that the unimodular angular functions can be cancelled out from each of the Eqs. (4.4), the following 60 linear equations among 64 constants t_b^L must hold (see Appendix II):

$$t_1^1 = t_1^2 = t_1^3 = t_1^4 = t_1^5 = t_1^6 = t_1^7 = t_1^8 = t_1^9 - 1 = t_1^{10} - 1 = t_1^{11} - 1 = t_1^{12} - 1 = t_1^{13} - 1 = t_1^{14} - 1 = t_1^{15} - 1 = t_1^{16} - 1,$$

$$t_2^1 = t_2^2 = t_2^3 = t_2^4 = t_2^5 - 1 = t_2^6 - 1 = t_2^7 - 1 = t_2^8 - 1 = t_2^9 = t_2^{10} = t_2^{11} = t_2^{12} = t_2^{13} - 1 = t_2^{14} - 1 = t_2^{15} - 1 = t_2^{16} - 1,$$

$$t_3^1 = t_3^2 = t_3^3 = t_3^4 - 1 = t_3^5 = t_3^6 = t_3^7 - 1 = t_3^8 - 1 = t_3^9 = t_3^{10} = t_3^{11} - 1 = t_3^{12} - 1 = t_3^{13} = t_3^{14} = t_3^{15} - 1 = t_3^{16} - 1,$$

$$t_4^1 = t_4^2 - 1 = t_4^3 = t_4^4 - 1 = t_4^5 = t_4^6 - 1 = t_4^7 = t_4^8 - 1 = t_4^9 = t_4^{10} - 1 = t_4^{11} = t_4^{12} - 1 = t_4^{13} = t_4^{14} - 1 = t_4^{15} = t_4^{16} - 1. \tag{4.5}$$

It is clear that four of t_b^L's can be chosen arbitrarily and the remaining 60 can be obtained via (4.5). The following t_b^L are chosen freely

$$t_1 \equiv t_1^1, \quad t_2 \equiv t_2^2, \quad t_3 \equiv t_3^3 - 1, \quad t_4 \equiv t_4^4 - 1. \tag{4.6}$$

With the above choices for t_b^L Eq. (4.4) goes over to

$$[(\alpha^{b+})_L^M(\partial_{\rho^b} + t_b^L/\rho^b) + (\alpha^{b-})_L^M(\partial_{\rho^b} - t_b^L/\rho^b) - im\delta_L^M]\,\chi^L(\rho) = 0. \tag{4.7}$$

The nonsingular basic solutions of the above equation are expressible in terms of Bessel functions of the first kind of the appropriate orders. These are exhibited below:

$$\chi^L(\rho) = u^L(\kappa) J_{t_1^L}(\kappa_1\rho^1) J_{t_2^L}(\kappa_2\rho^2) J_{t_3^L}(\kappa_3\rho^3) J_{t_4^L}(\kappa_4\rho^4) \quad (L \text{ not summed}), \tag{4.8}$$

where t_b^L are given by (4.5), (4.6) and $u^L(\kappa)$ are the undetermined amplitude functions. To solve for these functions, (4.8) is substituted into (4.7). The Bessel functions can all be cancelled out (see Appendix II). Furthermore, using $a^b = -i(\alpha^{b+} - \alpha^{b-})$, one arrives at a simple matrix equation for the column vector u (with components u^L)

$$[a^b\kappa_b - mI]u(\kappa) = 0. \tag{4.9}$$

Multiplying above from the left by $[a^b\kappa_b - mI]$ one obtains the following necessary condition for nonzero solutions

$$\kappa_4 = \pm E(\kappa) \equiv \pm\sqrt{\kappa_\alpha\kappa_\alpha + m^2}. \tag{4.10}$$

From the commutation rules (3.1) it is evident that the given representations (3.2) of a^b-matrices can be reduced to the four direct sums of Dirac matrices by a similarity transformation (needless to say, the same similarity transformation would not decompose v^b-matrices). Such a similarity transformation, which is unitary as well, is explicitly shown below:

$$4S \equiv \begin{pmatrix} I+\sigma^2 & I-\sigma^2 & \sigma^3+i\sigma^1 & \sigma^3-i\sigma^1 & i(I-\sigma^2) & i(I+\sigma^2) & -(\sigma^1+i\sigma^3) & \sigma^1-i\sigma^3 \\ -(\sigma^3+i\sigma^1) & \sigma^3-i\sigma^1 & I+\sigma^2 & -(I-\sigma^2) & \sigma^1+i\sigma^3 & \sigma^1-i\sigma^3 & i(I-\sigma^2) & -i(I+\sigma^2) \\ \sigma^3-i\sigma^1 & \sigma^3+i\sigma^1 & -(I-\sigma^2) & -(I+\sigma^2) & \sigma^1-i\sigma^3 & -(\sigma^1+i\sigma^3) & -i(I+\sigma^2) & -i(I-\sigma^2) \\ -(I-\sigma^2) & I+\sigma^2 & -(\sigma^3-i\sigma^1) & \sigma^3+i\sigma^1 & -i(I+\sigma^2) & i(I-\sigma^2) & -(\sigma^1-i\sigma^3) & -(\sigma^1+i\sigma^3) \\ -i(I+\sigma^2) & i(I-\sigma^2) & -(\sigma^1-i\sigma^3) & -(\sigma^1+i\sigma^3) & -(I-\sigma^2) & I+\sigma^2 & -(\sigma^3-i\sigma^1) & \sigma^3+i\sigma^1 \\ \sigma^1-i\sigma^3 & -(\sigma^1+i\sigma^3) & -i(I+\sigma^2) & -i(I-\sigma^2) & \sigma^3-i\sigma^1 & \sigma^3+i\sigma^1 & -(I-\sigma^2) & -(I+\sigma^2) \\ -(\sigma^1+i\sigma^3) & -(\sigma^1-i\sigma^3) & -i(I-\sigma^2) & i(I+\sigma^2) & \sigma^3+i\sigma^1 & -(\sigma^3-i\sigma^1) & -(I+\sigma^2) & I-\sigma^2 \\ -i(I-\sigma^2) & -i(I+\sigma^2) & \sigma^1+i\sigma^3 & -(\sigma^1-i\sigma^3) & -(I+\sigma^2) & -(I-\sigma^2) & -(\sigma^3+i\sigma^1) & -(\sigma^3-i\sigma^1) \end{pmatrix}, \tag{4.11}$$

$$S^\dagger S = I, \quad Sa^bS^{-1} = I\times\gamma^b, \quad \gamma^\alpha \equiv i\sigma^2\times\sigma^\alpha,$$

$$\gamma^4 \equiv \sigma^3\times I, \quad SAv^4S^\dagger = -\sigma^3\times I\times\sigma^2\times I.$$

The same symbol I has been used for the unit matrices of different sizes and can be understood from the context. Using this similarity transformation, (4.9) can be brought to the form

$$[(I\times\gamma^b)\kappa_b - m(I\times I)]u'(\kappa) = 0, \tag{4.12}$$

$$u' \equiv Su.$$

TABLE I. Baryon wave amiplitudes.

u'															
$\kappa_4 = E$								$\kappa_4 = -E$							
$u'_{(11)}$	$u'_{(12)}$	$u'_{(21)}$	$u'_{(22)}$	$u'_{(31)}$	$u'_{(32)}$	$u'_{(41)}$	$u'_{(42)}$	$v'_{(11)}$	$v'_{(12)}$	$v'_{(21)}$	$v'_{(22)}$	$v'_{(31)}$	$v'_{(32)}$	$v'_{(41)}$	$v'_{(42)}$
e_1	e_2	0	0	0	0	0	0	e'_1	e'_2	0	0	0	0	0	0
0	0	e_1	e_2	0	0	0	0	0	0	e'_1	e'_2	0	0	0	0
0	0	0	0	e_1	e_2	0	0	0	0	0	0	e'_1	e'_2	0	0
0	0	0	0	0	0	e_1	e_2	0	0	0	0	0	0	e'_1	e'_2

With the knowledge of the plane wave solutions of the Dirac equation it is now easy to construct a set of 16 linearly independent solutions of (4.12). These are explicitly listed in Tables I and II.

The normalization conditions satisfied by the above solutions, derived by direct computation from (3.2), (4.11), (4.12), are the following:

$$u'^{\dagger}_{(a\beta)}u'_{(c\gamma)} = u'^{\dagger}_{(a\beta)}u'_{(c\gamma)} = (E/m)\delta_{(ac)}\delta_{(\beta\gamma)}\,,$$
$$u'^{\dagger}_{(a\beta)}u_{(c\gamma)} = u^{\dagger}_{(a\beta)}u_{(c\gamma)} = (E/m)\delta_{(ac)}\delta_{(\beta\gamma)}\,,$$
$$\tilde{u}_{(a\beta)}v^4u_{(b\gamma)} = \tilde{u}_{(a\beta)}v^4u_{(b\gamma)} = 0\,,$$
$$\tilde{u}_{(a\beta)}v^4u_{(b\gamma)} = -\tilde{u}_{(b\gamma)}v^4u_{(\alpha\beta)} = i(\gamma^4)_{(ab)}\delta_{(\beta\gamma)}\,.$$

The general solution of (3.3) in polar coordinates can be written as

$$\psi^L(\rho,\theta) = (2\pi)^{-3/2}m^{1/2}\sum_{r=1}^{4}\sum_{\alpha=1}^{2}\sum_{t=-\infty}^{\infty}$$
$$\times\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3[\beta_{(r,\alpha)}(\underset{\sim}{\kappa},t)u^{(r\alpha)L} + \bar{\gamma}_{(r\alpha)}(\underset{\sim}{\kappa},t)v^{(r\alpha)L}]$$
$$\times J_{t^L_1}(\kappa_1\rho^1)J_{t^L_2}(\kappa_2\rho^2)J_{t^L_3}(\kappa_3\rho^3)J_{t^L_4}(E\rho^4)e^{it^L_b\theta^b}$$
$$(L \text{ not summed}), \qquad (4.14)$$

where $\beta_{(r\alpha)}$, $\bar{\gamma}_{(r\alpha)}$ are arbitrary 16 complex amplitudes and t^L_b satisfy (4.5). It will be convenient to subject $\beta_{(r\alpha)}$, $\bar{\gamma}_{(r\alpha)}$ to two invertible complex linear transformations (Appendix III, Eqs. AIII8, AIII36) to obtain other complex amplitudes $\alpha_{(R\alpha)}$ ($R = 1,...,8$; $\alpha = 1,2$) and introduce the distribution functions

$$f_{(R\alpha)}(\underset{\sim}{\kappa},t) \equiv (2/\pi)|\alpha_{(R\alpha)}(\underset{\sim}{\kappa},t)|^2 \geqslant 0\,,$$
$$N^{(t)}_{(R\alpha)} \equiv \int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3 f_{(R\alpha)}(k,t)\,. \qquad (4.15)$$

Then choosing a single t-mode in (4.14) the corresponding integral constants (2.12) turn out to be (see Appendix III)

$$K^{(t)}_\alpha = 0\,,$$
$$H^{(t)} = \sum_{R=1}^{8}\sum_{\alpha=1}^{2}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3[\epsilon_R f_{(R\alpha)}(\underset{\sim}{\kappa},t)]E(\underset{\sim}{\kappa})\,,$$
$$X^{(t)}_a = 0\,,$$
$$T^{(t)} = -\frac{\pi}{2}(t_1 + t_2 + t_3 + t_4 + \tfrac{1}{2})\sum_{R=1}^{8}\sum_{\alpha=1}^{2}\epsilon_R N^{(t)}_{(R\alpha)}\,,$$
$$B^{(t)} = \frac{\pi}{2}\sum_{\alpha=1}^{2}[N^{(t)}_{(1\alpha)} + N^{(t)}_{(2\alpha)} - N^{(t)}_{(3\alpha)} + N^{(t)}_{(4\alpha)} - N^{(t)}_{(5\alpha)}$$
$$- N^{(t)}_{(6\alpha)} - 3N^{(t)}_{(7\alpha)} + 3N^{(t)}_{(8\alpha)}]\,,$$
$$Q^{(t)} = -\frac{\pi}{2}e_0\sum_{\alpha=1}^{2}\{(t_1 + t_2 + t_3 + t_4 + 1)$$
$$\times[N^{(t)}_{(1\alpha)} + N^{(t)}_{(2\alpha)} - N^{(t)}_{(3\alpha)}]$$
$$+ (t_1 + t_2 + t_3 + t_4)[-N^{(t)}_{(4\alpha)} + N^{(t)}_{(5\alpha)} + N^{(t)}_{(6\alpha)}]$$
$$- (t_1 + t_2 + t_3 + t_4 + 2)N_{(7\alpha)}$$
$$- (t_1 + t_2 + t_3 + t_4 - 1)N^{(t)}_{(8\alpha)}\}\,, \qquad (4.16)$$

where $\epsilon_1 = \epsilon_2 = -\epsilon_3 = -\epsilon_4 = \epsilon_5 = \epsilon_6 = -\epsilon_7 = -\epsilon_8 = 1$.

In the above $f_{(R\alpha)}(\kappa,t)$ is the distribution function in momentum–space for the eight spin-1/2 basic particles at (t)-mode. $K^{(t)}_\alpha = X^{(t)}_\alpha = 0$ can be interpreted physically by recalling that in the basis of Bessel functions, incoming and outgoing waves are equally mixed. The charge constant $Q^{(t)}$ has a multiplier $(\pi/2)$ and factors $(t_1 + t_2 + t_3 + t_4 + b)$, $b = 0, \pm 1,2$. A possible antiparticle scheme is to consider the negative $(-t)$-mode $\psi^{(-t)}$. The various reflection properties of the $\psi^{(-t)}$-field (4.14) are the following

$$\pi_\alpha\psi^{(t)L}(\rho,\theta) \equiv \psi^{(t)L}(\rho,...,\theta^\alpha + \pi,\cdots) = \xi^L_\alpha\psi^{(t)}(\rho,\theta)$$
$$(L \text{ not summed}),$$
$$\xi^L_\alpha \equiv \epsilon^{it^L_\alpha\pi}\,,$$
$$\xi^L_P \equiv \xi^L_1\xi^L_2\xi^L_3 = e^{i(t^L_1 + t^L_2 + t^L_3)\pi}\,,$$
$$T'\psi^{(t)L}(\rho,\theta) \equiv \psi^{(t)L}(\rho,...,\theta^4 + \pi) = \xi^L_{T'}\psi^{(t)L}(\rho,\theta)\,,$$
$$\xi^L_{T'} \equiv e^{it^L_4\pi}\,, \qquad (4.17)$$

TABLE II. $v \equiv [(E + m)/2m]^{1/2}$. Dirac wave amplitudes.

e_1/v	e_2/v	e'_1/v	e'_2/v
1	0	$\kappa_3(E + m)^{-1}$	$(\kappa_1 - i\kappa_2)(E + m)^{-1}$
0	1	$(\kappa_1 + i\kappa_2)(E + m)^{-1}$	$-\kappa_3(E + m)^{-1}$
$-\kappa_3(E + m)^{-1}$	$-(\kappa_1 - i\kappa_3)(E + m)^{-1}$	1	0
$-(\kappa_1 + i\kappa_2)(E + m)^{-1}$	$\kappa_3(E + m)^{-1}$	0	1

where $t_a^L = t_a, t_a + 1$.

If one defines the (mass)^2-operator by the following:

$$M^2(\rho,\partial_\rho) \equiv -\eta^{ij}[\partial_{\rho^i}\partial_{\rho^j} + (1/\rho_i)\partial_{\rho^j}],$$

then from (4.4) one can derive a second-order equation

$$M^2(\rho,\partial_\rho)\chi = M^2_{(t)}(\rho)\chi,$$
$$M^2_{(t)}(\rho) \equiv m^2 I - \mathrm{diag}[\eta^{ij} t_i^L t_j^L/\rho^i\rho^j] \quad L = 1,\ldots,16. \tag{4.18}$$

For the particular (t)-mode $t_1 = t_2 = 0$ denoting $s \equiv 2t_4$, $e \equiv -(\pi/2)e_0$, one can obtain from (4.16), (4.18) the formulas:

$$H^{(t_3,s)} = \sum_{r=1}^{8}\sum_{\alpha=1}^{2}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3[\epsilon_R f_{(R\alpha)}(\kappa,t_3,s)]E(\kappa),$$
$$Q^{(t_3,s)} = \pm e\sum_{R=1}^{8}\sum_{\alpha=1}^{2}\left(t_3 + \frac{s}{2} + b\right)N_{(R\alpha)}(\kappa,t_3,s),$$
$$(mass)^2 \equiv m^2 + \phi_1(\rho^1,\rho^2) + (t_3 + c_1)^2/(\rho^3)^2 - (s + c_2)^2/4(\rho^4)^2, \tag{4.19}$$

where $b = 0, \pm 1,2$; $c_1 = 1,0$; $c_2 = 1,0$; $\phi_1 \equiv 0, 1/(\rho^1)^2$, $1/(\rho^2)^2$, $1/(\rho^1)^2 + 1/(\rho^2)^2$. Equations (4.19) bear analogy with Gell-Mann–Nishijima,[3] Okubo[6] formulas for the charge and mass of the particles if t_3, s are identified with isospin quantum number and the strangeness, respectively.

V. COMMENTS ON THE LEPTON FIELDS

For the lepton fields the 16-component BDY Eq. (3.3)

$$[\alpha^A\partial_A - imI]\psi(\xi) = 0 \tag{5.1}$$

is again chosen, but with a different mass value. Instead of the representations (3.2) the following equivalent irreducible representation of α_A-algebra is taken.

$$\alpha^a = -\gamma^a \times I, \quad \alpha^{a+4} = -\gamma^5\cdot\gamma^a,$$
$$\gamma^5 \equiv i\gamma^1\gamma^2\gamma^3\gamma^4, \quad A = \gamma^4\times(\sigma^2\times I),$$
$$= i\gamma^4\times(\gamma^4\gamma^5), \tag{5.2}$$

where γ^a's are Dirac matrices in (4.11).

Special solutions of (5.1) in the form

$$\psi(q,p) = \psi_{(1)}(q)\times\psi_{(2)}(p), \tag{5.3}$$

where $\psi_{(1)}(q)$, $\psi_{(2)}(p)$ are 4-component column vectors, will be investigated. Putting (5.3), (5.2) into (5.1) one obtains

$$\{[-i\gamma^a\partial_{q^a}\psi_{(1)}(q)]\times\psi_{(2)}(p)\}$$
$$+\{[\gamma^5{}_{(1)}(q)][-i\gamma^a\partial_{p^a}\psi_{(2)}(p)]\}$$
$$+[m\psi_{(1)}(q)\times\psi_{(2)(p)}] = 0. \tag{5.4}$$

By separating variables one can get

$$[-i\gamma^a\partial_{q^a} + mI - \lambda\gamma^5]\psi_{(1)}(q) = 0,$$
$$[-i\gamma^a\partial_{p^a} + \lambda I]\psi_{(2)}(p) = 0, \tag{5.5}$$

where λ is the constant of separation. The second-order equations derivable from (5.5) are, resectively,

$$[\eta^{ab}\partial_{q^a}\partial_{q^b} + (m^2 - \lambda^2)I]\psi_{(1)}(q) = 0,$$
$$[\eta^{ab}\partial_{p^a}\partial_{p^b} + \lambda^2 I]\psi_{(2)}(p) = 0. \tag{5.6}$$

For the case $\lambda = 0$ one obtains from (5.5)

$$[-i\gamma^a\partial_{q^a} + mI]\psi_{(e)}(q) = 0,$$
$$-i\gamma^a\partial_{p^a}\psi_{(\nu)}(p) = 0, \tag{5.7}$$

where $\psi_{(e)}(q) \equiv \psi_{(1)}(q)$, $\psi_{(\nu)}(p) \equiv \psi_{(2)}(p)$. Equation (5.7) can be identified with the Dirac equation for the electron-positron field $\psi_{(e)}$ and the neutrino–antineutrino field $\psi_{(\nu)}$.

The plane wave solutions of the Dirac equation satisfying the normalization conditions

$$e^\dagger_{(\alpha)}e_{(\beta)} = e'^\dagger_{(\alpha)}e'_{(\beta)} = 2E(\underline{k})\delta_{(\alpha\beta)},$$
$$e^\dagger_{(\alpha)}e'_{(\beta)} = e'^\dagger_{(\alpha)}e_{(\beta)} = 0, \tag{5.8}$$

are given in Table II, where a change of normalization from v to $v' \equiv (E+m)^{1/2}$ is still necessary. The plane wave solutions of the neutrino equation are furnished by (see Table III)

$$E_0 \equiv \sqrt{x_\alpha x_\alpha},$$
$$\epsilon^\dagger_{(\alpha)}\epsilon_{(\beta)} = \epsilon'^\dagger_{(\alpha)}\epsilon'_{(\beta)} = 2E_0\delta_{(\alpha\beta)},$$
$$\epsilon^\dagger_{(\alpha)}\epsilon'_{(\beta)} = \epsilon'^\dagger_{(\alpha)}\epsilon_{(\beta)} = 0. \tag{5.9}$$

With the knowledge of these plane wave solutions one can write now the special class (5.3) of the solutions of (5.1) as

$$\psi(q,p) = \psi_{(e)}(q)\times\psi^{(\nu)}(p)$$
$$= 2^{-1}(2\pi)^{-3}\int_{-\infty}^{\infty} d^3\underline{k}\,d^3\underline{x}(EE_0)^{-1/2}$$
$$\times\sum_{\alpha,\beta=1}^{2}\{b_{(\alpha)}(\underline{k})\beta_{(\beta)}(\underline{x})[e_{(\alpha)}\times\epsilon_{(\beta)}]$$
$$\times\exp[i(k_a q^a + x_a p^a)]$$
$$+\bar{d}_{(\alpha)}(\underline{k})\bar{\delta}_{(\beta)}(\underline{x})[e'_{(\alpha)}\times\epsilon'_{(\beta)}]$$
$$\times\exp[-i(k_a q^a + x_a p^a)]$$
$$+b_{(\alpha)}(\underline{k})\bar{\delta}_{(\beta)}(\underline{x})[e_{(\alpha)}\times\epsilon_{(\beta)}]$$
$$\times\exp[i(k_a q^a - x_a p^a)]$$
$$+\bar{d}_{(\alpha)}(\underline{k})\beta_{(\beta)}(\underline{x})\exp[i(-k_a q^a + x_a p^a)]\}, \tag{5.10}$$

$$k_4 \equiv E(\underline{k}), \quad x_4 \equiv E_0(\underline{x}).$$

TABLE III. Neutrino wave amplitudes.

$\epsilon_{(1)}E_0^{-1/2}$	$\epsilon_{(2)}E_0^{-1/2}$	$\epsilon'_{(1)}E_0^{-1/2}$	$\epsilon'_{(2)}E_0^{-1/2}$
1	0	$x_3E_0^{-1}$	$(x_1 - ix_2)E_0^{-1}$
0	1	$(x_1 + ix_2)E_0^{-1}$	$-x_3E_0^{-1}$
$-x_3E_0^{-1}$	$-(x_1 - ix_2)E_0^{-1}$	1	0
$-(x_1 + ix_2)E_0^{-1}$	$x_3E_0^{-1}$	0	1

VI. ACKNOWLEDGMENTS

The author wishes to thank his wife Mrs. Purabi Das, for checking out the matrix multiplications. He also thanks Mr. A. Vaze of the Computing Centre for figuring out the eigenvalues mentioned in the Appendix III. Furthermore, he acknowledges the support of NRC Grant No. 3993. Finally, he thanks Dr. M.J. Hamilton and Dr. S. Kloster for informal discussions on papers I, II, and III in this series.

APPENDIX I

To construct a suitable basis set for the solution space or the null space of the matrix $[v^b k_b + a^b x_b + mI]$ in (3.5), this equation is expressed by (3.2) in a more elaborate fashion, namely,

$$\begin{aligned}
&(\sigma^1 k_4 + \sigma^2 x_4 + mI)u_1 + (ik_3 + x_3)u_2 + (ik_2 + x_2)u_3 \\
&\quad + (ik_1 + x_1)u_5 = 0, \\
&(ik_3 - x_3)u_1 + (-\sigma^1 k_4 - \sigma^2 x_4 + mI)u_2 + (ik_2 + x_2)u_4 \\
&\quad + (ik_1 + x_1)u_6 = 0, \\
&(ik_2 - x_2)u_1 + (-\sigma^1 k_4 - \sigma^2 x_4 + mI)u_3 - (ik_3 + x_3)u_4 \\
&\quad + (ik_1 + x_1)u_7 = 0, \\
&(ik_2 - x_2)u_2 + (-ik_3 + x_3)u_3 + (\sigma^1 k_4 + \sigma^2 x_4 + mI)u_4 \\
&\quad + (ik_1 + x_1)u_8 = 0, \\
&(ik_1 - x_1)u_1 + (-\sigma^1 k_4 - \sigma^2 x_4 + mI)u_5 - (ik_3 + x_3)u_6 \\
&\quad - (ik_2 + x_2)u_7 = 0. \\
&(ik_1 - x_1)u_2 + (-ik_3 + x_3)u_5 + (\sigma^1 k_4 + \sigma^2 x_4 + mI)u_6 \\
&\quad - (ik_2 + x_2)u_8 = 0, \\
&(ik_1 - x_1)u_3 + (-ik_2 + x_2)u_5 + (\sigma^1 k_4 + \sigma^2 x_4 + mI)u_7 \\
&\quad + (ik_3 + x_3)u_8 = 0, \\
&(ik_1 - x_1)u_4 + (-ik_2 + x_2)u_6 + (ik_3 - x_3)u_7 \\
&\quad + (-\sigma^1 k_4 - \sigma^2 x_4 + mI)u_8 = 0, \qquad \text{(AI1)}
\end{aligned}$$

where $u_1, u_2, \ldots, u_8$ are 2-component fields which add up to the 16-component U. The first four and the last four equations in (AI1) are linearly dependent and the rank of $[v^b k_b + a^b x_b + mI]$ is eight. Therefore any four of $u_1, u_2, \ldots, u_8$ can be chosen arbitrarily and the other four can be solved from (AI1). With this understanding a set of 16 basic solutions has been arranged in Table IV.

The solutions in Table IV satisfy the normalization conditions (3.7).

TABLE IV. Plane wave amplitudes of BDY equation.

$\kappa_4 = \mathscr{E}$ U	$U^{(1)}, U^{(2)}$	$U^{(3)}, U^{(4)}$	$U^{(5)}, U^{(6)}$	$U^{(7)}, U^{(8)}$
u_1	$[(k_1^2 + x_1^2)/2m^2]^{1/2} \times \binom{1}{0}, \binom{0}{1}$	0	0	0
u_2	0	$[(k_1^2 + x_1^2)/2m^2]^{1/2} \times \binom{1}{0}, \binom{0}{1}$	0	0
u_3	0	0	$[(k_1^2 + x_1^2)/2m^2]^{1/2} \times \binom{1}{0}, \binom{0}{1}$	0
u_4	0	0	0	$[(k_1^2 + x_1^2)/2m^2]^{1/2} \times \binom{1}{0}, \binom{0}{1}$
u_5	$-(ik_1 + x_1)^{-1}[\sigma^1 \mathscr{E} + \sigma^2 x_4 + mI]u_1$	$-(ik_3 + x_3)(ik_1 + x_1)^{-1}u_2$	$-(ik_2 + x_2)(ik_1 + x_1)^{-1}u_3$	0
u_6	$-(ik_3 - x_3)(ik_1 + x_1)^{-1}u_1$	$(ik_1 + x_1)^{-1}[\sigma^1 \mathscr{E} + \sigma^2 x_4 - mI]u_2$	0	$-(ik_2 + x_2)(ik_1 + x_1)^{-1}u_4$
u_7	$-(ik_2 - x_2)(ik_1 + x_1)^{-1}u_1$	0	$(ik_1 + x_1)^{-1}[\sigma^1 \mathscr{E} + \sigma^2 x_4 - mI]u_3$	$(ik_3 + x_3)(ik_1 + x_1)^{-1}u_4$
u_8	0	$-(ik_2 - x_2)(ik_1 + x_1)^{-1}u_2$	$(ik_3 - x_3)(ik_1 + x_1)^{-1}u_3$	$-(ik_1 + x_1)^{-1}[\sigma^1 \mathscr{E} + \sigma^2 x_4 + mI]u_4$

U	$V^{(1)}, V^{(2)}$	$V^{(3)}, V^{(4)}$	$\kappa_4 = -\mathscr{E}$ $V^{(5)}, V^{(6)}$	$V^{(7)}, V^{(8)}$
u_1	$(ik_1 - x_1)^{-1}[-\sigma^1 \mathscr{E} + \sigma^2 x_4 - mI]u_5$	$(ik_3 + x_3)(ik_1 - x_1)^{-1}u_6$	$(ik_2 + x_2)(ik_1 - x_1)^{-1}u_7$	0
u_2	$(ik_3 - x_3)(ik_1 - x_1)^{-1}u_5$	$-(ik_1 - x_1)^{-1}[-\sigma^1 \mathscr{E} + \sigma^2 x_4 + mI]u_6$	0	$(ik_2 + x_2)(ik_1 - x_1)^{-1}u_8$
u_3	$(ik_2 - x_2)(ik_1 - x_1)^{-1}u_5$	0	$-(ik_1 - x_1)^{-1}[-\sigma^1 \mathscr{E} + \sigma^2 x_4 + mI]u_7$	$-(ik_3 + x_3)(ik_1 - x_1)^{-1}u_8$
u_4	0	$(ik_2 - x_2)(ik_1 - x_1)^{-1}u_6$	$-(ik_3 - x_3)(ik_1 - x_1)^{-1}u_7$	$(ik_1 - x_1)^{-1}[-\sigma^1 \mathscr{E} + \sigma^2 x_4 - mI]u_8$
u_5	$[(k_1^2 + x_1^2)/2m^2]^{1/2} \times \binom{1}{0}, \binom{0}{1}$	0	0	0
u_6	0	$[(k_1^2 + x_1^2)/2m^2]^{1/2} \times \binom{1}{0}, \binom{0}{1}$	0	0
u_7	0	0	$[(k_1^2 + x_1^2)/2m^2]^{1/2} \times \binom{1}{0}, \binom{0}{1}$	0
u_8	0	0	0	$[(k_1^2 + x_1^2)/2m^2]^{1/2} \times \binom{1}{0}, \binom{0}{1}$

The computation of

$$K_j = \int_{D_7} d^3k\, d^4x \sum_{R=1}^{8} [\epsilon_R |\alpha_R(\underline{k},x)|^2 - \epsilon_R |\beta_R(\underline{k},x)|^2] k_j, \tag{AI2}$$

where

$\epsilon_1 = \epsilon_2 = \epsilon_7 = \epsilon_8 = -\epsilon_3 = -\epsilon_4 = -\epsilon_5 = -\epsilon_6 = 1$, will be shown here.

From (3.8) it follows that

$$[\partial_{q^4}\tilde{\psi}]_{q^4=0} = (2\pi)^{-7/2} \int_{D_7} d^3k\, d^4x (m/\mathscr{E})^{1/2}(ik_j) \times \sum_{R=1}^{8} \{\overline{\alpha}_R \tilde{U}^{(R)} \times \exp[i(k_\mu q^\mu + x_a p^a)] - \beta_R \tilde{V}^{(R)} \times \exp[-i(k_\mu q^\mu + x_a p^a)]\}. \tag{AI3}$$

From (2.12), (3.7), (3.8), (AI3) and the usual representation of the δ-function one obtains

$$\begin{aligned}
K_j = [K_j]_{q^4=0} &= (2\pi)^{-7} \int_{V_7} d^3q\, d^4p \int_{D_7} d^3k\, d^4x (m/\mathscr{E})^{1/2}(ik_j) \\
&\times \int_{D_7} d^3k'\, d^4x' (m/\mathscr{E}')^{1/2} \sum_{R=1}^{8} \sum_{R=1}^{\infty} \{\overline{\alpha}_R \tilde{U}^{(R)} \exp[i(k_\mu q^\mu + x_a p^a)] - \beta_R \tilde{V}^{(R)} \exp[-i(k_\mu q^\mu + x_a p^a)]\} v^4 \\
&\times \{\alpha'_S U'^{(S)} \exp[-i(k'_\nu q^\nu + x'_b p^b)] + \overline{\beta}'_S V'^{(S)} \exp[i(k'_\nu q^\nu + x'_b p^b)]\} + (\text{h.c.}) \\
&= \tfrac{1}{2} \int_{D_7} d^3k\, d^4x (m/\mathscr{E})^{1/2} k_j \int_{D_7} d^3k'\, d^4x' (m/\mathscr{E}')^{1/2} \sum_{R=1}^{8} \sum_{S=1}^{8} \{[\overline{\alpha}_R \alpha'_S \tilde{U}^{(R)} v^4 U'^{(S)} - \overline{\beta}'_S \beta_R \tilde{V}^{(R)} v^4 V'^{(S)}] \\
&\times \delta^3(\underline{k} - \underline{k}')\delta^4(x - x') + [\overline{\alpha}_R \overline{\beta}'_S \tilde{U}^{(R)} v^4 V'^{(S)} - \alpha'_S \beta_R \tilde{V}^{(R)} v^4 U'^{(S)}] \delta^3(\underline{k} + \underline{k}')\delta^4(x + x')\} + (\text{h.c.}) \\
&= (m/2\mathscr{E}) \int_{D_7} d^3k\, d^4x\, k_j \sum_{R=1}^{8} \sum_{S=1}^{8} [\overline{\alpha}_R \alpha_S U^{(R)} v^4 U^{(S)} - \overline{\beta}_S \beta_R \tilde{V}^{(R)} v^4 V^{(S)}] + (\text{h.c.}) + (m/2\mathscr{E}) \int_{D_7} d^3k\, d^4x\, k_j \\
&\times \sum_{R=1}^{8} \sum_{S=1}^{8} [\overline{\alpha}_R(\underline{k},x) \overline{\beta}_S(-\underline{k},-x) \tilde{U}^{(R)}(\underline{k},x) v^4 V^{(S)}(-\underline{k},-x) \\
&- \overline{\alpha}_R(-\underline{k},-x) \overline{\beta}_S(\underline{k},x) \tilde{U}^{(R)}(-\underline{k},x) v^4 V^{(S)}(\underline{k},x) - \alpha_S(-\underline{k},-x) \beta_R(\underline{k},x) \tilde{V}_{(R)}(\underline{k},x) v^4 U^{(S)}(-\underline{k},-x) \\
&+ \alpha_S(\underline{k},x) \beta_R(-\underline{k},-x) \tilde{V}^{(R)}(-\underline{k},-x) v^4 U^{(S)}(\underline{k},x)].
\end{aligned}$$

The last integral in the above equation vanishes because whenever $(\underline{k},x) \in D_7$ also the point $(-\underline{k},-x) \in D_7$ and thus using (3.7) Eq. (AI2) is obtained.

APPENDIX II

From the matrix representations (3.2) and (4.2), Eq. (4.4) yield the following partial differential equation:

$$\begin{aligned}
&[\partial_{\rho^4}\chi^2 + (t^2_4/\rho^4)\chi^2] \exp[i(t^2_b\theta^b - \theta^4)] + i[\partial_{\rho^3}\chi^3 + (t^3_3/\rho^3)\chi^3] \exp[i(t^3_b\theta^b - \theta^3)] \\
&\quad + i[\partial_{\rho^2}\chi^5 + (t^5_2/\rho^2)\chi^5] \exp[i(t^5_b\theta^b - \theta^2)] + i[\partial_{\rho^1}\chi^9 + (t^9_1/\rho^1)\chi^9] \exp[i(t^9_b\theta^b - \theta^1)] - im\chi^1 \exp[i(t^1_b\theta^b)] = 0, \\
&[\partial_{\rho^4}\chi^1 - (t^1_4/\rho^4)\chi^1] \exp[i(t^1_b\theta^b + \theta^4)] + i[\partial_{\rho^3}\chi^4 + (t^4_3/\rho^3)\chi^4] \exp[i(t^4_b\theta^b - \theta^3)] \\
&\quad + i[\partial_{\rho^2}\chi^6 + (t^6_2/\rho^2)\chi^6] \exp[i(t^6_b\theta^b - \theta^2)] + i[\partial_{\rho^1}\chi^{10} + (t^{10}_1/\rho^1)\chi^1] \exp[i(t^{10}_b\theta^b - \theta^1)] - im\chi^2 \exp[i(t^2_b\theta^b)] = 0, \\
&i[\partial_{\rho^3}\chi^1 - (t^1_3/\rho^3)\chi^1] \exp[i(t^1_b\theta^b + \theta^3)] - [\partial_{\rho^4}\chi^4 + (t^4_4/\rho^4)\chi^4] \exp[i(t^4_b\theta^b - \theta^4)] \\
&\quad + i[\partial_{\rho^2}\chi^7 + (t^7_2/\rho^2)\chi^7] \exp[i(t^7_b\theta^b - \theta^2)] + i[\partial_{\rho^1}\chi^{11} + (t^{11}_1/\rho^1)\chi^{11}] \exp[i(t^{11}_b\theta^b - \theta^1)] - im\chi^3 \exp[i(t^3_b\theta^b)] = 0, \\
&i[\partial_{\rho^3}\chi^2 - (t^2_3/\rho^3)\chi^2] \exp[i(t^2_b\theta^b + \theta^3)] - [\partial_{\rho^4}\chi^3 - (t^3_4/\rho^4)\chi^3] \exp[i(t^3_b\theta^b + \theta^4)] \\
&\quad + i[\partial_{\rho^2}\chi^8 + (t^8_2/\rho^2)\chi^8] \exp[i(t^8_b\theta^b - \theta^2)] + i[\partial_{\rho^1}\chi^{12} + (t^{12}_1/\rho^1)\chi^{12}] \exp[i(t^{12}_b\theta^b - \theta^1)] - im\chi^4 \exp[i(t^4_b\theta^b)] = 0, \\
&i[\partial_{\rho^2}\chi^1 - (t^1_2/\rho^2)\chi^1] \exp[i(t^1_b\theta^b + \theta^2)] - [\partial_{\rho^4}\chi^6 + (t^6_4/\rho^4)\chi^6] \exp[i(t^6_b\theta^b - \theta^4)] - i[\partial_{\rho^3}\chi^7 + (t^7_3/\rho^3)\chi^7] \\
&\quad \times \exp[i(t^7_b\theta^b - \theta^3)] + i[\partial_{\rho^1}\chi^{13} + (t^{13}_1/\rho^1)\chi^{13}] \exp[i(t^{13}_b\theta^b - \theta^1)] - im\chi^5 \exp[i(t^5_b\theta^b)] = 0, \\
&i[\partial_{\rho^2}\chi^2 - (t^2_2/\rho^2)\chi^2] \exp[i(t^2_b\theta^b + \theta^2)] - [\partial_{\rho^4}\chi^5 - (t^5_4/\rho^4)\chi^5] \exp[i(t^5_b\theta^b + \theta^4)] - i[\partial_{\rho^3}\chi^8 + (t^8_3/\rho^4)\chi^8] \\
&\quad \times \exp[i(t^8_b\theta^b - \theta^3)] + i[\partial_{\rho^1}\chi^{14} + (t^{14}_1/\rho^1)\chi^{14}] \exp[i(t^{14}_b\theta^b - \theta^1)] - im\chi^6 \exp[i(t^6_b\theta^b)] = 0, \\
&i[\partial_{\rho^2}\chi^3 - (t^3_2/\rho^2)\chi^3] \exp[i(t^3_b\theta^b + \theta^2)] - i[\partial_{\rho^3}\chi^5 - (t^5_3/\rho^3)\chi^5] \exp[i(t^5_b\theta^b + \theta^3)] + [\partial_{\rho^4}\chi^8 + (t^8_4/\rho^4)\chi^8] \\
&\quad \times \exp[i(t^8_b\theta^b - \theta^4)] + i[\partial_{\rho^1}\chi^{15} + (t^{15}_1/\rho^1)\chi^{15}] \exp[i(t^{15}_b\theta^b - \theta^1)] - im\chi^7 \exp[i(t^7_b\theta^b)] = 0, \\
&i[\partial_{\rho^2}\chi^4 - (t^4_2/\rho^2)\chi^4] \exp[i(t^4_b\theta^b + \theta^2)] - i[\partial_{\rho^3}\chi^6 - (t^6_3/\rho^3)\chi^6] \exp[i(t^6_b\theta^b + \theta^3)] + [\partial_{\rho^4}\chi^7 - (t^7_4/\rho^4)\chi^7] \\
&\quad \times \exp[i(t^7_b\theta^b + \theta^4)] + i[\partial_{\rho^1}\chi^{16} + (t^{16}_1/\rho^1)\chi^{16}] \exp[i(t^{16}_b\theta^b - \theta^1)] - im\chi^8 \exp[i(t^8_b\theta^b)] = 0, \\
&i[\partial_{\rho^1}\chi^1 - (t^1_1/\rho^1)\chi^1] \exp[i(t^1_b\theta^b + \theta^1)] - [\partial_{\rho^4}\chi^{10} + (t^{10}_4/\rho^4)\chi^{10}] \exp[i(t^{10}_b\theta^b - \theta^4)] - i[\partial_{\rho^3}\chi^{11} + (t^{11}_3/\rho^3)\chi^{11}] \\
&\quad \times \exp[i(t^{11}_b\theta^b - \theta^3)] - i[\partial_{\rho^2}\chi^{13} + (t^{13}_2/\rho^2)\chi^{13}] \exp[i(t^{13}_b\theta^b - \theta^2)] - im\chi^9 \exp[i(t^9_b\theta^b)] = 0, \\
&-i[\partial_{\rho^1}\chi^2 - (t^2_1/\rho^1)\chi^2] \exp[i(t^2_b\theta^b + \theta^1)] - [\partial_{\rho^4}\chi^9 - (t^9_4/\rho^4)\chi^9] \exp[i(t^9_b\theta^b + \theta^4)] \\
&\quad - i[\partial_{\rho^3}\chi^{12} + (t^{12}_3/\rho^3)\chi^{12}] \exp[i(t^{12}_b\theta^b - \theta^3)] - i[\partial_{\rho^2}\chi^{14} + (t^{14}_2/\rho^2)\chi^{14}]
\end{aligned}$$

$$
\begin{aligned}
&\quad\times \exp[i(t_b^{14}\theta^b-\theta^2)]-im\chi^{10}\exp[i(t_b^{10}\theta^b)]=0,\\
&i[\partial_{\rho^1}\chi^3-(t_1^3/\rho^1)\chi^3]\exp[i(t_b^3\theta^b+\theta^1)]-i[\partial_{\rho^3}\chi^9-(t_3^9/\rho^3)\chi^9]\exp[i(t_b^9\theta^b+\theta^3)]\\
&\quad+[\partial_{\rho^4}\chi^{12}+(t_4^{12}/\rho^4)\chi^{12}]\exp[i(t_b^{12}\theta^b-\theta^4)]-i[\partial_{\rho^2}\chi^{15}+(t_2^{15}/\rho^2)\chi^{15}]\\
&\quad\times\exp[i(t_b^{15}\theta^b-\theta^2)]-im\chi^{11}\exp[i(t_b^{11}\theta^b)]=0,\\
&i[\partial_{\rho^1}\chi^4-(t_1^4/\rho^1)\chi^4]\exp[i(t_b^4\theta^b+\theta^1)]-i[\partial_{\rho^3}\chi^{10}-(t_3^{10}/\rho^3)\chi^{10}]\exp[i(t_b^{10}\theta^b+\theta^3)]\\
&\quad+[\partial_{\rho^4}\chi^{11}-(t_4^{11}/\rho^4)\chi^{11}]\exp[i(t_b^{11}\theta^b+\theta^4)]-i[\partial_{\rho^2}\chi^{16}+(t_2^{16}/\rho^2)\chi^{16}]\\
&\quad\times \exp[i(t_b^{16}\theta^b-\theta^2)]-im\chi^{12}\exp[i(t_b^{12}\theta^b)]=0,\\
&i[\partial_{\rho^1}\chi^5-(t_1^5/\rho^1)\chi^5]\exp[i(t_b^5\theta^b+\theta^1)]-i[\partial_{\rho^2}\chi^9-(t_2^9/\rho^2)\chi^9]\exp[i(t_b^9\theta^b+\theta^2)]\\
&\quad+[\partial_{\rho^4}\chi^{14}+(t_4^{14}/\rho^4)\chi^{14}]\exp[i(t_b^{14}\theta^b-\theta^4)]+i[\partial_{\rho^3}\chi^{15}+(t_3^{15}/\rho^3)\chi^{15}]\\
&\quad\times \exp[i(t_b^{15}\theta^b-\theta^3)]-im\chi^{13}\exp[i(t_b^{13}\theta^b)]=0,\\
&i[\partial_{\rho^1}\chi^6-(t_1^6/\rho^1)\chi^6]\exp[i(t_b^6\theta^b+\theta^1)]-i[\partial_{\rho^2}\chi^{10}-(t_2^{10}/\rho^2)\chi^{10}]\exp[i(t_b^{10}\theta^b+\theta^2)]\\
&+[\partial_{\rho^4}\chi^{13}-(t_4^{13}/\rho^4)\chi^{13}]\exp[i(t_b^{13}\theta^b+\theta^4)]+i[\partial_{\rho^3}\chi^{16}+(t_3^{16}/\rho^3)\chi^{16}]\exp[i(t_b^{16}\theta^b-\theta^3)]-im\chi^{14}\exp[i(t_b^{14}\theta^b)]=0,\\
&i[\partial_{\rho^1}\chi^7-(t_1^7/\rho^1)\chi^7]\exp[i(t_b^7\theta^b+\theta^1)]-i[\partial_{\rho^2}\chi^{11}-(t_2^{11}/\rho^2)\chi^{11}]\exp[i(t_b^{11}\theta^b+\theta^2)]\\
&+i[\partial_{\rho^3}\chi^{13}-(t_3^{13}/\rho^3)\chi^{13}]\exp[i(t_b^{13}\theta^b+\theta^3)]-[\partial_{\rho^4}\chi^{16}+(t_4^{16}/\rho^4)\chi^{16}]\exp[i(t_b^{16}\theta^b-\theta^4)]-im\chi^{15}\exp[i(t_b^{15}\theta^b)]=0,\\
&i[\partial_{\rho^1}\chi^8-(t_1^8/\rho^1)\chi^8]\exp[i(t_b^8\theta^b+\theta^1)]-i[\partial_{\rho^2}\chi^{12}-(t_2^{12}/\rho^2)\chi^{12}]\exp[i(t_b^{12}\theta^b+\theta^2)]\\
&+i[\partial_{\rho^3}\chi^{14}-(t_3^{14}/\rho^3)\chi^{14}]\exp[i(t_b^{14}\theta^b+\theta^3)]-[\partial_{\rho^4}\chi^{15}-(t_4^{15}/\rho^4)\chi^{15}]\exp[i(t_b^{15}\theta^b+\theta^4)]-im\chi^{16}\exp[i(t_b^{16}\theta^b)]=0.
\end{aligned}
\tag{AII1}
$$

In order to cancel out the angular functions, e.g., from the first equation, one must have

$$
\begin{aligned}
&t_b^2\theta^b-\theta^4\\
&\quad=t_b^3\theta^b-\theta^3\\
&\quad=t_b^5\theta^b-\theta^2=t_b^9\theta^b-\theta^1=t_b^1\theta^b,
\end{aligned}
\tag{AII2}
$$

for $0\leqslant\theta^b<2\pi$ (or else $0\leqslant\theta^b<4\pi$. Equating the coefficients of θ^b one obtains

$$
\begin{aligned}
&t_1^2=t_1^3=t_1^5=t_1^9-1=t_1^1,\\
&t_2^2=t_2^3=t_2^5-1=t_2^9=t_2^1,\\
&t_3^2=t_3^3-1=t_3^5=t_3^9=t_3^1,\\
&t_4^2-1=t_4^3=t_4^5=t_4^9=t_4^1.
\end{aligned}
\tag{AII3}
$$

From the remaining 15 equations in (AII1) similar relations follow. Arranging all this information in a concise manner yields Eq. (4.5). It may be argued here that the demand of cancelling the angular functions would restrict the general solutions to special cases. But that is not so. It can be strictly proved that the angular functions can be taken out. To give the essential ideas in the proof, a similar problem in a simpler setting is considered. Suppose that $F_1(r), F_2(r), F_3(r)$ are nonvanishing complex-valued functions in an interval of the real variable r, and m_1, m_2, m_3 are real constants and ϕ is another real variable representing an angle. Furthermore, $F_1(r)e^{im_1\phi}+F_2(r)e^{im_2\phi}+F_3(r)e^{im_3\phi}=0$. Dividing by $ie^{im_3\phi}$ and differentiating with respect to ϕ one obtains $(m_1-m_3)F_1(r)\times\exp[i(m_1-m_3)\phi]+(m_2-m_3)F_2(r)\times\exp[i(m_2-m_3)\phi]=0$. Dividing by $ie^{i(m_2-m_3)\phi}$ and differentiating with respect to ϕ one obtains $(m_1-m_2)(m_1-m_3)F_1(r)\exp[i(m_1-m_2)\phi]=0$. Therefore, either $m_1=m_2$ or $m_1=m_3$. Similarly, it follows that either $m_2=m_3$ or $m_2=m_1$ and either $m_3=m_1$ or $m_3=m_2$. Therefore $m_1=m_2=m_3$, and the angular functions can be cancelled out.

Now the functions χ^L can be expressed in terms of the Bessel functions as follows:

$$
\chi^L(\rho)=u^L(\underline{\kappa})J_{t_1}(\kappa_1\rho^1)J_{t_2}(\kappa_2\rho^2)J_{t_3}(\kappa_3\rho^3)J_{t_4}(\kappa_4\rho^4),
\tag{AII4}
$$

where $u^L(\underline{\kappa})$ is the amplitude factor.

Substituting the above into, e.g., the first of the equations, removing the angular functions, and using $(\partial_\rho\mp t/\rho)J_t(\kappa\rho)=\mp\kappa J_{t\pm1}(\kappa\rho)$ one obtains

$$
\begin{aligned}
&\kappa_4u^2J_{t_1^2}J_{t_2^2}J_{t_3^2}J_{t_4^2-1}+i\kappa_3u^3J_{t_1^3}J_{t_2^3}J_{t_3^3-1}J_{t_4^3}\\
&\quad+i\kappa_2u^5J_{t_1^5}J_{t_2^5-1}J_{t_3^5}J_{t_4^5}+i\kappa_1u^9J_{t_1^9-1}J_{t_2^9}J_{t_3^9}J_{t_4^9}\\
&\quad-imu^1J_{t_1^1}J_{t_2^1}J_{t_3^1}J_{t_4^1}=0.
\end{aligned}
\tag{AII5}
$$

Using the relations (4.5) among t_b^L's one can see that the Bessel functions can be cancelled out and the above equation reduces to

$$
\kappa_4u^2+i\kappa_3u^3+i\kappa_2u^5+i\kappa_1u^9-imu^1=0.
\tag{AII6}
$$

Similar simplifications occur in the other 15 equations in (AII1) so that Eq. (4.9) results.

APPENDIX III

The computations of the integral constants $N^{(t)}$, $H^{(t)}$, $K_\alpha^{(t)}$,$X_\alpha^{(t)}$, $T^{(t)}$,$B^{(t)}$ from the spinor field

$$
\begin{aligned}
\psi^{(t)L}=(2\pi)^{-3/2}m^{1/2}&\sum_{r=1}^{2}\sum_{\alpha=1}^{2}\int_0^\infty d^3\kappa\,\kappa_1\kappa_3\kappa_3\\
&\times[\beta_{(r\alpha)}(\underline{\kappa},t)u^{(r\alpha)L}+\bar{\gamma}_{(r\alpha)}(\underline{\kappa},t)v^{(r\alpha)L}]\\
&\times J_{t_1^L}(\kappa_1\rho^1)J_{t_2^L}(\kappa_2\rho^2)J_{t_3^L}(\kappa_3\rho^3)J_{t_4^L}(E\rho^4)\exp(it_b^L\theta^b),
\end{aligned}
\tag{AIII1}
$$

where L is not summed, will be shown. The ρ^4-integration will be restricted to the interval $0<\rho^4<\infty$. The integral constant $N^{(t)}$ computed at $\theta^4=\pi/2$ from (2.12) is

$$
N^{(t)}=[N^{(t)}]_{\theta^4=\pi/2}
$$

$$= (2\pi)^{-3} m \sum_{rs\alpha\beta LM} \int_0^{2\pi} d^3\theta \int_0^{\infty} d^3\rho\, \rho \rho^1\rho^2\rho^3$$
$$\times \int_0^{\infty} d\rho^4 \int_0^{\infty} d^3\kappa\, \kappa_1\kappa_2\kappa_3$$
$$\int_0^{\infty} d^3\kappa' \kappa'_1\kappa'_2\kappa'_3 \,[\bar{\beta}_{(r\alpha)} \bar{u}'^{(r\alpha)L} + \gamma_{(r\alpha)} \bar{v}'^{(r\alpha)L}]$$
$$\times (\Lambda v^4)_{LM} [\beta'_{(s\beta)} u'^{(s\beta)M} + \bar{\gamma}'_{(s\beta)} v'^{(s\beta)M}]$$
$$\times J_{t^L_1}(\kappa_1\rho^1) J_{t^M_1}(\kappa'_1 \rho^1) \cdots J_{t^L_4}(E\rho^4) J_{t^M_4}(E'\rho^4)$$
$$\times \exp[i(t^M_\mu - t^L_\mu)\theta^\mu] [(i)^{t^M_4 - t^L_4}]. \quad \text{(AIII2)}$$

Recalling the integrals

$$(2\pi)^{-1} \int_0^{2\pi} e^{i(t-t')\theta} d\theta = \delta_{tt'},$$
$$\int_0^{\infty} d\rho\, \rho J_t(\kappa\rho) J_t(\kappa'\rho) = (\kappa)^{-1} \delta(\kappa - \kappa'),$$
$$\int_0^{\infty} d\rho\, J_t(E\rho) J_{t+1}(E\rho) = (2E)^{-1}, \quad \text{(AIII3)}$$

and performing $d^3\theta$, $d^3\rho$, $d^3\kappa'$, $d\rho^4$ integrations (in that order), (AIII2) goes over into

$$N_{(t)} = \frac{m}{2} \sum_{rs\alpha\beta LM} d^3\kappa\, \kappa_1\kappa_2\kappa_3 (E)^{-1} [\bar{\beta}_{(r\alpha)} \bar{u}^{(r\alpha)L} + \gamma_{(r\alpha)} \bar{v}^{(r\alpha)L}]$$
$$\times [\Lambda v^4]_{LM} [\beta_{(s\beta)} u^{(s\beta)M} + \bar{\gamma}_{(s\beta)} v^{(s\beta)M}]$$
$$\times [(i)^{t^M_4 - t^L_4}]. \quad \text{(AIII4)}$$

Define a matrix M with entries

$$[M]_{LM} \equiv [(i)^{t^M_4 - t^L_4}][\Lambda v^4]_{LM} = -[I \times I \times I \times \sigma^2]_{LM}, \quad \text{(AIII5)}$$

where Eqs. (4.5) have been used.. From the solutions in Table I it follows that

$$u^\dagger_{(r\alpha)} M u_{(s\beta)} = v^\dagger_{(r\alpha)} M v_{(s\beta)} = -(E/m)[\sigma^2 \times I]_{rs} \delta_{\alpha\beta}, \quad \text{(AIII6)}$$
$$u^\dagger_{(r\alpha)} M v_{(s\beta)} = v^\dagger_{(r\alpha)} M u_{(s\beta)} = 0.$$

With (AIII5), (AIII6) the integral constant $N^{(t)}$ becomes

$$N^{(t)} = \frac{m}{2} \sum_{rs\alpha\beta} \int_0^{\infty} d^3\kappa\, \kappa_1\kappa_2\kappa_3 (E)^{-1} [\bar{\beta}_{(r\alpha)} u^{\dagger(r\alpha)} + \gamma_{(r\alpha)} v^{\dagger(r\alpha)}]$$
$$\times M [\beta_{(s\beta)} u^{(s\beta)} + \bar{\gamma}_{(s\beta)} v^{(s\beta)}]$$
$$= -\tfrac{1}{2} \sum_{rs\alpha} \int_0^{\infty} d^3\kappa\, \kappa_1\kappa_2\kappa_3$$
$$\times [\bar{\beta}_{(r\alpha)} \beta_{(s\alpha)} + \gamma_{(r\alpha)} \bar{\gamma}_{(s\alpha)}][\sigma^2 \times I]_{rs}. \quad \text{(AIII7)}$$

Making the following invertible linear transformation on the amplitudes $\beta_{(r\alpha)}, \gamma_{(r\alpha)}$:

$$\beta_{(1\alpha)} = \mu_{(1\alpha)} + i\mu_{(3\alpha)}, \quad \beta_{(3\alpha)} = \mu_{(3\alpha)} + i\mu_{(1\alpha)},$$
$$\beta_{(2\alpha)} = \mu_{(2\alpha)} + i\mu_{(4\alpha)}, \quad \beta_{(4\alpha)} = \mu_{(4\alpha)} + i\mu_{(2\alpha)}, \quad \text{(AIII8)}$$
$$\gamma_{(1\alpha)} = \bar{\mu}_{(5\alpha)} - i\bar{\mu}_{(7\alpha)}, \quad \gamma_{(3\alpha)} = \bar{\mu}_{(7\alpha)} - i\bar{\mu}_{(5\alpha)},$$
$$\gamma_{(2\alpha)} = \bar{\mu}_{(6\alpha)} - i\bar{\mu}_{(8\alpha)}, \quad \gamma_{(4\alpha)} = \bar{\mu}_{(8\alpha)} - i\bar{\mu}_{(6\alpha)},$$

Eq. (AIII7) yields

$$N^{(t)} = \sum_{R=1}^{8} \sum_{\alpha=1}^{2} \int_0^{\infty} d^3\kappa\, \kappa_1\kappa_2\kappa_3 [\epsilon_R |\mu_{R\alpha}(\underline{\kappa}, t)|^2], \quad \text{(AIII9)}$$

where

$\epsilon_1 = \epsilon_2 = -\epsilon_3 = -\epsilon_4 = \epsilon_5 = \epsilon_6 = -\epsilon_7 = -\epsilon_8 = 1.$

Now computing $H^{(t)}$ from (2.12), (AIII1), (II-AII1c) one obtains

$$H^{(t)} = [H^{(t)}]_{\theta^4 = \pi/2}$$
$$= (2i)^{-1} m (2\pi)^{-3} \sum_{rs\alpha\beta LM} \int_0^{2\pi} d^3\rho\, \rho \rho^1\rho^2\rho^2$$
$$\times \int_0^{\infty} d\rho^4 \int_0^{\infty} d^3\kappa\, \kappa_1\kappa_2\kappa_3 \int_0^{\infty} d^3\kappa'\, \kappa'_1\kappa'_2\kappa'_3$$
$$\times [\bar{\beta}_{(r\alpha)} \bar{u}^{(r\alpha)L} + \gamma_{(r\alpha)} \bar{v}^{(r\alpha)L}]$$
$$\times (\Lambda v^4)_{LM} [\beta'_{(s\beta)} u'^{(s\beta)M} + \bar{\gamma}'_{(s\beta)} v'^{(s\beta)M}]$$
$$\times J_{t^L_1}(\kappa_1\rho^1) J_{t^M_1}(\kappa'_1\rho^1) \cdots J_{t^L_4}(E\rho^4) J_{t^M_4}(E'\rho^4)$$
$$\times \exp[i(t^M_\mu - t^L_\mu)\theta^\mu]$$
$$\times [(i)^{t^M_4 - t^L_4}](it^L_4/\rho^4) + (\text{c.c.}). \quad \text{(AIII10)}$$

Performing $d^3\theta, d^3\rho, d^3\kappa', d\rho^4$ integrations respectively and remembering (AIII3) and also the integral

$$\int_0^{\infty} d\rho\, (\rho)^{-1} J_t(E\rho) J_{t+1}(E\rho) = [\pi(t+\tfrac{1}{2})]^{-1}, \quad \text{(AIII11)}$$

one gets from (AIII10)

$$H^{(t)} = \frac{m}{\pi} \sum_{rs\alpha\beta LM} \int_0^{\infty} d^3\kappa\, \kappa_1\kappa_2\kappa_3 [\bar{\beta}_{(r\alpha)} \bar{u}^{(r\alpha)L} + \gamma_{(r\alpha)} \bar{v}^{(r\alpha)L}]$$
$$\times [\Lambda v^4]_{LM} [\beta_{(s\beta)} u^{(s\beta)M} + \bar{\gamma}_{(s\beta)} v^{(s\beta)M}] t^L_4 (t^L_4 + t^M_4)^{-1}$$
$$\times [(i)^{t^M_4 - t^L_4}] + (\text{c.c.}). \quad \text{(AIII12)}$$

Now another matrix M' can be defined as

$$[M']_{LM} \equiv [(i)^{t^M_4 - t^L_4}] t^L_4 (t^L_4 + t^M_4)^{-1} [\Lambda v^4]_{LM}. \quad \text{(AIII13)}$$

Using (4.5), (4.6), (AIII5) the matrix M' can be expressed as

$$M' = (1/2)M - (i/2)(2t_4 + 1)^{-1} [\Lambda v^4]. \quad \text{(AIII14)}$$

By (AIII14), (4.13), (AIII6), Eq. (AIII12) becomes

$$H^{(t)} = [m/\pi(2t_4+1)] \sum_{rs\alpha\beta} \int_0^{\infty} d^3\kappa\, \kappa_1\kappa_2\kappa_3$$
$$\times [\bar{\beta}_{(r\alpha)} u^{\dagger(r\alpha)} + \gamma_{(r\alpha)} v^{\dagger(r\alpha)}] M'$$
$$\times [\beta_{(s\beta)} u^{(s\beta)} + \bar{\gamma}_{(s\beta)} v^{(s\beta)}] + (\text{h.c.})$$
$$= [m/\pi(2t_4+1)] \sum_{rs\alpha} \int_0^{\infty} d^3\kappa\, \kappa_1\kappa_2\kappa_3$$
$$\times ((t_4 + \tfrac{1}{2})(-E/m)$$
$$\times \{[\sigma^2 \times I]_{rs} [\bar{\beta}_{(r\alpha)} \beta_{(s\alpha)} + \gamma_{(r\alpha)} \bar{\gamma}_{(s\alpha)}]\}$$
$$+ \tfrac{1}{2} \{(\gamma^4)_{rs} [\bar{\beta}_{(r\alpha)} \bar{\gamma}_{(s\alpha)} - \gamma_{r\alpha} \beta_{s\alpha}]\}) + (\text{c.c.}). \quad \text{(AIII15)}$$

Recalling (AIII8) and noticing that the last curly bracket is cancelled out by the counterpart in (c.c.), one finally arrives at the expression

$$H^{(t)} = \frac{2}{\pi} \sum_{R=1}^{8} \sum_{\alpha=1}^{2}$$
$$\times \int_0^{\infty} d^3\kappa\, \kappa_1\kappa_2\kappa_3 [\epsilon_R |\mu_{(R\alpha)}(\underline{\kappa}, t)|^2] E(\kappa). \quad \text{(AIII16)}$$

For the integral constants $K^{(t)}_\alpha$, $X^{(t)}_\alpha$, one finds

$$K^{(t)}_\alpha = X^{(t)}_\alpha = 0. \quad \text{(AIII17)}$$

This is because the $d^3\theta$ integrations bring in some restrictions on t^L_μ, t^M_μ, which in turn require that only the zero entries of $[\Lambda v^4]_{LM}$ can contribute.

Next the integral constant $T^{(t)}$ is computed. From (2.12) and (AIII1) one obtains

$$\begin{aligned}
T^{(t)} &= [T^{(t)}]_{\theta^4=\pi/2}\\
&= \frac{im}{2}(2\pi)^{-3}\sum_{rs\alpha\beta LM}\int_0^{2\pi} d^3\theta \int_0^\infty d^3\rho\,\rho^1\rho^2\rho^3\\
&\quad\times\int_0^\infty d\rho^4 \int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3\int_0^\infty d^3\kappa'\,\kappa'_1\kappa'_2\kappa'_3\\
&\quad\times[\bar\beta_{(r\alpha)}\bar u^{(r\alpha)L} + \gamma_{(r\alpha)}\bar v^{(r\alpha)L}][\Lambda v^4]_{LM}\\
&\quad\times[\beta'_{(s\beta)}u'^{(s\beta)M} + \bar\gamma'_{(s\beta)}v'^{(s\beta)M}]\\
&\quad\times J_{t_1^L}(\kappa_1\rho^1)\cdots J_{t_4^L}(E\rho^4)[e^{-it^L_b\theta^b}]_{\theta^4=\pi/2}\\
&\quad\times[(\partial_{\theta^1}+\partial_{\theta^2}+\partial_{\theta^3}+\partial_{\theta^4})J_{t_1^M}(\kappa'_1\rho^1)\cdots J_{t_4^M}(E'\rho^4)\\
&\quad\times e^{it^M_a\theta^a}]_{\theta^4=\pi/2} + (\text{c.c})\\
&= -\frac{m}{2}\sum_{rs\alpha\beta LM}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3(2E)^{-1}[\bar\beta_{(r\alpha)}\bar u^{(r\alpha)L}\\
&\quad+\gamma_{(r\alpha)}\bar v^{(r\alpha)L}][\Lambda v^4]_{LM}[\beta_{(s\beta)}u^{(s\beta)M}+\bar\gamma_{(s\beta)}v^{(s\beta)M}]\\
&\quad\times(t_1^M+t_2^M+t_3^M+t_4^M)[(i)^{t_4^M-t_4^L}] + (\text{c.c}),
\end{aligned}\tag{AIII18}$$

where in the last line $d^3\theta$, $d^3\rho$, $d^3\kappa'$, $d\rho^4$ integrations have been carried out with the help of (AIII3). Now the following matrix is introduced:

$$\begin{aligned}
&[M'']_{LM}\equiv[(i)^{t_4^M-t_4^L}](t_1^M+t_2^M+t_3^M+t_4^M)[\Lambda v^4]_{LM},\\
&M''=(t_1+t_2+t_3+t_4+\tfrac{1}{2})M+(i/2)[\Lambda v^4],
\end{aligned}\tag{AIII19}$$

where in the last equation, the formulas (4.5), (4.6) have been used. Applying (AIII19), (4.13), (AIII6), the integral constant $T^{(t)}$ in (AIII18) reduces to

$$\begin{aligned}
T^{(t)}
&= -\frac{m}{4}\sum_{rs\alpha}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3(E)^{-1}((t_1+t_2+t_3+t_4+\tfrac{1}{2})\\
&\quad\times(-E/m)\{(\sigma^2\times I)_{rs}[\bar\beta_{(r\alpha)}\beta_{(s\alpha)}+\gamma_{(r\alpha)}\bar\gamma_{(s\alpha)}]\}\\
&\quad-\tfrac{1}{2}\{(\gamma^4)_{rs}[\bar\beta_{(r\alpha)}\bar\beta_{(s\alpha)}-\gamma_{(r\alpha)}\beta_{(s\alpha)}]\}) + (\text{c.c.}).
\end{aligned}\tag{AIII20}$$

The last curly bracket is cancelled by the corresponding term in (c.c.) so that by using (AIII8), (AIII9) one has

$$T^{(t)}=-(t_1+t_2+t_3+t_4+\tfrac{1}{2})N^{(t)}.\tag{AIII21}$$

The calculation of $B^{(t)}$ in (2.12) will be done now.

$$\begin{aligned}
B^{(t)} &= [B^{(t)}]_{\theta^4=\pi/2}\\
&= \frac{im}{4}(2\pi)^{-3}\sum_{rs\alpha\beta LM}\int_0^{2\pi} d^3\theta\int_0^\infty d^3\rho\,\rho^1\rho^2\rho^3\\
&\quad\times\int_0^\infty d\rho^4\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3\int_0^\infty d^3\kappa\,\kappa'_1\kappa'_2\kappa'_3\\
&\quad\times[\bar\beta_{(r\alpha)}\bar u^{(r\alpha)L}+\gamma_{(r\alpha)}\bar v^{(r\alpha)L}][\Lambda v^4(v^i a_i - a^i v_i)]_{LM}\\
&\quad\times[\beta'_{(s\beta)}u'^{(s\beta)M}+\bar\gamma'_{(s\beta)}v'^{(s\beta)M}]J_{t_1^L}(\kappa_1\rho^1)\\
&\quad\times J_{t_1^M}(\kappa'_1\rho^1)\cdots J_{t_4^L}(E\rho^4)J_{t_4^M}(E'\rho^4)\\
&\quad\times\exp[i(t^M_\mu-t^L_\mu)\theta^\mu][(i)^{t_4^M-t_4^L}] + (\text{c.c}).
\end{aligned}\tag{AIII22}$$

Performing $d^3\theta$, $d^3\rho$, $d^3\kappa'$, $d\rho^4$ integrations one has

$$\begin{aligned}
B^{(t)} &= \frac{im}{8}\sum_{rs\alpha\beta LM}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3(E)^{-1}\\
&\quad\times[\bar\beta_{(r\alpha)}\bar u^{(r\alpha)L}+\gamma_{(r\alpha)}\bar v^{(r\alpha)L}][\Lambda v^4(v^i a_i - a^i v_i)]_{LM}\\
&\quad\times[\beta_{(s\beta)}u^{(s\beta)M}+\bar\gamma_{s\beta}v^{(s\beta)M}][(i)^{t_4^M-t_4^L}] + (\text{c.c.}).
\end{aligned}\tag{AIII23}$$

One can introduce a matrix

$$\begin{aligned}
&[M_1]_{LM}\equiv-[(i)^{1+t_4^M-t_4^L}][\Lambda v^4(v^i a_i - a^i v_i)]_{LM},\\
&M_1 = 2M'_1 - 2i[\Lambda v^4],\\
&M'_1\equiv\text{diag}[-3\sigma^2,-\sigma^2,-\sigma^2,\sigma^2,-\sigma^2,\sigma^2,\sigma^2,3\sigma^2].
\end{aligned}\tag{AIII24}$$

From Table I it follows that

$$\begin{aligned}
&N_{(r\alpha)(s\beta)}\equiv u^\dagger_{(r\alpha)}M'_1u_{(s\beta)}=\begin{bmatrix}A & 0\\ 0 & A\end{bmatrix},\\
&N'_{(r\alpha)(s\beta)}\equiv v^\dagger_{(r\alpha)}M'_1v_{(s\beta)}=-N_{(r\alpha)(s\beta)},\\
&N''_{(r\alpha)(s\beta)}\equiv u^\dagger_{(r\alpha)}M'_1v_{(s\beta)}=\begin{bmatrix}B & 0\\ 0 & B\end{bmatrix},\\
&A\equiv\begin{bmatrix}a & \beta & \gamma & \delta\\ \bar\beta & -a & \delta' & -\gamma\\ \bar\gamma & \bar\delta' & -a & -\beta\\ \bar\delta & -\bar\gamma & -\bar\beta & a\end{bmatrix},\\
&B\equiv\begin{bmatrix}l & 0 & \lambda & 0\\ 0 & l & 0 & \lambda\\ \bar\lambda & 0 & -\lambda & 0\\ 0 & \bar\lambda & 0 & -l\end{bmatrix},\\
&a\equiv\kappa_2\kappa_3[m(E+m)]^{-1},\\
&\beta\equiv i(m)^{-1}[E-\kappa_2(\kappa_2+i\kappa_1)(E+m)^{-1}],\\
&\gamma\equiv(m)^{-1}[-E+\kappa_3(\kappa_3+i\kappa_1)(E+m)^{-1}],\\
&\delta\equiv(m)^{-1}[-iE+(\kappa_1-i\kappa_2)(\kappa_3+i\kappa_1)(E+m)^{-1}],\\
&\delta'\equiv(m)^{-1}[-iE+(\kappa_1+i\kappa_2)(\kappa_3+i\kappa_1)(E+m)^{-1}],\\
&l=-\kappa_2(m)^{-1},\quad\lambda=-(m)^{-1}(\kappa_3+i\kappa_1).
\end{aligned}\tag{AIII25}$$

By (AIII24), (AIII25) the expression for $B^{(t)}$ becomes

$$\begin{aligned}
B^{(t)} &= -\frac{m}{4}\sum_{rs\alpha\beta}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3(E)^{-1}\\
&\quad\times\{[\bar\beta_{(r\alpha)}\beta_{(s\beta)}-\gamma_{(r\alpha)}\bar\gamma_{(s\beta)}]N_{(r\alpha)(s\beta)}\\
&\quad+\bar\beta_{(r\alpha)}\bar\gamma_{(s\beta)}N''_{(r\alpha)(s\beta)}+\beta_{(r\alpha)}\gamma_{(s\beta)}\bar N''_{(r\alpha)(s\beta)}\}\\
&\quad+\frac{im}{4}\sum_{rs\alpha\beta LM}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3(E)^{-1}\\
&\quad\times[\bar\beta_{(r\alpha)}\bar u^{(r\alpha)L}+\gamma_{(r\alpha)}v^{(r\alpha)L}][\Lambda v^4]_{LM}\\
&\quad\times[\beta_{(s\beta)}u^{(s\beta)M}+\bar\gamma_{(s\beta)}v^{(s\beta)M}] + (\text{c.c.}).
\end{aligned}\tag{AIII26}$$

By (4.13) the second integral of the above expression is cancelled by its counterpart in (c.c.). Therefore, using (AIII8) one obtains

$$\begin{aligned}
B^{(t)} &= -\frac{m}{2}\sum_{rs\alpha\beta}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3(E)^{-1}\\
&\quad\{[\bar\beta_{(r\alpha)}\beta_{(s\beta)}-\gamma_{(r\alpha)}\bar\gamma_{(s\beta)}]N_{(r\alpha)(s\beta)}\\
&\quad+\bar\beta_{(r\alpha)}\bar\gamma_{(s\beta)}N''_{(r\alpha)(s\beta)}+\beta_{(r\alpha)}\gamma_{(s\beta)}\bar N''_{(r\alpha)(s\beta)}\}\\
&= -m\sum_{RS\alpha\beta}\int_0^\infty d^3\kappa\,\kappa_1\kappa_2\kappa_3(E)^{-1}\\
&\quad\times[\bar\mu_{(R\alpha)}N_{(R\alpha)(s\beta)}\mu_{(S)}],
\end{aligned}\tag{AIII27}$$

$$[N_{(R\alpha)(s\beta)}] \equiv \begin{bmatrix} N_{(r\alpha)(s\beta)} & N''_{(r\alpha)(s\beta)} \\ \bar{N}''_{(r\alpha)(s\beta)} & -N_{(r\alpha)(s\beta)} \end{bmatrix}. \qquad \text{(AIII27)}$$

The integrand of $B^{(t)}$ is a Hermitian form on the 16 ordered tuples $\mu_{(R\alpha)}$. It is not in the convenient diagonal form. A similar situation arises with the Dirac wave field if one tries to compute a single component of the total spin angular momentum with the usual plane wave solutions. The situation can be remedied by a diagonalization which would not disturb the already diagonalized Hermitian form of the integrand of $N^{(t)}$. It would suffice for the present purpose to show the existence of such a diagonalization. To do that, the already diagonalized Hermitian form of the integrand of $N^{(t)}$ yields the corresponding matrix ("metric tensor") $[I'] \equiv \text{diag}[\epsilon_1,\epsilon_2,...,\epsilon_8]$, where $\epsilon_R = \pm 1$ are given in (AIII9). The pseudounitary matrices W acting on V_{16} are defined by the equation

$$W^\dagger I' W = I'. \qquad \text{(AIII28)}$$

The eigenvalue equation for the Hermitian matrix $N \equiv [N_{(R\alpha)(s\beta)}]$ relative to the metric I' is

$$[N - \lambda I']\mathbf{u} = \mathbf{0}, \quad \det[N - \lambda I'] = 0. \qquad \text{(AIII29)}$$

An "orthonormal" basis of V_{16} must satisfy

$$\mathbf{e}_{(R\alpha)} I' \mathbf{e}_{(s\beta)} = I'_{(R\alpha)(s\beta)}. \qquad \text{(AIII30)}$$

A matrix which has for the columns 16 "orthonormal" vectors

$$W \equiv [\mathbf{e}_{(1,1)},\mathbf{e}_{(1,2)},...,\mathbf{e}_{(8,2)}], \qquad \text{(AIII31)}$$

must be pseudounitary. Let the first vector of a pseudounitary matrix W_1 be a "normalized" eigenvector $\mathbf{u}_{(1,1)}$ of N corresponding to the real eigenvalue $\lambda_{(1,1)}$ which need not be simple. It can be asserted that 15 other "orthonormal" vectors $\mathbf{z}_{(2,1)},...,\mathbf{z}_{(8,2)}$ exist such that $[\mathbf{u}_{(1,1)},\mathbf{z}_{(2,1)},...,\mathbf{z}_{(8,2)}]$ is an "orthonormal" basis set. Therefore W_1 can be expressed as

$$W_1 = [\mathbf{u}_{(1,1)},\mathbf{z}_{(2,1)},...,\mathbf{z}_{(8,2)}] \equiv [\mathbf{u}_{(1,1)},Z].$$

The condition (AIII28) can be expressed as

$$\begin{bmatrix} u^\dagger_{(1,1)} \\ Z^\dagger \end{bmatrix} [I'u_{(1,1)}, I'Z] = \begin{bmatrix} \mathbf{u}^\dagger_{(1,1)} I' \mathbf{u}_{(1,1)}, & \mathbf{u}^\dagger_{(1,1)} I' Z \\ Z^\dagger I' \mathbf{u}_{(1,1)} & Z^\dagger I' Z \end{bmatrix} = \begin{bmatrix} \epsilon_1 & 0 \\ 0 & I'' \end{bmatrix}, \qquad \text{(AIII33)}$$

implying that $Z^\dagger I' u_{(1,1)} = 0 = u^\dagger_{(1,1)} I' Z$.

Now consider

$$W_1^\dagger N W_1 = \begin{bmatrix} \mathbf{u}^\dagger_{(1,1)} \\ Z^\dagger \end{bmatrix} [N\mathbf{u}_{(1,1)} NZ] = \begin{bmatrix} \lambda_{(1,1)} [\mathbf{u}^\dagger_{(1,1)} I' \mathbf{u}_{(1,1)}], & \mathbf{u}^\dagger_{(1,1)} NZ \\ \lambda_{(1,1)} [Z^\dagger I' \mathbf{u}_{(1,1)}] & Z^\dagger NZ \end{bmatrix}. \qquad \text{(AIII34)}$$

Using (AIII30), (AIII33), and noticing that $W_1^\dagger N W_1$ must be a Hermitian matrix, one obtains from (AIII34)

$$W_1^\dagger N W_1 = \begin{bmatrix} \lambda_{(1,1)}\epsilon_1 & 0 \\ 0 & Z^\dagger NZ \end{bmatrix}.$$

Repeating similar arguments, one can show the existence of a pseudounitary matrix W such that

$$W^\dagger N W = \text{diag}[\lambda_{(1,1)}\epsilon_1, \lambda_{(1,2)}\epsilon_1, ..., \lambda_{(8,1)}\epsilon_8, \lambda_{(8,2)}\epsilon_8],$$
$$W^\dagger I' W = \text{diag}[\epsilon_1,\epsilon_1,...,\epsilon_8,\epsilon_8], \qquad \text{(AIII35)}$$

where all eigenvalues $\lambda_{(R\alpha)}$ may not be distinct.

Now writing

$$\mu_{(R\alpha)} = \sum_{s,\beta} W_{(R\alpha)(s\beta)} \alpha_{(s\beta)}, \qquad \text{(AIII36)}$$

the integrands $\mathscr{I}$ of $N^{(t)}$ and $B^{(t)}$ respectively become

$$\mathscr{I}[N^{(t)}] = \sum_{R,\alpha} \epsilon_R |\alpha_{(R\alpha)}|^2,$$
$$\mathscr{I}[B^{(t)}] = \sum_{R,\alpha} \epsilon_R \lambda_{(R\alpha)} |\alpha_{(R\alpha)}|^2. \qquad \text{(AIII37)}$$

The eigenvalues $\lambda_{(R\alpha)}$ have been computed on an IBM-360 and they come out to be

$$(mE)^{-1}\lambda_{(R\alpha)} = (1)^6, (-1)^6, (3)^2, (-3)^2, \qquad \text{(AIII38)}$$

where the exponents stand for the multiplicities.

[1]H. Yukawa, Phys. Rev. **76**, 300 (1949); **77**, 849 (1950); **80**, 1047 (1950).

[2]A. Das, J. Math. Phys. **7**, 45 (1966).

[3]A. Das, J. Math. Phys. **21**, 1506 (1980).

[4]M. Gell-Mann, California Inst. of Tech. Lab. Report CTSL-20 (1961).

[5]It should be *emphasized* here that the usual mass operator in Eq. (6.1) of the paper by L. O'Raifartaigh, Phys. Rev. **139**, 258 (1965), for the proof of the NO-GO theorem is *not assumed* in the present paper.

[6]S. Okubo, Prog. Theor. Phys. (Kyoto) **27**, 949 (1962).

[7]G.N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed. (Cambridge U.P., 1966) p. 248.

[8]H. Weyl and R. Brauer, Am. J. Math. **57**, 425 (1935); N. Jacobson, *Lie Algebras* (Wiley, New York, 1962), p. 229.

Large coupling expansions for eigenenergies and Regge trajectories of the general even-power potential with applications

L. K. Sharma and Jyoti Choubey
Department of Applied Physics, Government Engineering College, Jabalpur-482011, India
H. J. W. Müller-Kirsten
Department of Physics, University of Kaiserslautern, 6750 Kaiserslautern, Federal Republic of Germany

(Received 25 April 1979; accepted for publication 10 July 1979)

Large coupling expansions of eigenenergies, wave functions and Regge trajectories of the generalized even power potential $V(r) = -g^2\Sigma_{j=0}^{\infty}N_{2j}r^{2j}$ are obtained. These general expansions are then used to obtain eigenenergy expansions and Regge trajectories for the anharmonic oscillator, Gauss, and similar potentials.

1. INTRODUCTION

Recently there has been a great deal of interest in large coupling solutions to various problems arising in particle physics. In field theory, however, this search has not yeilded useful information about specific problems such as nucleon structure and therefore most of those working on this model have confined themselves to studying simple models using different forms of static sources in the Hamiltonian formalism[1-4] and properties of noncompact strong coupling groups.[5-7] Attempts have also been made to explain pion–nucleon scattering[8] and nuclear forces[9] in the context of a static model. Large coupling solutions of the Bethe–Salpeter equation of scalar ϕ^3 theory have been studied by Cheng and Wu.[10] Large coupling solutions of the Bethe–Salpeter equation of the Wick–Cutkosky model have also been derived.[11]

Various difficulties in quantum field theory have led to a revival of interest in analogous aspects of nonrelativistic potential theory which is flexible enough to serve as a simple prototype of almost any kind of model theory. In elementary particle physics, nonrelativistic potential models have contributed substantially to a deeper understanding of the analytic behavior of scattering amplitudes. Several authors have investigated nonrelativistic wave equations for large coupling constants. Thus, Cheng and Wu[12] calculated the approximate behavior of Regge trajectories for the Yukawa potential. Iafrate and Mendelsohn[13] and Zauderer[14] have studied the energy eigenvalues for the Yukawa potential. Müller-Kirsten and Vahedi-Faridi[15] have investigated the Yukawa potential for large coupling constants. Other cases for which large coupling solutions have been determined are the Gauss[16] and the superposition of inverse square and Yukawa potentials.[17]

The main result of the present investigation is the derivation of the eigenenergy and Regge trajectory expansions for a general even-power potential with large coupling constant. The general even-power potential is of particular interest in potential theory since such well-known potentials as the harmonic oscillator, the Gauss potential, and anharmonic oscillator potentials (with even anharmonicities) may be derived from this as particular cases. Thus the results of this paper give a unified treatment of these potentials. In our derivation we use the perturbation technique which has been used in numerous other investigations.[11,15,16] In Sec. 2, we derive the large coupling expansion of the eigenenergy and Regge trajectories for the general even-power potential. In Sec. 3, we give applications of the general eigenenergy expansion to the harmonic oscillator, the Gauss potential, and the general anharmonic oscillator. In Sec. 4, the general expression for Regge trajectories is utilized to derive the Regge trajectories for the Gauss potential and for a potential which can be assumed to be the superposition of an inverse square and an anharmonic oscillator potential. Finally in Sec. 5, we give a brief discussion of our results.

2. EIGENENERGIES FOR THE GENERAL EVEN-POWER POTENTIAL

We consider the radial Schrödinger equation

$$\left[\frac{d^2}{dr^2} + k^2 - \frac{l(l+1)}{r^2} - V(r)\right]\psi = 0$$
$$(\hbar = c = 1, \quad m = \tfrac{1}{2}), \tag{2.1}$$

and take a potential of the form

$$V(r) = -g^2 \sum_{j=0}^{\infty} N_{2j} r^{2j}, \tag{2.2}$$

where the coefficents N_{2j} can be negative. In particular we require N_2 to be negative so that the eigenvalues to be derived below are real. It is convenient at times to express the overall coupling constant g^2 in terms of parameter λ defined by

$$g = |\lambda|e^{i\pi/2}. \tag{2.3}$$

We wish to determine the eigenenergies k^2 under normal bound state boundary conditions for large values of the coupling constant g^2.

Substituting (2.2) into (2.1) and changing the independent variable to

$$Z = (2ig\sqrt{N_2})^{1/2}r,$$

we obtain

$$\frac{d^2\psi}{dz^2} + \left[\frac{k^2 + g^2N_0}{2ig\sqrt{N_2}} - \frac{l(l+1)}{z^2} - \frac{z^2}{4}\right]\psi$$
$$= -\tfrac{1}{2}\sum_{j=2}^{\infty}\frac{N_{2j}(z^2/2)^j\psi}{(g)^{j-1}(i\sqrt{N_2})^{j+1}}. \tag{2.4}$$

We now assume that as $g \to \infty$ $(k^2 + g^2N_0)/2ig\sqrt{N_2} \sim$ finite and nonzero, i.e., of $O(g^0)$. (This ensures also that the correct result for the harmonic oscillator is reproduced.) Then the right-hand side of this equation is of $O(1/g)$.

Hence in the limit $g \to \infty$, (2.4) may be approximated by

$$\frac{d^2\psi_0}{dz^2} + \left[\frac{k^2 + g^2 N_0}{2ig\sqrt{N_2}} - \frac{l(l+1)}{z^2} - \frac{z^2}{4} \right]\psi_0 = 0. \quad (2.5)$$

Setting

$$\psi_0(z) = z^{l+1} e^{-z^2/4} X_0(z) \quad \text{and} \quad s = \tfrac{1}{2}z^2, \quad (2.6)$$

one gets

$$s\frac{d^2X_0}{ds^2} + (b-s)\frac{dX_0}{ds} - aX_0(s) = 0, \quad (2.7)$$

where

$$a = \frac{l}{2} + \frac{3}{4} - \left(\frac{k^2 + g^2N_0}{4ig\sqrt{N_2}}\right) \quad \text{and} \quad b = l + \tfrac{3}{2}. \quad (2.8)$$

A solution of Eq. (2.7) is

$$X_0(s) = \phi(a,b;s), \quad (2.9)$$

where ϕ is a confluent hypergeometric function.

The solution of Eq. (2.5)

$$\psi_0(z) = z^{l+1} e^{-z^2/4} \phi(a,b;z^2/2) \quad (2.10)$$

will be a normalizable bound state wave function, if

$$a = -n \quad \text{for} \quad n = 0,1,2,\cdots. \quad (2.11)$$

Setting $q = 4n+3$ gives $(k^2 + g^2N_0) = ig(\sqrt{N_2})(2l+q)$.

Hence in our original problem we may write

$$(k^2 + g^2N_0) = ig\sqrt{N_2}(2l+q) - 2N_2\Delta, \quad (2.12)$$

where Δ is an (as yet) undetermined expansion in descending powers of g. Inserting (2.12) into (2.4) we have

$$D_q\psi = \left[\frac{\Delta h}{2} + \frac{1}{4}\sum_{j=2}^{\infty} \frac{N_{2j}h^{j-1}(z^2/2)^j}{N_2^j}\right]\psi, \quad (2.13)$$

$$D_q = \frac{1}{2}\left(\frac{d^2}{dz^2} + l + \frac{q}{2} - \frac{l(l+1)}{z^2} - \frac{z^2}{4}\right), \quad (2.14)$$

and

$$h = \sqrt{N_2}/g. \quad (2.15)$$

As a first approximation to ψ we have (apart from an overall normalization factor)

$$\psi \approx \psi^{(0)} = \psi_q(z). \quad (2.16)$$

This approximation obviously leaves uncompensated terms on the right-hand side of (2.13) amounting to

$$R_q^{(0)} = \left[\frac{1}{2}\Delta h + \frac{1}{4}\sum_{j=2}^{\infty}\frac{N_{2j}h^{j-1}}{N_2^j}\left(\frac{1}{2}z^2\right)^j\right]\psi_q(z). \quad (2.17)$$

For convenience we set $\psi_q(z) \equiv \psi(a,b;z) \equiv \psi(a)$ and write the recurrence relation for $\psi(a)$ in the form

$$\tfrac{1}{2}z^2\psi(a) = (a,a+1)\psi(a+1) + (a,a)\psi(a) + (a,a-1)\psi(a-1), \quad (2.18)$$

where

$$(a,a+1) = a = -\tfrac{1}{4}(q-3),$$
$$(a,a) = b - 2a = l + q/2,$$

and

$$(a,a-1) = a - b = -\tfrac{1}{4}(q+3) - l. \quad (2.19)$$

By repeated application of (2.19), we obtain the following general relation:

$$(\tfrac{1}{2}z^2)^m\psi(a) = \sum_{j=-m}^{m} s_m(a,a+j)\psi(a+j), \quad (2.20)$$

where the coefficients $s_m(a,a+r)$ satisfy the recurrence relation

$$\begin{aligned} s_m(a,a+r) = {} & s_{m-1}(a,a+r-1)(a+r-1,a+r) \\ & + s_{m-1}(a,a+r)(a+r,a+r) \\ & + s_{m-1}(a,a+r+1)(a+r+1,a+r), \end{aligned} \quad (2.21)$$

with $s_0(a,a) = 1$; all $s_0(a,a+i) = 0$, for $i \neq 0$ and $s_m(a,a+r) = 0$, for $|r| > m$. The expansion $R_q^{(0)}$ may now be written

$$R_q^{(0)} = \sum_{j=0}^{\infty} h^{j+1} \sum_{k=-(j+2)}^{j+2} [a,a+K]_{j+1}\psi(a+K), \quad (2.22)$$

where $[a,a]_1 = \Delta/2 + (N_4/4N_2^2)s_2(a,a)$

and

$$[a,a+K]_{j+1} = (N_2(j+2)/4N_2^{j+2})s_{j+2}(a,a+K), \quad (2.23)$$

for j and K not zero simultaneously.

We now observe that $D_q\psi(a+K) = K\psi(a+K)$, so that a term $\mu\psi(a+K)$ may be removed by adding to $\psi^{(0)}$ the contribution $(\mu/K)\psi(a+K)$ except, of course, when $K=0$.

Hence the next contribution to $\psi^{(0)}$ becomes

$$\psi^{(1)} = \sum_{j=0}^{\infty} h^{j+1} \sum_{\substack{K=-(j+2)\\ \neq 0}}^{j+2} \frac{[a,a+K]_{j+1}}{K}\psi(a+K). \quad (2.24)$$

This contribution leaves uncompensated a sum of terms $R_q^{(1)}$ which again lead to another contribution $\psi^{(2)}$. Repeating this process successively and adding these contribution to $\psi^{(0)}$ we obtain

$$\psi = \psi^{(1)} + \psi^{(2)} + \psi^{(3)} + \cdots. \quad (2.25)$$

However, (2.25) will be a solution of our equation only if the sum of all terms containing $\psi(a)$ in $R_q^{(0)}, R_q^{(1)}, \ldots$, left uncompensated so far, is set equal to zero. Thus,

$$\begin{aligned} 0 = {} & h[a,a]_1 + h^2\Big\{[a,a]_2 + [a,a-2]_1\frac{[a-2,a]_1}{-2} \\ & + [a,a-1]_1\frac{[a-1,a]_1}{-1} + \frac{[a,a+1]_1}{1}[a+1,a]_1\Big\} \\ & + O(h^3). \end{aligned} \quad (2.26)$$

The expansion (2.25) is then an eigensolution and (2.26) the appropriate secular equation which enables us to calculate Δ and hence the eigenenergy k^2. Explicit calculation of terms up to $O(h^3)$ yields the following expression for the eigenenergies for large coupling constant g^2:

$$(k^2+g^2N_0)=ig\sqrt{N_2}(2l+q)+\frac{N_4}{2^3N_2}[3(q^2+1)+4(3q-1)l+8l^2]$$
$$+\frac{N_6}{2^4N_2^{3/2}(ig)}[5q(q^2+5)+2(15q^2-6q+25)l+24(2q-1)l^2+16l^3]$$
$$-\frac{N_4}{2^6N_2^{5/2}(ig)}[q(17q^2+67)+2(51q^2-18q+67)l+24(7q-3)l^2+64l^3]$$
$$-\frac{N_8}{2^7N_2^2g^2}[(35q^4+490q^2+315)+280lq^3-384l^3+128l^4]$$
$$+\frac{N_4N_6}{2^8N_2^3g^2}[(165q^4+1770q^2+945)+1320lq^3+504lq^2+7080lq-1368l$$
$$+3456l^2q^2-2016l^2q+5856l^2+3264l^3q+1928l^3+768l^4]$$
$$-\frac{N_4^3}{2^{16}g^2N_2^4}[(24000q^4+218496q^2+9896)+2078lq^3-66048lq^2-873984lq$$
$$-139776l+56995l^2q^2-264192l^2q+745472l^2+504108l^3q-241664l^3+131072l^4]+O(1/g^3). \quad (2.27)$$

In terms of $|\lambda|$ Eq. (2.27) can be written [this expression is only written up to $O(1/|\lambda|)$]:

$$k^2-|\lambda^2|N_0=-|\lambda|\sqrt{N_2}(2l+q)+(N_4/2^3N_2)\{3(q^2+1)+4(3q-1)l+8l^2\}+O(1/|\lambda|). \quad (2.28)$$

The Regge trajectories follow with the help of (2.8) and (2.11), and we obtain

$$\alpha_n(k)=l_n(k)=-2n-\frac{3}{2}-\frac{P}{2}-\frac{N_4}{2^3\lambda N_2^{3/2}}(-8n^2-8n-1-P^2-2P-24nP)$$
$$+\frac{N_6}{2^4\lambda^2N_2^2}(960n^3-120n^2-140n-P^3+6P^2-8P+12P^2n+24Pn+24Pn^2)$$
$$+\frac{N_4^2}{2^6N_2^3\lambda^2}(1088n^3+792n^2+648n-4P^3+18P^2-27P+378+36P^2n+522Pn-24Pn^2)+O(P^4), \quad (2.29)$$

where

$$P=(k^2-\lambda^2N_0)/\lambda\sqrt{N_2}. \quad (2.30)$$

3. APPLICATIONS OF THE GENERAL EIGENENERGY EXPANSION

We now apply the eigenenergy expansion (2.28) to three cases.

A. Harmonic oscillator

The harmonic oscillator is given as

$$V(r)=\alpha^2r^2, \quad (3.1)$$

Implying

$$N_0=0,\quad N_{2j}=0\quad \text{for}\quad j\geqslant 2, \quad (3.2)$$

and so

$$k^2=ig\sqrt{N_2}(2l+q)=-\lambda\sqrt{N_2}(2l+q), \quad (3.3)$$

which is the well-known result.

B. Gauss potential

The Gauss potential is given by

$$V(r)=-g^2e^{-\alpha^2r^2}, \quad (3.4)$$

so that

$$N_{2j}=(-1)^j\alpha^{2j}/j!, \quad (3.5)$$

and we obtain

$$(k^2+g^2)$$
$$=g\alpha(2l+q)-\frac{\alpha^4}{2^4}[3(q^2+1)+4(3q-1)l+8l^2]$$
$$-\frac{\alpha^3}{3\times2^8\times g}[q(11q^2+1)+2(33q^2$$
$$-6q+1)l+24(5q-1)l^2+64l^3]$$
$$+\frac{\alpha^4}{3\times2^{15}\times g^2}[4(85q^4+2q^2-423)$$
$$+l(2720q^3-71q^2+32q+2976)$$
$$+32l^2(252q^2-12q+64)$$
$$+256l^3(41q-9)+4096l^4]+O(1/g^3). \quad (3.6)$$

Expansion (3.6) is identical with the expression derived previously.[16]

C. Generalized anharmonic oscillator

Next we consider the application of (2.27) in evaluating the eigenenergies for the generalized anharmonic oscillator with anharmonicities of the form f, where $m=2,3$, and 4, and the following Hamiltonian

$$H=\left(-\frac{d^2}{dr^2}+\frac{1}{4}r^2+\frac{1}{4}fr^{2m}\right). \quad (3.7)$$

Evaluating the eigenenergies of this oscillator for $m = 2$ (r^4 anharmonicity), we make the following substitutions in (2.27):

$$N_0 = 0, \quad -g^2N_2 = \tfrac{1}{4}, \quad -g^2N_4 = \tfrac{1}{4}f,$$
$$N_6 = N_8 = 0, \quad l = 0. \qquad (3.8)$$

For consistency of these equations under the condition that N_2 is in general not simply related to N_4 we set $g = 1$. For the eigenenergies, we then obtain

$$k^2 = (2n + \tfrac{3}{2}) + 3f(2n^2 + 3n + \tfrac{5}{4}) - f^2[34n^3 + 153n^2 + (263n/4) + (165/8)] + O(f^3). \qquad (3.9)$$

The eigenenergies for $m = 3$ (r^6 anharmonicity) and $m = 4$ (r^8 anharmonicity) can be evaluated similarly. Bender and Wu,[18] while discussing the anharmonic oscillator problem defined by the differential equation

$$\left(-\frac{d^2}{dr^2} + \frac{1}{4}r^2 + \frac{1}{4}\lambda r^4\right)\phi(r) = \underset{\substack{\lim\phi(r) = 0 \\ r\to\pm\infty}}{E(\lambda)}\ \phi(r), \qquad (3.10)$$

observed that the perturbation series for the ground state energy $E_0(f)$ has the form

$$E_0(f) = \tfrac{1}{2}m + \sum_{n=1}^{\infty} mA_n\left(\frac{f}{m^3}\right)^n. \qquad (3.11)$$

The coefficients A_n calculated by Bender and Wu[18] agrees with our calculations.

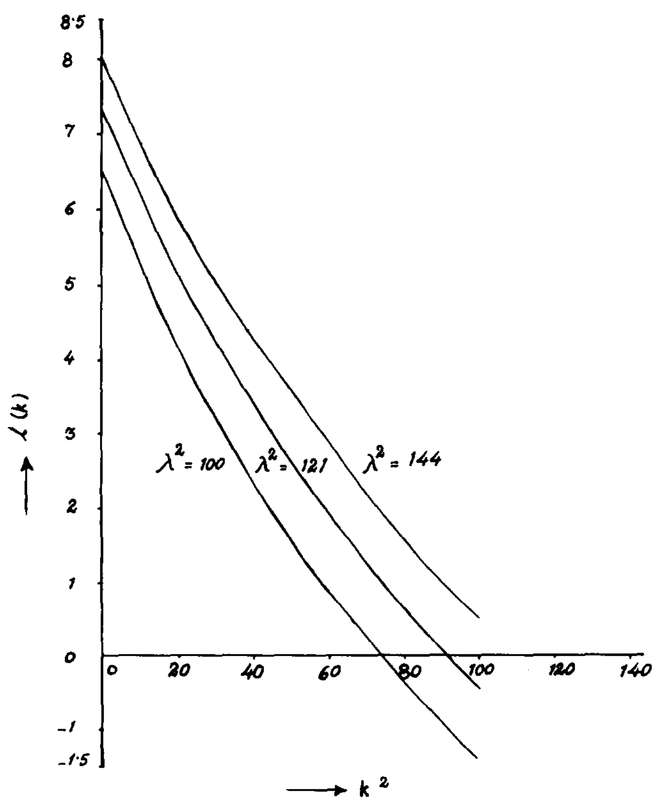

FIG. 1. Regge trajectories for the even power ($N_{2j} = 1$) potential in the ground state for different values of coupling constant λ^2.

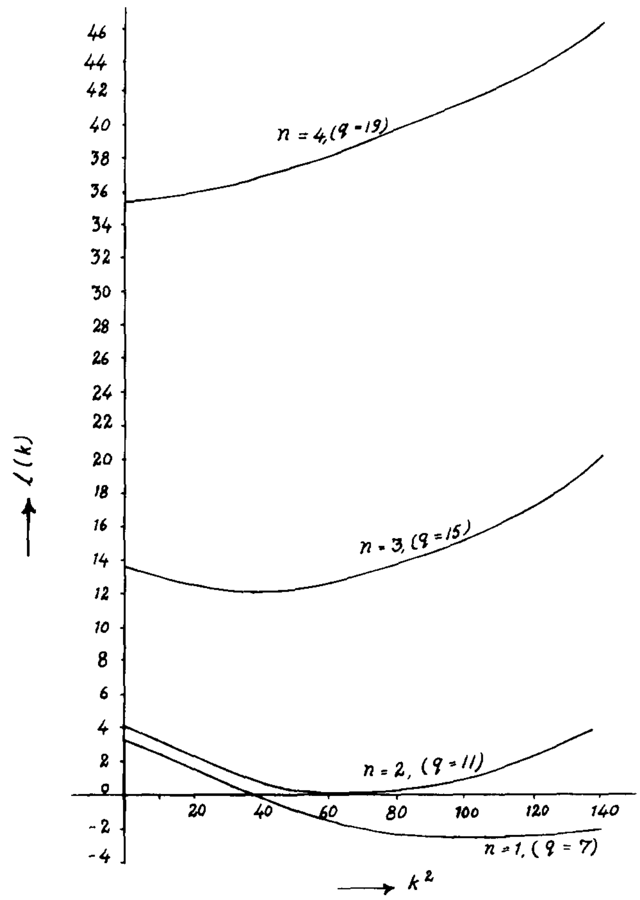

FIG. 2. Regge trajectories for the even power potential ($N_{2j} = 1$) for different values of n.

4. REGGE TRAJECTORIES

We consider the following examples.

A. Harmonic oscillator

In this case we have

$$\alpha_n = l_n(k) = -2n - \tfrac{3}{2} - k^2/2ig\sqrt{N_2} \qquad (4.1)$$

as is well known.

B. Gauss potential

The expansion for the ground state has the form

$$l_0(k) = -\frac{3}{2} + \frac{k^2+g^2}{2\alpha g} - \frac{\alpha}{2^3 g}\left\{-\frac{1}{2} + \frac{k^2+g^2}{\alpha g} - \frac{(k^2+g^2)^2}{2\alpha^2 g^2}\right\} + O(\alpha^2). \qquad (4.2)$$

Aly *et al.*[19] have calculated the expression for the Regge trajectories for this potential [$V(r) = \mathfrak{G}e^{-\alpha^2 r^2}$] up to $O(k^4)$. By setting $g = \mathfrak{G}^{1/2}/i$ in expansion (4.2), we now get the following series for the Regge trajectories

$$l_0(k) = -\frac{3}{2} + \frac{(k^2-\mathfrak{G})i}{2\alpha\sqrt{\mathfrak{G}}} - \frac{1}{8\mathfrak{G}}(k^2-\mathfrak{G}) - \frac{i}{16\mathfrak{G}^{3/2}}\left[-\frac{4}{3}(k^2-\mathfrak{G})\alpha + (k^2-\mathfrak{G})^2\frac{1}{\alpha}\right]$$

$$+\cdots+\frac{3i}{2^4\times 10^2}\left[\frac{(k^2-\circledS)^4}{\alpha}\right]+\cdots. \quad (4.3)$$

This expression is an improvement over that derived by Aly *et al.*[19] (Since it contains terms up to those of $O(k^8)$.

C. Superposition of inverse square and anharmonic oscillator potentials

Finally we derive the Regge trajectories for the potential described by the following Hamiltonian

$$H=\left(-\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2}+r^2+\lambda r^{2m}\right), \quad (4.4)$$

where $m = 2,3,4$, etc. Thus with the value of $m = 2$ (r^4 anharmonicity) the Regge trajectories (for the ground state) are given by:

$$l_0(k)=-\frac{3}{2}+\frac{k^2}{2}+\lambda\left(-\frac{1}{2^3}-\frac{k^4}{2^3}+\frac{k^2}{2^2}\right)$$
$$+\lambda^2\left(\frac{k^6}{2^4}+\frac{3^2k^4}{2^5}+\frac{3^3k^2}{2^6}\right). \quad (4.5)$$

Similarly for $m = 3$ (r^6 anharmonicity) and $m = 4$ (r^8 anharmonicity), we can write the following expressions for the Regge trajectories (for the ground state)

$$l_0(k)=-\frac{3}{2}+\frac{k^2}{2}+\frac{\lambda}{2^3}(4k^2+3k^4)$$
(for r^6 anharmonicity), (4.6)

$$l_0(k)=-\frac{3}{2}+\frac{k^2}{2}+\frac{\lambda}{2}(-10+8k^2+3k^4)$$
(for r^8 anharmonicity). (4.7)

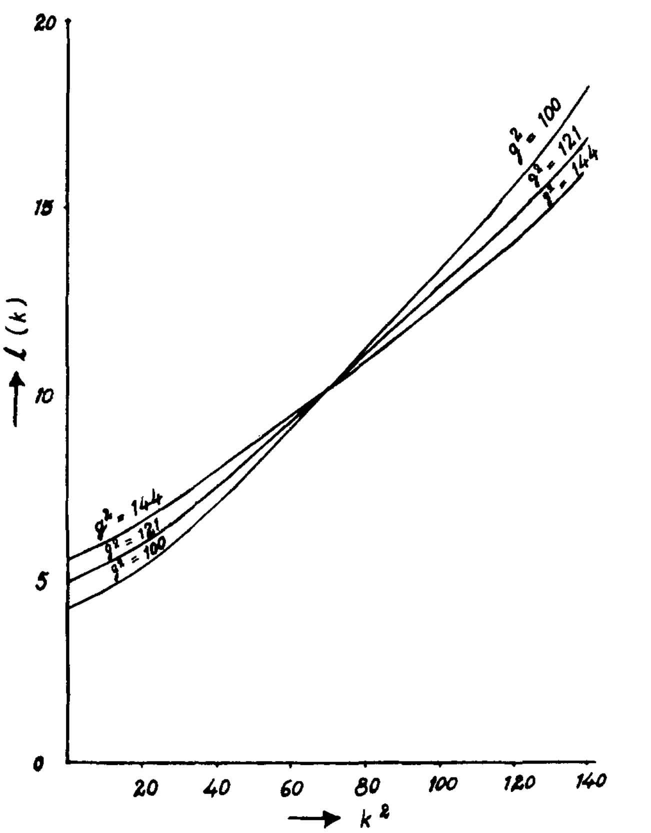

FIG. 3. Ground state Regge trajectories for Re(l) for the Gauss potential for different values of coupling parameter g^2.

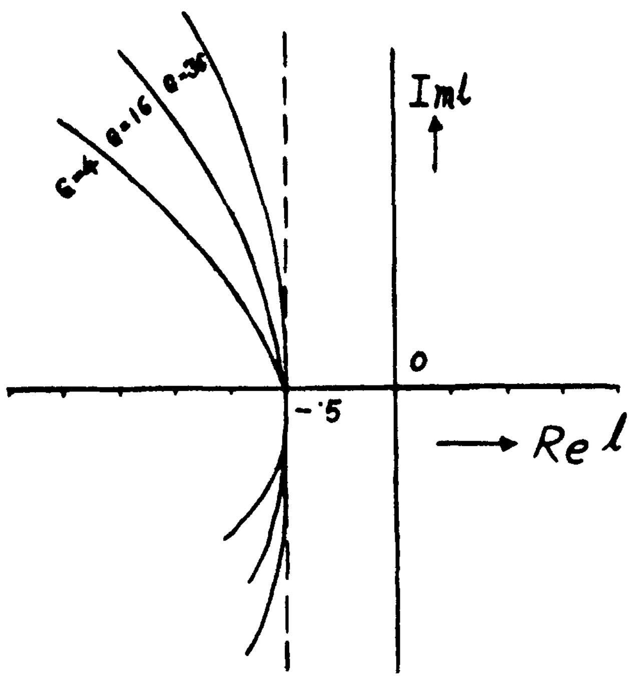

FIG. 4. Ground state Regge trajectories for complex l values for the Gauss potential for various values of g^2.

Graphical representation of Regge trajectories:

(1) In Fig. 1, we show the ground state Regge trajectories for the even-power potential with all coefficients $N_{2j} = 1$ in (2.2) for different values of the coupling constant λ^2.

(2) In Fig. 2, these trajectories for the potential which contains only even powers of r for different values of n viz: $n = 1$ ($q = 7$), $n = 2$ ($q = 11$) $n = 3$ ($q = 15$) and $n = 4$ ($q = 19$) have been plotted.

(3) Ground state Regge trajectories for the Gauss potential for different values of g^2 have been displayed in Fig. 3.

(4) Regge trajectories for complex l values for the Gauss potential are plotted in Fig. 4. This work has also been done by Aly *et al.*[19]; our graph, however, has been drawn with the improved approximation obtained in Eq. (4.3).

(5) Figure 5 depicts the behavior of Regge trajectories for different anharmonicities (r^4, r^6, and r^8) for the potential which is the superposition of inverse square and anharmonic oscillator potential in the ground state for Re(l).

5. DISCUSSION

(A) It is well known that the linear potential gives a rather rapid rise of Regge trajectories, while the oscillator potential yields linearly rising trajectories. For a Yukawa potential the trajectoreis do not rise but fall off rapidly with increasing energy. This behavior also persists in the strong coupling domain. It has been conjectured that a Coulomb/Yukawa-like force is responsible for the power law fall-off of form factors and vertex functions.[20] It is therefore of interest to understand the wave functions and Regge trajectories when the potential contains an anharmonic as well as Yukawa-like part. However, for such a potential the wave equation is, in general, difficult to solve except in two cases:

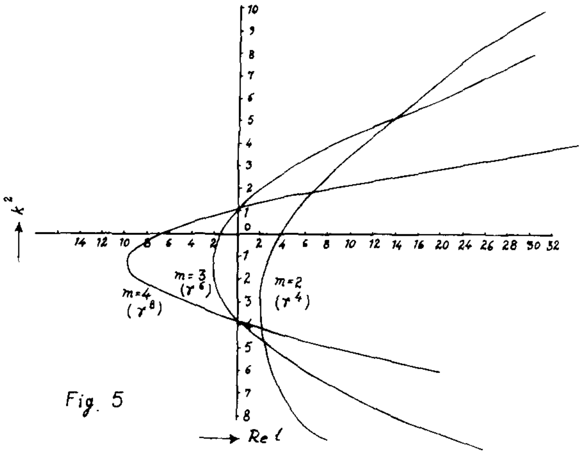

FIG. 5. Ground state Regge trajectories for Re(l) for the superposition of inverse square and anharmonic oscillator for different anharmonicities of the oscillator.

1. the case of the Coulomb potential perturbed by anharmonic contributions, and 2. the case of the oscillator (quark confining) potential perturbed by Coulomb or Yukawa contributions.

Other types of potentials have been discussed in this context; one that has attracted considerable interest recently is a potential with a finite range singularity i.e., a potential $V_\lambda(r)$ which is singular at a point $r=\lambda$ where $\lambda \neq 0$ or ∞.

The potential of this type discussed by Filippov[21] has the form

$$V_\lambda(r) = -\mathfrak{G}^2/r^2 - \lambda^2, \tag{5.1}$$

with the property that $\lim_{r\to\infty} V_\lambda(r) = 0$.

For $r \neq \lambda$ one can write

$$V_\lambda(r) = \frac{\mathfrak{G}^2}{\lambda^2}\left\{1 + \frac{r^2}{\lambda^2} + \left(\frac{r^2}{\lambda^2}\right)^2 + \cdots\right\}. \tag{5.2}$$

Thus for $r > \lambda$ potential (5.1) behaves as a modified harmonic oscillator, and one can easily obtain its eigenvalues and Regge trajectories from Eqs. (2.27) and (2.29). The Regge trajectories for the potential are given by

$$l_n(k) = -2n - \frac{3}{2} + \frac{(k^2\lambda^2 - \mathfrak{G}^2)}{2\mathfrak{G}} + \frac{1}{\mathfrak{G}}\left\{ -n^2 - n - \frac{1}{2^3} - \frac{(k^2\lambda^2 - \mathfrak{G}^2)^2}{2\mathfrak{G}} + 3n\,\frac{(k^2\lambda^2 - \mathfrak{G}^2)}{\mathfrak{G}^2}\right\} + O\left(\frac{1}{\mathfrak{G}^2}\right) \tag{5.3}$$

and demonstrate their rise with k^2. However, for λ approaching zero the linear rise of trajectory is lost, the potential therby reducing to that of the centrifugal type.

(B) We also observe from Fig. 2, that for the general even-power potential (with $N_{2j} = 1$), the spacing between two successive Regge trajectories increases with n or q.

ACKNOWLEDGMENT

One of us (J.C.) acknowledges with thanks the University Grants Commission, India, for the award of a Junior fellowship.

[1] W. Pauli and S.M. Dancoff, Phys. Rev. **62**, 85 (1942).
[2] G. Wentzel, Helv. Phys. Acta **13**, 269 (1940); **14**, 633 (1941).
[3] G. Dullemond, Ann. Phys. N.Y. **33**, 214 (1965); F.J.M. Von der Linden and C. Dullemond, Nuovo Cimento A **63**, 615 (1966); C. Dullemond and F.J.M. Von der Linden, Ann. Phys. N.Y. **41**, 372 (1967).
[4] T.W. Ruijgrok, Nucl. Phys. B **8**, 591 (1968).
[5] C.J. Goebel, Phys. Rev. Lett. **16**, 1130 (1966); C.J. Goebel, *Noncompact Groups in Particle Physics*, edited by Y. Chow (Benjamin, New York, 1967); T. Cook, C.J. Goebel, and B. Sakita, Phys. Rev. Lett. **15**, 35 (1965); T.Cook and B. Sakita, J. Math. Phys. **8**, 708 (1967); B. Sakita, Phys. Rev. **170**, 1453 (1968).
[6] J.G. Kuriyan and E.C.G. Sudarshan, Phys. Rev. **162**, 1650 (1967).
[7] N.G. Deshpande, Phys. Rev. **163**, 1629 (1967).
[8] R.G. Sachs, Phys. Rev. **95**, 1065 (1954); H. John, *ibid*, **126**, 824 (1962).
[9] S. Wagner, Phys. Rev. **177**, 2278 (1969).
[10] H. Cheng and T.T. Wu, Phys. Rev. D **5**, 3192 (1972).
[11] H.J.W. Müller-Kirsten, Nuovo Cimento A **10**, 76 (1972).
[12] H. Cheng and T.T. Wu, Phys. Rev. D **5**, 3189 (1972).
[13] G.J. Iafrate and L.B. Mendelsohn, (a) Phys. Rev. **182**, 244 (1969); (b) Phys. Rev. A **2**, 561 (1970).
[14] E. Zauderer, J. Math. Phys. **13**, 42 (1972).
[15] H.J.W. Müller-Kirsten and N. Vahedi-Faridi, J. Math. Phys. **14**, 1291 (1973).
[16] H.J.W. Müller-Kirsten, J. Math. Phys. **11**, 355 (1970).
[17] S.S. Raghuwanshi and L.K. Sharma, Indian. J. Phys. **19**, (1978).
[18] C.M. Bender and T.T. Wu, Phys. Rev. **184**, 1231 (1969).
[19] H.H. Aly, H.J.W. Müller-Kirsten, and K. Schilcher, Nucl. Phys. B **3**, 401 (1967).
[20] H.J.W. Müller-Kirsten, Phys. Rev. D **12**, 1103 (1975).
[21] A.T. Filippov, Dubna preprints E2-6936 (1973), E2-7563 (1974), E2-7929 (1974).

Path-dependent quantum formulation of electromagnetism with magnetic charges

R. Gambini [a] and A. Trias
Departamento de Física, Universidad Simón Bolívar, Apartado Postal:80659, Caracas 108, Venezuela

(Received 20 August 1979; accepted for publication 30 November 1979)

A path-dependent quantum formulation of electromagnetism with magnetic charges is proposed. One starts from a path-dependent nonsingular potential which generalizes the Wu–Yang potential. Then a Lagrangian formulation is proposed manifestly selfdual and without singular terms or kinematical constraints. The quantization of this Lagrangian leads to a linear space of path-dependent wavefunctions where the dynamical equations are stated. The angular momentum usually considered in the field of a static monopole is identified with the generator of spatial rotations in the space of the path-dependent wavefunctions.

1. INTRODUCTION

Path-dependent objects were first considered by Mandelstam[1] in ordinary electromagnetism with the introduction of the path-dependent field operator:

$$\Psi(x,P(x)) = \psi(x) \exp\left\{ie\int_{P(x)}^{x} dx'^{\lambda}A_{\lambda}(x')\right\}, \qquad (1.1)$$

where $\psi(x)$ is the ordinary matter field which under gauge transformations:

$$A_{\mu}(x) \rightarrow A_{\mu}(x) + \Lambda_{,\mu}(x) \qquad (1.2)$$

transforms according to:

$$\psi(x) \rightarrow \psi(x) \exp\{-ie\Lambda(x)\}. \qquad (1.3)$$

Obviously $\Psi(x,P(x))$ is a gauge-independent operator. The gauge dependence of the ordinary matter field has been changed into path dependence of the Mandelstam operator.

In first quantization $\Psi(x,P(x))$ is a path-dependent wavefunction. It is then a natural object to consider since both its modulus and its phase may be considered as physical. Moreover, by introducing the Mandelstam covariant derivative:

$$D_{\mu}\Psi(x,P(x)) = \lim_{\epsilon\rightarrow 0}\frac{\Psi(x+\epsilon,P_E(x+\epsilon)) - \Psi(x,P(x))}{\epsilon}, \qquad (1.4)$$

where ϵ is an infinitesimal displacement along the "μ" direction and $P_E(x+\epsilon)$ is the path $P(x)$ with a rectilinear extension to reach the point $x+\epsilon$, the quantum equations for $\psi(x)$ coupled to the electromagnetic field may be rewritten as formally free equations with D derivatives for the path-dependent wavefunction.

On the other hand, as most clearly remarked by Wu and Yang,[2] the Aharonov–Bohm experiment may be considered as the experimental detection of the phase factor

$$\exp\left\{ie\oint dx'^{\lambda}A_{\lambda}(x')\right\}, \qquad (1.5)$$

which in turn may be considered as the relative phase in a linear superposition of two Mandelstam path-dependent wavefunctions. It turns out that a linear space of path-dependent wavefunctions could be the natural framework for a quantum description of general electromagnetism.

This is the point of view advocated in the present paper where path dependence is recognized as the key element to have a nonsingular description of electromagnetism in the presence of magnetic poles. One starts in Sec. 2 by introducing a path-dependent potential which furnishes a globally nonsingular description of the electromagnetic field. The potential has a natural relationship with the Wu–Yang[2] potential of which it may be thought as the limit to a continuous number of regions. Stated in more geometrical terms, the Wu–Yang potential furnishes a connection in a principal fiber bundle over space–time. This structure may be used to induce a principal fiber bundle over the manifold of paths of space–time. The induced bundle is globally trivial. This is the geometrical explanation of the "free looking" aspect of the dynamical equations expressed in terms of path-dependent wavefunctions and the geometrical motivation of the whole path-dependent approach.

In Sec. 3 the path-dependent potential is used to discuss a Lagrangian formulation for a general system of point charges in electromagnetic interaction. This formulation is free from anomalous singularities and kinematical constraints. The complete set of the Lorentz–Maxwell equations is obtained as stationarity conditions of the action which is manifestly covariant and selfdual.

In Sec. 4, the quantization of this Lagrangian is considered. One obtains a quantum formulation in a linear space of path-dependent wavefunctions, which reduces to the Mandelstam formulation in the case of ordinary electromagnetism.

Finally, in Sec. 5 the general properties of the angular momentum of the formulation are considered. The angular momentum of Wu–Yang[3] in the case of a static monopole is recovered now as the local version of the generator of spatial rotations in the path-dependent linear space.

2. THE PATH-DEPENDENT POTENTIAL

Let us consider some field strength $F_{\mu\nu}(x)$ verifying the Bianchi identities

$$F_{\mu\nu,\lambda}(x) + F_{\nu\lambda,\mu}(x) + F_{\lambda\mu,\nu}(x) = g_{\mu\nu\lambda}(x), \qquad (2.1)$$

where $g_{\mu\nu\lambda}(x)$ is the dual of the magnetic current. In ordinary electromagnetism $g\equiv 0$ and these identities guarantee

[a] This research was supported in part by the Consejo Nacional de Investigaciones Cientificas y Tecnológicas de Venezuela.

the existence of the familiar representation

$$F_{\mu\nu}(x) = A_{\nu,\mu}(x) - A_{\mu,\nu}(x). \quad (2.2)$$

If one tries, following Dirac,[4,5] to maintain this decomposition when magnetic charges are present, the potentials exhibit string singularities. The position of the strings is completely arbitrary thus suggesting the existence of new degrees of freedom in the potential. Other formulations,[6] free from anomalous singularities, need also to consider generalized objects in the process of quantization.[7]

In this paper one introduces path-dependent potentials as the general and natural way to process the information contained in the Bianchi identities. For this purpose let us define the path-dependent object

$$\mathbb{K}_\mu(x,P(x)) = \int_{P(x)}^{x} dx'^\lambda F_{\lambda\mu}(x'), \quad (2.3)$$

where $P(x)$ goes from spatial infinity, where $F_{\mu\nu}(x)$ is assumed to vanish, to the point x.

Let us also introduce the differential operator, which we shall refer hereafter as the parallel derivative, given for an arbitrary path-dependent functional by

$$\partial_{:\mu} G(x,P(x)) = \lim_{\epsilon\to 0} \frac{G(x+\epsilon,P_:(x+\epsilon)) - G(x,P(x))}{\epsilon}, \quad (2.4)$$

where $P_:(x+\epsilon)$ is obtained by parallel transporting $P(x)$ from x to $x+\epsilon$ and ϵ is some infinitesimal displacement along the "μ" direction.

Using this definition in (2.3) one immediately obtains

$$\mathbb{K}_{\mu:\nu}(x,P(x)) = \int_{P(x)}^{x} dx'^\lambda F_{\lambda\mu,\nu}(x'). \quad (2.5)$$

Hence we may integrate (2.1) along some path $P(x)$ to obtain the representation

$$F_{\mu\nu}(x) = \mathbb{K}_{\nu:\mu}(x,P(x)) - \mathbb{K}_{\mu:\nu}(x,P(x)) + \int_{P(x)}^{x} dx'^\lambda g_{\mu\nu\lambda}(x'). \quad (2.6)$$

This decomposition of the point function field strength in terms of path-dependent potentials is in fact the coordinate version of a natural geometric construction which represents nonclosed k forms in a manifold by $k-1$ forms in the manifold of paths of the original manifold. Then (2.6) arises as a very general homotopy associated with the construction.

The path-dependent potential is essentially a nonsingular object. According to its definition (2.3), the potential is singular for paths going through the sources of the electromagnetic field. But these singularities are avoidable by considering appropriate limits on small deformations of the path. Only when the end point is a source itself the singularity becomes unavoidable. But these singularities should be associated with the ordinary source-induced point singularities of ordinary electromagnetism rather than anomalous string singularities present in the point-function potentials of electromagnetism with magnetic charges.

Ordinary point-function potentials may be obtained from the path-dependent potential by selecting a fixed path for every point x. For instance let us consider some fixed reference path C going from spatial infinity to the origin of coordinates. One may then select for every point x the path $C(x)$ obtained from C by parallel transport from the origin of coordinates to the point x. The point function object which arises from this path specialization will be denoted as

$$A_\mu(x) = \mathbb{K}_\mu(x,P(x))\big|_{P(x)=C},$$

or simply

$$A_\mu(x) = \mathbb{K}_\mu(x,C). \quad (2.7)$$

It is an immediate consequence of the definition given in (2.4) that

$$\partial_{:\mu}\mathbb{K}_\nu(x,P(x))\big|_{P(x)=C} = A_{\nu,\mu}(x) \quad (2.8)$$

and then it follows from (2.6) that $\mathbb{K}_\mu(x,C)$ is an ordinary potential with its string singularities related to C in an obvious geometrical way. For instance, in the case of a static monopole at the origin of coordinates, one may obtain from (2.8) the Dirac's potential with a generic string[8]:

$$\mathbf{A}(\mathbf{x}) = -\int_{C'} d\mathbf{a}\times\mathbf{B}(\mathbf{x}-\mathbf{a}), \quad (2.9)$$

where the string C' is symmetric to the reference path C by rapport to the origin. It is evident from this example the role played by the path dependence of the potential in avoiding the anomalous singularities. Once the degrees of freedom associated with the path are frozen, by selecting fixed paths for every point, the string singularity becomes unavoidable.

Since parallel derivatives are obviously commutative one may generalize (2.3) to

$$\mathbb{K}(x,P(x)) = \int_{P(x)}^{x} dx'^\lambda F_{\lambda\mu}(x') + \Lambda_{:\mu}(x,P(x)), \quad (2.10)$$

without changing the decomposition (2.6). Thus the gauge freedom of this path-dependent formulation may be stated in complete analogy with the ordinary case. The path-dependent formulation is related in a very natural way to the Wu–Yang[2] formulation of gauge theories. To see this, let us consider some finite covering of space–time by open sets $D_1,D_2,\ldots,D_N$ with nonvanishing intersections D_{ij}. Let us assign a reference path C_i to each region D_i and then let us fix parallel paths $C_i(x)$ for each point in D_i. Within a region D_i one must then have the usual representation of the field strength in terms of the ordinary point-function potential $\mathbb{K}_\mu(x,C_i)$. For x belonging to some intersection region D_{ij} the potential is doubly defined but one immediately sees that the two potentials are related by a gauge transformation. In fact, a simple calculation using the Bianchi identities and Stokes theorem shows that for x in D_{ij} one has

$$\mathbb{K}_\mu(x,C_i) - \mathbb{K}_\mu(x,C_j) = \partial_\mu\left\{\tfrac{1}{2}\int_{\Sigma_{ij}}^{x} d\sigma^{\nu\lambda} F_{\nu\lambda}\right\}, \quad (2.11)$$

where Σ_{ij} is some two-dimensional surface drawn from $C_j(x)$ to $C_i(x)$. A change of this surface will change the integral in (2.11) by the amount of the magnetic charges enclosed in the variation. Hence the potentials in adjacent regions are related in the intersection by the gradient of a multiply-valued function and we recognize all the elements of an ordinary Wu-Yang formulation.

Since the reference paths C_i may be changed at will, the whole Wu–Yang construction is equivalent to a path-dependent potential which only exhibits the degrees of freedom associated with the path when going from region to region in some covering of space–time. One may then consider more refined coverings and in the limit where the regions reduce to points the Wu–Yang potential reduces to the path-dependent potential as defined in (2.10).

3. LAGRANGIAN FORMULATION

Let us now consider a general system of classical particles in electromagnetic interaction. There is a number of well-known Lagrangian formulations for this problem. Some of them[5,9] contain highly singular string–string interaction terms associated with the use of Dirac potentials. This problem will be avoided in the present formulation as a result of the nonsingular nature of the path-dependent formulation. Other formulations are not manifestly selfdual. The particles may have a single type of charge[5,10] or part of the Maxwell equations is imposed externally as kinematical constraints.[10] In the present formulation the selfduality of the action is strongly emphasized. Complex notation is adopted throughout which allows one to assign completely symmetric roles to the electric and magnetic charges. Moreover the structure equation for the complex field strength

$$\mathbb{F}_{\mu\nu}(x) = F_{\mu\nu}(x) + i\bar{F}_{\mu\nu}(x), \tag{3.1}$$

is assumed only as a boundary condition and then obtained as part of the dynamical equations.

Dirac's[5] well-known veto of Wentzel's[11] weaker condition will play now the role of appropriate restrictions on the possible choices of the gauge.

The action for a system of particles which may be both electrically and magnetically charged is given by

$$\begin{aligned} S &= RE\Sigma,\\ \Sigma &= \int_{\sigma_1}^{\sigma_2} d^4x \,\{ \tfrac{1}{8}\mathbb{F}^{\mu\nu}(x)\mathbb{F}^*_{\mu\nu}(x)\\ &\quad + \tfrac{1}{4}\,\mathbb{F}^{*\mu\nu}(x)\Big[\mathbb{K}_{\mu:\nu}(x,P(x))\\ &\quad - \mathbb{K}_{\nu:\mu}(x,P(x)) - \int_{P(x)}^{x} dx'^{\lambda}\, h_{\mu\nu\lambda}(x')\Big]\\ &\quad + \tfrac{1}{2}\bar{\mathbb{H}}^{\mu\nu}(x)\mathbb{B}^*_{\mu,\nu}(x) - (i/2)\mathbb{B}^*_{\mu}(x)\, J_{\mu}(x)\}\\ &\quad + \tfrac{1}{4}\sum_i G_i^* \int_{1\Sigma_i}^{2} d\sigma^{\mu\nu}\,(A_{\nu,\mu} - A_{\mu,\nu} + \mathbb{H}_{\mu\nu})\\ &\quad + \sum_i m_i \int_1^2 ds\,\{\dot z_\mu^{(i)}(s)\dot z^{\mu(i)}(s)\}^{1/2}. \end{aligned} \tag{3.2}$$

The action integral is taken in a 4-dimensional region bounded by the two spatial surfaces σ_1 and σ_2.

$\mathbb{F}_{\mu\nu}(x)$ is a point function antisymmetric complex Lorentz tensor. The selfduality condition (3.1) is assumed to hold only as a boundary condition in the surfaces σ_1 and σ_2.

$\mathbb{K}_\mu(x,P(x))$ is a complex path-dependent 4-vector. Its general structure is assumed to be

$$\mathbb{K}_\mu(x,P(x)) = \int_{P(x)}^{x} dx'^{\lambda}\, \mathbb{H}_{\lambda\mu}(x') + A_\mu(x) + \Lambda_{:\mu}(x,P(x)), \tag{3.3}$$

where $\mathbb{H}_{\mu\nu}(x)$ is a complex antisymmetric Lorentz tensor, $A_\mu(x)$ is a complex point function Lorentz vector and $\Lambda(x,P(x))$ is an arbitrary path-dependent scalar. $h_{\mu\nu\lambda}(x)$ is a notation for

$$h_{\mu\nu\lambda}(x) \equiv \mathbb{H}_{\mu\nu,\lambda}(x) + \mathbb{H}_{\nu\lambda,\mu}(x) + \mathbb{H}_{\lambda\mu,\nu}(x). \tag{3.4}$$

The structure of the path-dependent potential has been chosen to guarantee the path independence for the free part of the action which is a rather obvious requirement since path dependence is only relevant for a interacting theory.

$\mathbb{B}_\mu(x)$ is a complex point function Lorentz vector which will be identified later on as a potential *à la* Dirac.

$J^\mu(x)$ is the complex electromagnetic current of the system. The ith particle contributes to this current by

$$J_\mu^{(i)}(x) = G_i \int ds\, \dot z_\mu^{(i)}(s)\delta^4(x - z^{(i)}(s)), \tag{3.5}$$

where G_i is the complex charge of the particle:

$$G_i = e_i + ig_i. \tag{3.6}$$

The next-to-last term in (3.2) is expressed in terms of two-dimensional surface integrals. Each surface Σ_i is made up with the worldline Π_i of the ith particle and a selection of paths $P(x)$ for each point x on Π_i. All the paths are assumed to have their origin in the same point at spatial infinity to avoid undesirable contributions in the asymptotic region due to the radial nature of the magnetic field of magnetic poles. It is also required that the worldlines of the particles do not cross. This is a well-known[12,10] condition for the proper definition of the action. It may be seen that this term is the continuous limit of the interaction part of the Wu–Yang[10] action.

The last term in (3.2) is the ordinary real free action for a system of material particles of mass m_i.

The dynamical variables are defined to be: the complex Lorentz tensor $\mathbb{F}_{\mu\nu}(x)$, the path dependent potential $\mathbb{K}_\mu(x,P(x))$, the point function Dirac's potential $\mathbb{B}_\mu(x)$ together with their complex conjugates and the surfaces Σ_i which embody the information of the worldlines Π_i and the path dependence of the action. It is interesting to realize that according to the definition (3.3) the path-dependence potential contains independent degrees of freedom associated with the antisymmetric generator $\mathbb{H}_{\mu\nu}(x)$, the ordinary potential $A_\mu(x)$ and the parallel 4-gradient. All these elements have then to be varied independently.

The combined information furnished by the variations of $\mathbb{F}_{\mu\nu}(x)$, $\mathbb{K}_\mu(x,P(x))$ and $\mathbb{B}_\mu(x)$ leads to

(a) The Maxwell equations:

$$\mathbb{F}_{\mu\nu}(x) = F_{\mu\nu}(x) + i\bar{F}_{\mu\nu}(x), \quad \mathbb{F}^{\mu\nu}{}_{,\nu}(x) = J^\mu(x); \tag{3.7}$$

(b) The representation of $\mathbb{F}_{\mu\nu}(x)$ in terms of the path-dependent potential:

$$\begin{aligned}\mathbb{F}_{\mu\nu}(x) &= \mathbb{K}_{\nu:\mu}(x,P(x)) - \mathbb{K}_{\mu:\nu}(x,P(x))\\ &\quad - i\int_{P(x)}^{x} dx'^{\lambda}\, J_{\mu\nu\lambda}(x');\end{aligned} \tag{3.8}$$

(c) The Dirac representation

$$\mathbb{F}_{\mu\nu}(x) = \mathbb{F}^0_{\mu\nu}(x) + \epsilon_{\mu\nu}{}^{\rho\sigma}\mathbb{B}_{\rho,\sigma}(x), \tag{3.9}$$

where

$$\mathbb{F}^0_{\mu\nu}(x) = \sum_i \mathbb{F}^{0(i)}_{\mu\nu}(x)$$

$$= \sum_i G_i \int_{-\infty}^{\infty} ds \int_{\infty}^{0} dt \, \{\partial_t z^{(i)}_\mu \partial_s z^{(i)}_\nu - \partial_t z^{(i)}_\nu \partial_s z^{(i)}_\mu\} \times \delta^4(x - z^{(i)}) \quad (3.10)$$

is the well-known Dirac string written in complex notation;

(d) The representation

$$\mathbb{F}_{\mu\nu}(x) = \mathbb{H}_{\mu\nu}(x) + A_{\nu,\mu}(x) - A_{\mu,\nu}(x), \quad (3.11)$$

which in fact contains many other interesting representations.[6,9]

The variation of the surfaces Σ_i, which contain the world lines of the particles Π_i and the selection of paths attached to them, leads to

(a) The Lorentz equations:

$$m_i \dot{u}^{(i)}_\mu = (e_i F_{\mu\nu} + g_i \bar{F}_{\mu\nu}) u^{\nu(i)}; \quad (3.12)$$

(b) The Wentzel condition:

$$\partial_t z^{\nu(i)} \partial_s z^{\lambda(i)} (e_i \, g_{\mu\nu\lambda} - g_i \, j_{\mu\nu\lambda}) = 0. \quad (3.13)$$

Hence the whole set of the Maxwell–Lorentz equations is obtained as stationary conditions for the action (3.2).

The equation (3.13) may be considered as an equation of motion for the path[9] or, if the path is considered as a gauge variable, as a restriction on the possible choices of the gauge which is automatically satisfied assuming Dirac's veto or Wentzel's weaker condition.

It is also interesting to discuss the Lagrangian formulation for the restricted problem of a point charge in some general externally fixed electromagnetic field. The action may be written by considering the appropriate restriction of (3.2):

$$S_{12} = \int_{t_1}^{t_2} dt \, L(t), \quad (3.14)$$

where

$$L(t) = m\{u^\mu(t) u_\mu(t)\}^{1/2} + RE\{G^* A_\mu(x) u^\mu(t) + G^* \int_{\infty}^{0} ds \mathbb{H}_{\mu\nu}(z) \partial_t z^\mu(s,t) \partial_s z^\nu(s,t). \quad (3.15)$$

Here the dynamical variable is the surface $z^\mu(s,t)$ which has been parametrized in such a way that for fixed $t, z^\mu(s,t)$ describes the path $P(x(t))$ when s moves from ∞ to 0, and for $s = 0$ one has

$$z^\mu(0,t) = x^\mu(t), \quad \partial_t z^\mu(0,t) = u^\mu(t). \quad (3.16)$$

The Lorentz tensor $\mathbb{H}_{\mu\nu}(x)$ and the potential $A_\mu(x)$ are externally fixed complex objects which reproduce the field strength by

$$\mathbb{F}_{\mu\nu}(x) = \mathbb{H}_{\mu\nu}(x) + A_{\nu,\mu}(x) - A_{\mu,\nu}(x). \quad (3.17)$$

The variations of the action lead now to the equations

$$m\dot{u}_\mu(t) = \omega_{\mu\nu}(x) u^\nu(t), \quad (3.18)$$

where we have introduced the notation

$$\omega_{\mu\nu}(x) \equiv e F_{\mu\nu}(x) + g \bar{F}_{\mu\nu}(x), \quad (3.19)$$

together with the Wentzel condition (3.13).

4. THE STRUCTURE OF THE WAVEFUNCTION

Let us consider in this section the quantum description of a point charge $G = (e,g)$ moving in some externally fixed general field strength $\mathbb{F}_{\mu\nu}(x)$.

A satisfactory classical Lagrangian formulation for this problem has been obtained in the last section. The Lagrangian (3.15) reproduces the correct classical equations of motion using the surface $z^\mu(s,t)$ as the dynamical variable.

In general, if the dynamical degrees of freedom of some classical system are $q_1(t), q_2(t), ..., q_N(t)$, the wavefunction is defined as a function on the classical configuration space: $\Psi = \Psi(q_1, q_2, ..., q_N)$. In the present case, the dynamical variable is $z^\mu(t,s)$ and the wavefunction should be defined as a functional over the function $z^\mu(s)$ which describes the path $P(x(t))$. Hence, the first step towards the quantum description of the system should be the identification of the wavefunction as a path dependent object:

$$\Psi = \Psi\{z^\mu(s)\} = \Psi(x, P(x)). \quad (4.1)$$

On the other hand one may compute the total canonical momentum of the system as:

$$P_\mu(t) = \int_{\infty}^{0} ds \frac{\delta L}{\delta(\partial_t z^\mu(s,t))}. \quad (4.2)$$

Carrying on the calculation with the Lagrangian (3.15) it is easy to obtain

$$P_\mu = m u_\mu + \Omega_\mu(x, P(s)), \quad (4.3)$$

where we have introduced

$$\Omega_\mu(x, P(x)) \equiv \int_{P(x)}^{x} dx'^\lambda \, \omega_{\lambda\mu}(x') + A_{;\mu}(x, P(x)). \quad (4.4)$$

$A_{;\mu}$ is an arbitrary parallel 4-gradient and the field strength $\omega_{\mu\nu}(x)$ has been introduced in (3.19).

Let us now consider the differential operator

$$P_{;\mu} \equiv i \partial_{;\mu}, \quad (4.5)$$

which obviously verifies the commutation relations

$$\{P_{;\mu}, P_{;\nu}\} = 0. \quad (4.6)$$

From the definition (2.4) of the parallel derivative, it is evident that the change induced in the path-dependent wavefunction by a global infinitesimal translation

$$\Psi(x, P(x)) \rightarrow \Psi(x + \epsilon, P_{\epsilon}(x + \epsilon)) \quad (4.7)$$

is generated by $P_{;\mu}$.

Hence one must have the correspondence rule

$$p_\mu = m u_\mu \rightarrow P_{;\mu} - \Omega_\mu(x, P(x)). \quad (4.8)$$

Now the dynamical equation is obtained as usual. From the relation

$$p_\mu p^\mu = m^2, \quad (4.9)$$

one obtains the path-dependent Klein–Gordon equation

$$\{P_{;\mu} - \Omega_\mu(x, P(x))\}\{P^\mu_{;} - \Omega^\mu(x, P(x))\} \Psi(x, P(x)) = m^2 \Psi(x, P(x)), \quad (4.10)$$

and a similar procedure leads to the path-dependent Dirac equation or any other desired dynamical equation.

The structure of the path-dependent wavefunction is severely limited by the invariance requirements of the formulation under gauge transformations. The Lagrangian formulation clearly identifies two types of gauge freedom in the path-dependent potential:

1. The path-dependent potential may be changed by an arbitrary parallel 4-gradient

$$\Omega_\mu(x,P(x)) \to \Omega_\mu(x,P(x)) + \Lambda_{:\mu}(x,P(x)); \qquad (4.11)$$

2. Since the path selection in the Lagrangian is arbitrary, the transformation

$$\Psi(x,P(x)) \to \Psi(x,P'(x)) \qquad (4.12)$$

must also be considered as a gauge transformation.

These two types of gauge transformations are in fact closely related as we shall see in the course of this section.

The wavefunction must transform in order to ensure the invariance of the electromagnetic current:

$$\Psi^*(x,P(x))\{P_{:\mu} - \Omega_\mu(x,P(x))\}\Psi(x,P(x)) - \text{c.c.} \qquad (4.13)$$

For gauge transformations of type 1, the problem is formally identical to the corresponding problem in ordinary electromagnetism. Hence, under these transformations, one must have

$$\Psi(x,P(x)) \to \Psi(x,P(x)) \exp\{-i\Lambda(x,P(x))\}. \qquad (4.14)$$

One may then choose the path-dependent potential in the specific form

$$\Omega_\mu(x,P(x)) = \int_{P(x)}^x dx'^\lambda\, \omega_{\lambda\mu}(x') \qquad (4.15)$$

and concentrate in the transformation properties of the wavefunctions under gauge transformations of type 2.

Let us then consider two arbitrary paths $P(x)$ and $P'(x)$ with the same finite end at the point x and the same asymptotic direction. Let us also introduce an auxiliary system C of parallel paths $C(x')$ each one attached to a point x' in some region D containing the point x. The system C is chosen to have the same direction of asymptoticity as $P(x)$ and $P'(x)$ and it is required that the paths $C(x')$ do not cut any source. It is evident that if x is not a source itself, this construction is always possible in a sufficiently small neighborhood of the point x.

One may now introduce the path-dependent function

$$\Lambda_C(x,P(x)) = -\tfrac{1}{2}\int_{\Sigma_C(P)} d\sigma^{\mu\nu}\, \omega_{\mu\nu}, \qquad (4.16)$$

where $\Sigma_C(P)$ is a surface with edges in the paths $P(x)$ and $C(x)$. Its positive sense is defined by the sense of circulation along $P(x)$ and it is required that the surface does not contain any source.

The field strength $\omega_{\mu\nu}(x)$ verifies the Maxwell equations

$$\partial^\nu \overline{\omega}_{\mu\nu}(x) = e g_\mu^s(x) - g j_\mu^s(x), \qquad (4.17)$$

where j_μ^s and g_μ^s are, respectively, the external electric and magnetic currents of the sources. According to the quantization condition, given in Zwanzinger's[13] chiral invariant form

$$eg^s - ge^s = 2\pi N, \qquad (4.18)$$

the source charges of $\omega_{\mu\nu}(x)$ must be entire multiples of 2π. Hence a change in the surface in the definition (4.16) can only change $\Lambda_C(x,P(x))$ by a multiple of 2π. This multiformity is not important as we shall see that the final result does not depend of the chosen branch of $\Lambda_C(x,P(x))$.

One may now compute the parallel derivative of this function and using (4.17) and Stokes theorem it is easy to obtain

$$\Lambda_{C:\mu}(x,P(x)) = \Omega_\mu(x,P(x)) - \Omega_\mu(x,C(x)). \qquad (4.19)$$

Hence the change of path from $C(x)$ to $P(x)$ may be considered as a gauge transformation of type 1. Then one must have

$$\Psi(x,P(x)) = \Psi(x,C(x)) \exp\{-i\Lambda_C(x,P(x))\}. \qquad (4.20)$$

This result does not depend of the specific chosen branch as indicated. The same result holds with $P(x)$ replaced by $P'(x)$ and one must have

$$\Psi(x,P'(x)) = \Psi(x,P(x)) \exp\{-i\Lambda_C(x,P'(x)) + i\Lambda_C(x,P(x))\}. \qquad (4.21)$$

Let us introduce a surface $\Sigma(P,P')$ with edges at $P(x)$ and $P'(x)$ and with its positive sense defined by the sense of circulation along $P'(x)$. Now using the Gauss theorem and charge quantization it is easy to obtain

$$\Psi(x,P'(x)) = \Psi(x,P(x)) \times \exp\left\{\frac{i}{2}\int_{\Sigma(P,P')} d\sigma^{\mu\nu}\omega_{\mu\nu}\right\}. \qquad (4.22)$$

It follows from this relation that the norm of the wavefunction is path independent. It also follows from (4.22) a very useful relation between the Mandelstam and the parallel derivatives when acting on an electromagnetic wavefunction. From the definitions (1.4) and (2.4) it immediately follows that

$$\Psi(x+\epsilon, P_{:}(x+\epsilon)) = \{\P + \epsilon^\mu \partial_{:\mu}\}\Psi(x,P(x)), \qquad (4.23)$$

$$\Psi(x+\epsilon, P_E(x+\epsilon)) = \{\P + \epsilon^\mu D_\mu\}\Psi(x,P(x)). \qquad (4.24)$$

But $P_{:}(x+\epsilon)$ and $P_E(x+\epsilon)$ are paths with the same finite end and the same asymptotic direction. Hence, using (4.22) one may relate the wavefunction in the two paths and it is easy to obtain the expression

$$D_\mu \Psi(x,P(x)) = \{\partial_{:\mu} - i\Omega_\mu(x,P(x))\}\Psi(x,P(x)). \qquad (4.25)$$

In the linear space of the path-dependent wavefunctions, (4.22) defines a linear subspace which we shall refer hereafter as the electromagnetic subspace. Only the wavefunctions in this subspace are acceptable to describe the quantum motion of the particle. An operator which leaves this subspace invariant will be known as physical or gauge invariant and only these operators are suitable to represent physical quantities. It follows from (4.25) that the Mandelstam derivatives D_μ leave invariant the electromagnetic subspace. Hence any operator constructed with D derivatives must be physical.

According to (4.10) and (4.25) the dynamical equation may be written in terms of Mandelstam covariant derivatives. Hence a wavefunction initially electromagnetic will remain in the electromagnetic subspace in the course of its dynamical evolution.

The path dependence of the wavefunction is completely determined by (4.22). The dynamical equation may then be used to find the x-dependence of the wavefunction. For this reason this quantum formulation is in fact a description of the degrees of freedom associated with a point particle.

In the case of ordinary electromagnetism, this formulation should reduce explicitly to the usual description in terms of point functions. This is indeed the case as it may be seen in a rather trivial way. When magnetic poles are not present one must have

$$\omega_{\mu\nu}(x) = eF_{\mu\nu}(x) = e(A_{\nu,\mu}(x) - A_{\mu,\nu}(x)), \qquad (4.26)$$

this relation holding everywhere except on the sources. We may then use this decomposition in (4.22) and use Stokes theorem to obtain

$$\Psi(x,P'(x)) \exp\left\{-ie\int_{P'(x)}^{x} dx'^{\lambda} A_{\lambda}(x')\right\} = \Psi(x,P(x)) \exp\left\{-ie\int_{P(x)}^{x} dx'^{\lambda} A_{\lambda}(x')\right\}, \qquad (4.27)$$

for arbitrary $P(x)$ and $P'(x)$.

Hence both terms must be path independent:

$$\Psi(x,P(x)) = \psi(x) \exp\left\{ie\int_{P(x)}^{x} dx'^{\lambda} A_{\lambda}(x')\right\} \qquad (4.28)$$

and one recovers the Mandelstam path-dependent wavefunction. It is immediately seen that

$$D_{\mu}\Psi(x,P(x)) = \{(\partial_{\mu} - ieA_{\mu}(x))\psi(x)\} \times\exp\left\{ie\int_{P(x)}^{x} dx'^{\lambda} A_{\lambda}(x')\right\}. \qquad (4.29)$$

Hence in the dynamical equation, or in any physical operator constructed with D derivatives, the path dependence of $\Psi(x,P(x))$ simply factors out and becomes irrelevant. One obtains then a projected dynamical equation for the ordinary wavefunction $\psi(x)$ which reproduces the usual quantum formulation for a particle in an external field.

When magnetic poles are present, (4.26) cannot be valid globally and one must retain the general relation (4.22) to determine the path-dependent structure of the wavefunction, but it remains true that the wavefunction continues to describe the degrees of freedom associated with the motion of a point charge.

To close this section let us now generalize (4.22) to the case where $P(x)$ and $P'(x)$ do not have the same direction of asymptoticity. In this case (4.22) cannot be valid as it stands. Two different surfaces $\Sigma(P,P')$ and $\Sigma'(P,P')$ will fail to close in the asymptotic region since $P(x)$ and $P'(x)$ are no longer asymptotically parallel. Then a change in the surface in the phase factor of (4.22) will contribute by the magnetic flux at spatial infinity through the asymptotic hole between the two surfaces. If the electromagnetic field is created by a finite system of charges, the asymptotic field may be considered as generated by a point charge of total electric charge e_T^s and magnetic charge g_T^s. In general electromagnetism where one may have $g_T^s \neq 0$, there will be a radial magnetic field giving a finite contribution to the asymptotic flux. Accordingly, (4.22) depends on the particular surface drawn through $P(x)$ and $P'(x)$ and becomes undetermined.

Under this general gauge transformation, $P(x) \to P'(x)$, the wavefunction must transform to ensure the invariance of the electromagnetic current and of its norm. Both requirements can be embodied into a single expression by writing

$$\Psi^*(x,P(x))(\P + \epsilon^{\mu}D_{\mu})\Psi(x,P(x)) = \Psi^*(x,P'(x))(\P + \epsilon^{\mu}D_{\mu})\Psi(x,P'(x)), \qquad (4.30)$$

where ϵ must be considered as an arbitrary infinitesimal displacement. According to (1.4) this relation is equivalent to

$$\Psi^*(x,P(x))\Psi(x_0,P_E(x_0)) = \Psi^*(x,P'(x))\Psi(x_0,P'_E(x_0)), \qquad (4.31)$$

where $x_0 = x + \epsilon$. By iterating (4.31) it is not difficult to obtain the following finite relation:

Let x and y be two arbitrary points which do not coincide with any source. Let $P(x)$ and $P'(x)$ be two arbitrary paths attached to the point x with asymptotic directions k_{μ} and k'_{μ}, respectively. Let $\Pi(y)$ and $\Pi'(y)$ be another arbitrary couple of paths attached to the point y. It is required that $\Pi(y)$ and $\Pi'(y)$ be asymptotically parallel to $P(x)$ and $P'(x)$. An electromagnetic wavefunction must verify the relation

$$\Psi(x,P(x))\Psi^*(y,\Pi'(y)) = \Psi(x,P(x))\Psi^*(y,\Pi(y)) \cdot\exp\left\{\frac{i}{2}\int_{\Sigma(P,P')} d\sigma^{\mu\nu}\,\omega_{\mu\nu} - \frac{i}{2}\int_{\Sigma(\Pi,\Pi')} d\sigma^{\mu\nu}\,\omega_{\mu\nu}\right\}, \qquad (4.32)$$

where the same cutoff curve must be used to evaluate the two surface integrals.

(4.32) is the desired generalization of (4.22) to the case where the two paths are not asymptotically parallel. It looks as a duplicate version of it where factorization is not possible due to the asymptotic problem already discussed. It will be useful in the next section in discussing the properties of the angular momentum.

5. ANGULAR MOMENTUM

Let us consider some spatial path $z_i(s)$ attached to some point $x_i = z_i(0)$. Its asymptotic direction is given by some unitary vector k_i. Under some infinitesimal rotation of angle $\delta\phi$ and axis n_i, the path will transform into

$$z_i'(s) = z_i(s) + \delta\phi\epsilon_{ijk} n_j z_k(s). \qquad (5.1)$$

Then one may introduce the differential operator

$$in{\cdot}J\Psi(x,P(x)) \equiv \lim_{\delta\phi\to 0} \frac{\Psi\{z_i'(s)\} - \Psi\{z_i(s)\}}{\delta\phi}, \qquad (5.2)$$

which obviously verifies the commutation relations

$$\{J_i,J_j\} = i\epsilon_{ijk} J_k. \qquad (5.3)$$

Obviously, J_i is the generator of spatial rotations in the linear space of path-dependent wavefunctions.

The transformation (5.1) may be divided into two parts

$$z_i'(s) = z_i(s) + \delta\phi\epsilon_{ijk} n_j x_k + \delta\phi\epsilon_{ijk} n_j \{z_k(s) - x_k\}. \qquad (5.4)$$

In the first part, the path is parallel transported to the rotated end point $z_i'(0)$. The corresponding change in the wavefunction is generated by

$$L_{:i} \equiv i\epsilon_{ijk} x_j P_{:k}. \tag{5.5}$$

In the second part, the path is rotated around the finite end point to reach the final path. One may then introduce the generator

$$in \cdot O\Psi(x,P(x)) \equiv \lim_{\delta\phi \to 0} \frac{\Psi\{z_i + \delta\phi\epsilon_{ijk} n_j (z_k - x_k)\} - \Psi\{z_i\}}{\delta\phi}, \tag{5.6}$$

which also verifies

$$\{O_i, O_j\} = i\epsilon_{ijk} O_k, \tag{5.7}$$

and one has

$$J_i = L_{:i} + O_i. \tag{5.8}$$

The $L_{:}$ operator describes the rotational degrees of freedom associated with the finite end. It reduces to the ordinary angular momentum operator when acting on a point function. The O operator takes into account the new degrees of freedom associated with the path itself. It has no analogue in the usual Hilbert space of point functions.

Since the O operator does not affect the end point, its action on an electromagnetic function must be completely determined by (4.32).

Let us then consider two arbitrary points x and y and two arbitrary paths $P(x)$ and $\Pi(y)$ attached to them. It is required that the paths do not cut any source and that both paths have the same asymptotic direction described by the unitary vector k_i. Under an infinitesimal rotation of angle $\delta\phi$ and axis along the "i" direction the paths $P(x)$ and $\Pi(y)$ will transform respectively into $P'(x)$ and $\Pi'(y)$. According to the definition (5.6) of the O operator one must have

$$\Psi(x,P'(x))\Psi^*(y,\Pi'(y)) = \{\P + i\delta\phi O_i(x) - i\delta\phi O_i^\dagger(y)\} \times \Psi(x,P(x))\Psi^*(y,\Pi(y)), \tag{5.9}$$

where we have written $O_i(x)$, $O_i^\dagger(y)$ to emphasize the wavefunction on which these operators must act.

On the other hand one may use (4.32) to relate the values of the wavefunction on the four paths. The result is

$$\begin{aligned}\Psi(x,P'(x))\Psi^*(y,\Pi'(y)) &= \{1 + i\delta\phi\epsilon_{ijk} x_j \Omega_k(x,P(x)) \\ &\quad - i\delta\phi\epsilon_{ijk} y_j \Omega_k(y,\Pi(y)) \\ &\quad + i\delta\phi \Lambda_i(x,P(x)) - i\delta\phi\Lambda_i(y,\Pi(y))\} \\ &\quad \times \Psi(x,P(x))\Psi^*(y,\Pi(y)),\end{aligned} \tag{5.10}$$

where we have introduced the path-dependent function

$$\Lambda_i(x,P(x)) \equiv \epsilon_{ijk} \int_{P(x)}^{x} dz_m\, \omega_{mj}(z) z_k. \tag{5.11}$$

Comparing (5.9) and (5.10) one must have

$$\begin{aligned}&\Psi^{-1}(x,P(x))\{O_i(x) - \epsilon_{ijk} x_j \Omega_k(x,P(x)) \\ &\quad - \Lambda_i(x,P(x))\}\Psi(x,P(x)) \\ &= \Psi^{*-1}(y,\Pi(y))\{O_i^\dagger(y) - \epsilon_{ijk} y_j \Omega_k(y,\Pi(y)) \\ &\quad - \Lambda_i(y,\Pi(y))\}\Psi^*(y,\Pi(y)) \\ &= C_i(k),\end{aligned} \tag{5.12}$$

where $C_i(k)$ must be a real constant which may only depend on the common asymptotic direction of $P(x)$ and $\Pi(y)$. Hence one must have

$$O_i\Psi(x,P(x)) = \{\epsilon_{ijk} x_j \Omega_k(x,P(x)) + \Lambda_i(x,P(x)) + C_i(k)\}\, \Psi(x,P(x)). \tag{5.13}$$

Since the O operators are known to verify the commutation relations (5.7) the constant $C_i(k)$ can be evaluated to make (5.13) compatible with them. After a tedious but straightforward calculation one finds

$$C_i(k) = \mu k_i, \quad \mu \equiv \frac{eg_T^s - ge_T^s}{4\pi}, \tag{5.14}$$

where e_T^s and g_T^s are respectively the total electric and magnetic charges of the sources of the external electromagnetic field.

From (5.8) and the relation (4.25) between the Mandelstam and the parallel derivatives one may write down the desired restriction of the angular momentum operator to the electromagnetic subspace:

$$J_i\Psi(x,P(x)) = \{\epsilon_{ijk} x_j(-iD_k) \\ \mu k_i + \Lambda_i(x,P(x))\}\Psi(x,P(x)). \tag{5.15}$$

From this expression it follows that J_i does not leave invariant the electromagnetic subspace unless Λ_i is path independent and this is only possible for rotationally invariant field strengths. Hence only in this case J_i becomes physical or gauge invariant.

In the case of a static source at the origin of coordinates one finds

$$\Lambda_i(x) = (\mu x_i / r) - \mu k_i \tag{5.16}$$

and the angular momentum becomes

$$J_i\Psi(x,P(x)) = \{\epsilon_{ijk} x_j(-iD_k) + \mu x_i/r\}\Psi(x,P(x)). \tag{5.17}$$

This is the well-known operator used to represent the angular momentum in the field of a static monopole. The new point is that this operator has been identified with the generator of spatial rotations in the linear space of the path-dependent wavefunctions while this operator cannot be associated with any coordinate transformation in the ordinary space of the point functions.

[1]S. Mandelstam, Ann. Phys. **19**, 1 (1962).
[2]T.T. Wu and C.N. Yang, Phys. Rev. D **12**, 3845 (1975).
[3]T.T. Wu and C.N. Yang, Nuc. Phys. B 107, 365 (1976).
[4]P.A.M. Dirac, Proc. R. Soc. A **133**, 60 (1931).
[5]P.A.M. Dirac, Phys. Rev. V **74**, 817 (1948).
[6]N. Cabibbo and E. Ferrari, Nuovo Cimento **23**, 1147 (1962).
[7]P. Vinciarelli, Phys. Rev. D **6**, 3419 (1972).
[8]P. Jordan, Annalen der Physik, 5 **32**, 66 (1938).
[9]A.O. Barut, J. Phys. A **11**, 2073 (1978).
[10]T.T. Wu and C.N. Yang, Phys. Rev. D **14**, 437 (1976).
[11]G. Wentzel, Supp. Prog. Theo. Phys. **37–38**, 163 (1966).
[12]D. Rosenbaum, Phys. Rev. D **147**, 891 (1966).
[13]D. Zwanziger, Phys. Rev. D **3**, 880 (1971).

Multipoles and excitons in higher order coherence

Andrei N. Weiszmann
College of Staten Island of The City University of New York, Staten Island, New York 10301

(Received 7 November 1979; accepted for publication 1 February 1980)

Considering the coherence properties of an inelastically scattered quantum radiation by an ensemble of N identical three level atoms in a solid-like array, we addressed the question whether this process is actually mediated by a series of multipole interactions and/or a series of virtual intermediate-exciton transitions and how the type of the considered mediation is reflected in the higher order coherence properties of the scattered radiation. Using a projector technique the reduced field density operator, the atomic and field correlation functions were found, the higher order coherence properties were studied and the arrival-time and phase correlation conditions of the light quanta were obtained and analysed. The results suggest that not only the arrival-time and phase correlation conditions are mathematically independent from the type of the considered mediating process, but that the multipole interaction approximation of a resonant Raman scattering, in the above considered system, is physically equivalent with the virtual intermediate-exciton transitions too.

I. INTRODUCTION

Higher order coherence phenomena in scattered radiation are a subject of considerable interest for modern optical physics. Although higher order coherence effects have been investigated fairly extensively since the early 1960's, they have been mainly studied in areas such as photon statistics in lasers near the threshold of oscillations and in connection with photoelectric detection of light fluctuations.

Higher order photon-counting measurements performed on a scattered quantum radiation can yield information on the scattering medium not available with conventional first order experiments. The theory behind these experiments was developed based in part on the concepts of higher order optical coherence and correlations. Discrepancies between the experimental findings and theoretical predictions point to the fact that there is a clear need for further refinement of the existent theory.

The question we would like to address in this paper is:

(a) whether a resonant Raman scattering process on a system of frozen or nearly frozen N identical three level atoms is actually mediated by a series of dipole, quadrupole, and multipole interactions and/or by a series of virtual intermediate-exciton transitions;

(b) how this mediation is reflected in the higher order coherence properties of the scattered radiation, namely in the arrival-time and phase correlation conditions of light quanta.

The basic Raman effect[1,2] is usually considered as an inelastic light-scattering process in which an incident quantum $h\nu_i$ *is scattered into a quantum* $h\nu_s$ while the difference in energy is absorbed (or emitted) by the material scattering center. In principle the excitation $h\nu_{is} = h(\nu_i - \nu_s)$ of the material system may be considered a pure electronic excitation, a vibrational, rotational or other kind of excitation of the scattering center, depending on the nature of the center or the technique we approach the process.[3-8] Depending on the treatment and the correlations involved the coherence properties of the scattered quantum radiation can be investigated in various ways.[9-12]

In three recent papers[13-15] the author considered the coherence properties of a quantum radiation inelastically scattered by an ensemble of N identical three level atoms. Considering dipole–dipole interactions in the Hamiltonian the atomic correlation functions were found, the higher order coherence properties were analyzed and the correlation conditions in the arrival-times and phases of light quanta were obtained. In the present paper the above mentioned investigation is intended to be expanded by considering the questions addressed above.

II. THE HAMILTONIAN

The scattering center will consist of a system of N fixed three level atoms at uncorrelated spatial positions—a so-called frozen or nearly frozen "dense" gas—on which some of the features of solids can be carried over.[16] Let $|l\rangle$ be an eigenstate of the atomic Hamiltonian

$$H_A = \sum_l \epsilon_l \mathcal{N}_l , \tag{2.1}$$

where

$$\mathcal{N}_l = \sum_{m=1}^{N} |l\rangle\langle l|_m , \tag{2.2}$$

with $|g\rangle$ staying for ground state $|h\rangle$ for intermediate state and $|f\rangle$ for final state. The Hamiltonian of the field—the incident coherent radiation—is

$$H_F = \hbar \sum_\kappa \omega_\kappa a_\kappa^+ a_\kappa , \tag{2.3}$$

where

$$[a_\kappa, a_{\kappa'}^+] = \delta_{\kappa\kappa'},$$

where the interaction between the field and atomic system will depend on the type of mediation considered.

 0022-2488/80/061546-06$1.00

Single-photon interactions of atoms and molecules at optical or even ultraviolet frequencies are adequately described in the dipole–dipole approximation. Basically, this stems from the fact that the atomic dimensions are much smaller than the wavelength of optical radiation. It was shown recently[17,18] that for certain ranges of photon frequencies contributions from electric quadrupole transitions will be larger than the corresponding dipole contributions. This will occur when the photon frequency is approximately equal to the energy difference between two states connected with a quadrupole transition.[17] In most cases quadrupole resonances occur at frequencies around which the pure dipole contributions is near a minimum. As a result, quadrupole contributions will often exceed the pure dipole ones by several order of magnitude.

In this case the interaction part of the Hamiltonian of the atomic system + field is

$$V= -\left(\frac{e}{mc}\right)\bar{p}\cdot\bar{A}(\bar{r}) + \left(\frac{e^2}{2mc^2}\right)\bar{A}^2(\bar{r}). \tag{2.4}$$

Power and Zienau[19] have shown that the interaction V, in all its generality, can be written in terms of multipole expansion

$$V\equiv V^D + V^Q + \cdots \equiv -e\bar{r}\cdot\bar{E}(0) - \tfrac{1}{2}e\sum_{\alpha\beta} Q_{\alpha\beta}\nabla_\beta \bar{E}_\alpha(0) + \cdots, \tag{2.5}$$

where

$$Q_{\alpha\beta} = x_\alpha x_\beta - \tfrac{1}{3}r^2\delta_{\alpha\beta}, \quad \alpha,\beta = 1,2,3, \tag{2.6}$$

is the quadrupole dyadic.

The contribution of the first term in (2.5) to the higher order coherence properties of the scattered radiation was analyzed in Refs. 13, 14, 15. It is well known[20] that $V^D = -e\bar{r}\cdot\bar{E}(0)$ can be written in second quantized form.

$$V^D = \sum_\kappa \sum_m i\hbar \mu_{hf}[a_\kappa^+(t)|h\rangle\langle f|_m - a_\kappa(t)|f\rangle\langle h|_m], \tag{2.7}$$

where

$$\mu_{hf} = \left(\frac{\omega_\kappa}{2\hbar\epsilon_g V}\right)^{1/2} \bar{\mu}_{hf}\cdot\hat{e}^{(\kappa)}, \tag{2.8}$$

with

$$\bar{\mu}_{hf} = \langle h|e\bar{r}|f\rangle = \langle h|\bar{d}|f\rangle \tag{2.9}$$

being the dipole matrix element between the atomic states $|h\rangle$ and $|f\rangle$ and $\hat{e}^{(\kappa)}$ the electric field polarization vector. All the other notations are self-evident.

To write V^Q in a suitable second quantized form we first apply the ∇_β operator in (2.5) on the α component of the electric field, than taking into account the double summation, collect the corresponding terms and we have

$$V^Q = \tfrac{1}{2}e\sum_{\alpha\beta} Q_{\alpha\beta}\hat{e}_\alpha^{(\kappa)}K_\beta \sum_\kappa \hbar\left(\frac{\omega_\kappa}{2\hbar\epsilon_g V}\right)^{1/2} \times [a_\kappa e^{-i\omega_\kappa t} + a_\kappa^+ e^{i\omega_\kappa t}], \tag{2.10}$$

which is identical with Birman's decomposition of the multipole term into a symmetrical quadrupole $\frac{1}{2}\bar{k}\cdot(\overline{rp}+\overline{pr})\cdot\bar{\epsilon}$,[17] terms and an antisymmetrical magnetic dipole term, if only we label

$$Q_\alpha = \tfrac{1}{2}e\left(\frac{\omega_\kappa}{2\hbar\epsilon_g V}\right)^{1/2}\sum_\beta Q_{\alpha\beta}K_\beta \tag{2.11}$$

as the α component of the $\bar{Q}^{(\kappa)}$ vector and take its scalar product with $\hat{e}_\alpha^{(\kappa)}$. In the interaction picutre (I.P.)

$$\bar{Q}(t) = e^{i/\hbar(H_F+H_A)t}\bar{Q}e^{-i/\hbar(H_F+H_A)t} \tag{2.12}$$

and using the completeness relation

$$I = \sum_j |j\rangle\langle j| \tag{2.13}$$

twice, we have

$$Q(t) = \sum_{j\neq l} Q_{jl}|j\rangle\langle l|e^{i\omega_{jl}t}, \tag{2.14}$$

where

$$Q_{jl} = \langle j|\bar{Q}|l\rangle \tag{2.15}$$

is the quadrupole matrix element between the atomic states $|j\rangle$ and $|l\rangle$ and $\omega_{jl} = \omega_j - \omega_l$. (2.14) and (2.10) leads immediatley to the quadrupole term of the interaction Hamiltonian

$$V^Q(t) = \sum_\kappa\sum_m\sum_{l\neq j} \hbar Q_{jl}[a_\kappa^+(t)|j\rangle\langle l|_m + a_\kappa(t)|l\rangle\langle j|_m], \tag{2.16}$$

where

$$a_\kappa(t) = a_\kappa(0)e^{-i(\omega_\kappa - \omega_{jl})t}, \quad a_\kappa^+(t) = a_\kappa^+(0)e^{i(\omega_\kappa - \omega_{jl})t}. \tag{2.17}$$

Thus (2.5) has the suitable form

$$V(t) = \sum_\kappa\sum_m\sum_{l\neq j} \hbar[(i\mu_{jl} + Q_{jl})a_\kappa^+(t)|j\rangle\langle l|_m - (i\mu_{lj} - Q_{lj})a_\kappa(t)|l\rangle\langle j|_m] \tag{2.18}$$

and the Hamiltonian of the multipole interaction mediated resonant Raman process will be

$$H = H_0 + H', \tag{2.19}$$

with

$$H_0 = H_A + H_F, \tag{2.20}$$

from (2.1) and (2.3), and

$$H' = V, \tag{2.21}$$

from (2.18).

However, in the Raman effect, when a photon is scattered producing a change in the vibrational state of the atomic system the virtual intermediate states involve the excitation of the electrons and this intermediate states can be considered as exciton states. Let assume therefore that the Raman scattering by phonons is an inelastic process mediated by virtual intermediate-exciton transitions.[17]

The material scattering center will be the same system of N fixed three level atoms as above, the atomic Hamiltonian being (2.1) with (2.2). The Hamiltonian of the field—the incident coherent radiation—is given by (2.3) and the exciton Hamiltonian representing the so-called weak-binding Wannier–Mott model will be given as[4]

$$H_{ex} = \sum_{\substack{j\neq l \\ \nu\bar{x}}} \epsilon_{\nu\bar{x}}(j,l) b^+_{\nu\bar{x}}(j,l) b_{\nu\bar{x}}(j,l), \quad (2.22)$$

with

$$b^+_{\nu\bar{x}} = \mathcal{N}^{1/2} \sum_k \bar{e}^{i\bar{\nu}\cdot\bar{k}} \sum_k C^+_{j\bar{k}} C_{l,\bar{k}-\bar{x}} \quad (2.23)$$

the exciton creation, and

$$b_{\nu\bar{x}} = \mathcal{N}^{-1/2} \sum_k e^{i\bar{\nu}\bar{k}} C^+_{l,\bar{k}-\bar{x}} c_{j\bar{k}} \quad (2.24)$$

the annihilation operators obeying "almost" boson commutation relations[22]

$$[b_{\nu\bar{x}}, b^+_{\nu'\bar{x}'}] = \delta_{\nu\nu'}\delta_{\bar{x}\bar{x}'} + \hat{O}\left(\frac{\text{no.of.el-hole pairs}}{\mathcal{N}}\right). \quad (2.25)$$

The excitons having inner quantum number ν and wave vector $\bar{x}$ are formed from the $j-l$ "bands". This is to say that we consider the electrons as being represented in Wannier exciton representation. The operator $C_{l,\bar{k}-\bar{x}}$ will destroy an electron (creates a hole) in band 1 with wave vector $\bar{k}-\bar{x}$ and similarly $C^+_{j\bar{k}}$ creates an electron in band j with wave vector $\bar{k}$. Thus an exciton state is given alternately by

$$\Psi_{jl}(\bar{k}_e, \bar{k}-\bar{x}_h) = C^+_{j\bar{k}_e} C_{l,\bar{k}-\bar{x}_h}\phi_0, \quad (2.26)$$

with ϕ_0 being our vacuum state. The $C^+_{j\bar{k}}$ and $C_{l,\bar{k}-\bar{x}}$ obey the usual anticommutation relations.

This noninteracting boson approximation is correct as long as the exciton density does not become too high. It is only overlaps and interactions which make excitons fail to behave as independent bosons.

In these assumptions—for the exciton mediated Raman process—the interaction part of the Hamiltonian for the above described atomic system + field is

$$V = V^1 + V^2, \quad (2.27)$$

where V^1 is linear in both exciton and photon operators and V^2 is linear in photon and bilinear in exciton operators[4]

$$V^1 = \sum_{l\neq j}\sum_{\nu\bar{x}}\sum_{\bar{\kappa}} [\{f_{\bar{\kappa}}(l,j,\bar{x}) b^+_{\nu\bar{x}}(l,j) a_{\bar{\kappa}} \delta_{\bar{x},\bar{\kappa}} + f^*_{\bar{\kappa}}(l,j,\bar{x}) b_{\nu\bar{x}}(l,j) a_\kappa \delta_{\bar{x},-\bar{\kappa}}\} - \{f^*_{\bar{\kappa}}(l,j,\bar{x}) b_{\nu\bar{x}}(l,j) a^+_{\bar{\kappa}} \delta_{\bar{x},\bar{\kappa}} + f_{\bar{\kappa}}(l,j,\bar{x}) b^+_{\nu\bar{x}}(l,j) a^+_{\bar{\kappa}} \delta_{\bar{x},-\bar{\kappa}}\}], \quad (2.28)$$

$$V^2 = \sum_{\substack{l\neq j \\ l'\neq j'}}\sum_{\substack{\nu\bar{x} \\ \nu'\bar{x}'}}\sum_{\bar{\kappa}} [\{F_{\bar{\kappa}}(l,j,\bar{x},l'j',\bar{x}') b^+_{\nu\bar{x}}(l,j) b_{\nu'\bar{x}'}(l',j') a_{\bar{\kappa}} \delta_{\bar{x}-\bar{x}',\bar{\kappa}} + F^*_{\bar{\kappa}}(l,j,\bar{x},l',j',\bar{x}') b_{\nu\bar{x}}(l,j) b^+_{\nu\bar{x}'}(l',j') a_{\bar{\kappa}} \delta_{\bar{x}-\bar{x}',-\bar{\kappa}}\} - \{F^*_{\bar{\kappa}}(l,j,\bar{x},l',j',\bar{x}') b_{\nu\bar{x}}(l,j) b^+_{\nu\bar{x}'}(l',j) a^+_{\bar{\kappa}} \delta_{\bar{x}-\bar{x}',\bar{\kappa}} + F_{\bar{\kappa}}(l,j,\bar{x},l',j',\bar{x}') b^+_{\nu\bar{x}}(l,j) b_{\nu\bar{x}'}(l',j) a^+_{\bar{\kappa}} \delta_{\bar{x}-\bar{x}',-\bar{\kappa}}\}], \quad (2.29)$$

with the coupling parameters $f_{\bar{\kappa}}$ and $F_{\bar{\kappa}}$ extensively analyzed in Ref. 4:

$$f_{\bar{\kappa}}(l,j,\bar{x}) \equiv -\tfrac{1}{2}l\left(\frac{\omega_\kappa}{2\hbar\epsilon_g V}\right)^{1/2} U^*_{l,j,\bar{x}}(0)\langle j|\hat{e}^{(\kappa)}\cdot\bar{d}\,|l\rangle,$$

$$F_{\bar{\kappa}}(l,j,\bar{x},l',j',\bar{x}') \equiv -\tfrac{1}{2}l\left(\frac{\omega_\kappa}{2\hbar\epsilon_g V}\right)^{1/2} \times\Big[\sum_{\bar{\beta}} U^*_{l,j,\bar{x}}(\bar{\beta}) U_{l',j',x'}(\bar{\beta}) e^{-i(\bar{x}-\bar{x}')\cdot\bar{\beta}} \langle j|\hat{e}^{(\kappa)}\cdot\bar{d}\,|j'\rangle\delta_{ll'} - \sum_{\bar{\beta}} U^*_{l,j,\bar{x}}(\bar{\beta}) U_{l',j',\bar{x}'}(\bar{\beta}) \langle l|\hat{e}^{(\kappa)}\cdot\bar{d}\,|l'\rangle\delta_{ll'}\Big]. \quad (2.30)$$

The wavefunction for the relative motion of the electron and hole is $U_{l,j,\bar{x}}(\bar{\beta})$ with $\bar{\beta}$ being the electron-hole separation. Ganguly and Birman[4] shows that $F_{\bar{\kappa}}$ vanishes if simultaneously $j=j'$ and $l=l'$. In interaction picture

$$V(t) = V^1(t) + V^2(t) = e^{i/\hbar(H_F+H_A+H_{ex})t}(V^1+V^2) \times e^{-i/\hbar(H_F+H_A+H_{ex})t} \quad (2.31)$$

and using the completeness relation (2.13) twice we have

$$V(t) = \sum_{\substack{l\neq j \\ l'\neq j'}}\sum_{\substack{\nu\bar{x} \\ \nu'\bar{x}'}}\sum_{\bar{\kappa}}\sum_m \{[G_m(l,j)|l\rangle\langle j|_m a_{\bar{\kappa}}(t) - G_m(j,l)|j\rangle\langle l|_m a^+_{\bar{\kappa}}(t)] + [\mathcal{G}_m(l,j)|l\rangle\langle j|_m a_{\bar{\kappa}}(t) - \mathcal{G}_m(j,l)|j\rangle\langle l|_m a^+_{\bar{\kappa}}(t)]\}, \quad (2.32)$$

where

$$G_m(l,j) = G^+_-(l,j)_m + G_+(l,j)_m,$$
$$G(j,l)_m = G_-(j,l)_m + G^+_+(j,l)_m, \quad (2.33a)$$
$$\mathcal{G}_m(l,j) = \mathcal{G}^+_-(l,j)_m + \mathcal{G}_+(l,j)_m,$$
$$\mathcal{G}_m(j,l) = \mathcal{G}_-(j,l)_m + \mathcal{G}^+_+(j,l)_m, \quad (2.33b)$$

with

$$G^+_-(l,j)_m = \langle l|f_{\bar{\kappa}} e^{i(\omega_{\bar{x}}-\omega_{\bar{\kappa}})t} b_{\nu\bar{x}}\delta_{\bar{x},\bar{\kappa}}|j\rangle_m,$$
$$G^+_+(j,l)_m = \langle j|f_{\bar{\kappa}} e^{i(\omega_{\bar{x}}+\omega_{\bar{\kappa}})t} b^+_{\nu\bar{x}}\delta_{\bar{x},-\bar{\kappa}}|l\rangle_m, \quad (2.34a)$$

$$G_+(l,j)_m = \langle l|f^*_{\bar{\kappa}} e^{-i(\omega_{\bar{x}}+\omega_{\bar{\kappa}})t} b_{\nu\bar{x}}\delta_{\bar{x},-\bar{\kappa}}|j\rangle_m,$$
$$G_-(j,l)_m = \langle l|f^*_{\bar{\kappa}} e^{-i(\omega_{\bar{x}}-\omega_{\bar{\kappa}})t} b_{\nu\bar{x}}\delta_{\bar{x},\bar{\kappa}}|l\rangle_m,$$
$$\mathcal{G}^+_-(l,j)_m = \langle l|F_{\bar{\kappa}} e^{i(\omega_{\bar{x}}-\omega_{\bar{x}'}-\omega_{\bar{\kappa}})t} b^+_{\nu\bar{x}} b_{\nu'\bar{x}'}\delta_{\bar{x}-\bar{x}',\bar{\kappa}}|j\rangle_m,$$
$$\mathcal{G}^+_+(j,l)_m = \langle j|F_{\bar{\kappa}} e^{i(\omega_{\bar{x}}-\omega_{\bar{x}'}+\omega_{\bar{\kappa}})t} b^+_{\nu\bar{x}} b_{\nu'\bar{x}'}\delta_{\bar{x}-\bar{x}',-\bar{\kappa}}|l\rangle_m, \quad (2.34b)$$

$$\mathcal{G}_+(l,j)_m = \langle l|F^*_{\bar{\kappa}} e^{-i(\omega_{\bar{x}}-\omega_{\bar{x}'}+\omega_{\bar{\kappa}})t} b_{\nu\bar{x}} b^+_{\nu'\bar{x}'}\delta_{\bar{x}-\bar{x}',-\bar{\kappa}}|j\rangle_m,$$
$$\mathcal{G}_-(j,l)_m = \langle j|F^*_{\bar{\kappa}} e^{-i(\omega_{\bar{x}}-\omega_{\bar{x}'}-\omega_{\bar{\kappa}})t} b_{\nu\bar{x}} b^+_{\nu'\bar{x}'}\delta_{\bar{x}-\bar{x}',\bar{\kappa}}|l\rangle_m,$$

and because the Kronecker symbols in (2.34), the photon creation and annihilation operators will have only ω_{jl} in the exponent of (2.17). Thus (2.27) has the suitable form

$$V(t) = -\sum_{\substack{l\neq j \\ l'\neq j'}}\sum_{\substack{\nu,\bar{x} \\ \nu',\bar{x}'}}\sum_{\bar{\kappa}}\sum_m [(G(jl)_m + \mathcal{G}(jl)_m) a^+_{\bar{\kappa}}(t)|j\rangle\langle l|_m - (G(lj)_m + \mathcal{G}(lj)_m) a_{\bar{\kappa}}(t)|l\rangle\langle j|_m] \quad (2.35)$$

and the Hamiltonian of the exciton mediated resonant Raman process will be

$$H = H_0 + H^1, \quad (2.36)$$

with

$$H_0 = H_A + H_F + H_{ex}, \quad (2.37)$$

from (2.1), (2.2), (2.3), (2.22), and

$$H' = V(t), \quad (2.38)$$

from (2.35).

III. TIME AND PHASE CORRELATION CONDITIONS

The question whether an initially coherent radiation preserves its coherence in the Glauber sense[11,12] (and if so, to what order) after being inelastically scattered by an atomic ensemble was studied in three recent papers[13–15] assuming a dipole–dipole interaction among the atoms. Using the projection technique the reduced field density operator was found, and with it the first and higher order field correlation functions were computed. It was proved that the inelastically scattered radiation is for the most part incoherent in any order at any time interval after collision, except some specific time intervals and positions, where the coherence is preserved, the order of coherence being determined by the atomic correlations considered. The time and phase correlation conditions were given and analyzed.

Let's consider at first the extension of these considerations for the multipole interactions among the atoms in the scattering center and later considering the virtual intermediate-exciton transitions too.

The necessary and sufficient conditions for coherence in the above mentioned sense are connected with some properties—namely normalization and factorization properties—of the jth order field corelation functions

$$G^{(j)}(x_1,\dots,x_{2j}) = \mathrm{Tr}\{\rho E^{(-)}(x_1)\dots E^{(-)}(x_j)E^{(+)}(x_{j+1})\dots E^{(+)}(x_{2j})\}, \tag{3.1}$$

where $x_j = (\bar{r}_j;t)$ and $E^{(-)}(x_j)$; $E^{(+)}(x_j)$ are the negative and positive frequency parts of $\bar{E}(\bar{r}_j;t_j)$ respectively, while ρ is the density operator of the whole system in Heisenberg representation. $\rho(t)$ satisfies the equation

$$i\hbar\frac{\partial\rho(t)}{\partial t} = \mathscr{H}\rho(t), \tag{3.2}$$

where $\mathscr{H}$ is the Liouville operator for the Hamiltonian given in (2.19). The reduced field density operator—which is the only relevant in our problem—can be found more easily in the interaction picture where (3.2) takes the form

$$i\hbar\frac{\partial\chi(t)}{\partial t} = \gamma\chi(t), \tag{3.3}$$

with γ the Liouville operator for the time-dependent interaction operator (2.18), and the reduced field density operator is

$$\sigma(t) = \mathrm{Tr}_A\chi(t). \tag{3.4}$$

The Tr_A meaning that the trace is applied over the atomic ensemble only. We can write the density operator of the whole system in (I.P) as

$$\chi(t) = f(H_A)\sigma(t) + \eta(t), \tag{3.5}$$

where

$$f(H_A) = \frac{e^{-\beta H_A}}{\mathrm{Tr}_A e^{-\beta H_A}} \tag{3.6}$$

and

$$\eta(t) = \mathscr{D}\chi(t), \tag{3.7}$$

with $\mathscr{D} = \mathscr{D}^2 = [1 - f(H_A)\mathrm{Tr}_A]$, a projector in the operator space of the entire system.

Then the equation of motion for $\sigma(t)$ is[14]

$$\frac{d\sigma(t)}{dt} = -i\,\mathrm{Tr}_A\gamma f(H_A)\sigma(t) - \mathrm{Tr}_A\gamma\int_0^t d\tau\exp[-i\mathscr{D}\gamma\tau]\mathscr{D}\gamma f(H_A)\sigma(\tau), \tag{3.8}$$

which is a non-Markoffian equation valid for all times and for any orders in $V(t)$, and has in the second term on the right-hand side a generalized collision operator containing the memory of the system.

Maintaining the second order terms in $V(t)$ for the collision operator, after performing the indicated integrals, we have for the reduced density operator

$$\frac{d\sigma(t)}{dt} = \sum_\kappa\{B'[a_{\bar\kappa}\sigma(t)a_{\bar\kappa}^+ - a_{\bar\kappa}^+a_\kappa\sigma(t)] + F'[a_{\bar\kappa}^+\sigma(t)a_{\bar\kappa} - a_{\bar\kappa}a_{\bar\kappa}^+\sigma(t)]\}, \tag{3.9}$$

with

$$B' = \hbar^2(\mu_{hf}\mu'_{fh} - Q_{hf}Q'_{fh})2\pi g(\omega_\kappa)(g,h),$$
$$F' = \hbar^2(\mu_{hf}\mu'_{fh} - Q_{hf}Q'_{fh})2\pi g(\omega_\kappa)(g,f), \tag{3.10}$$

where $g(\omega_\kappa)$ is a weigh function[20] and (g,h), (g,f) the atomic correlation functions

$$(g,h) = \frac{\mathscr{N}e^{-\beta\epsilon_h}[e^{-\beta\epsilon_g} + e^{-\beta\epsilon_h} + e^{-\beta\epsilon_f}]^{\mathscr{N}-1}}{\mathrm{Tr}_A[\exp(-\beta H_A)]}, \tag{3.11}$$

while (g,f) the same expression having ϵ_h interchanged with ϵ_f. The numerical values of these functions can be easily obtained by a relatively simply computer program for any desired value of $\mathscr{N}$. The solution of (3.9) is given in Ref. 14

$$\sigma(a_{\bar\kappa},a_{\bar\kappa}^+,t) = \prod\sum_{k^2}\sum_{\bar\kappa}e^{-(F'/4)k^2t}e^{-[(B'-F')/F']a_{\bar\kappa}a_{\bar\kappa}^+}\times\mathscr{I}_0[k(a_{\bar\kappa}a_{\bar\kappa}^+)^{1/2}], \tag{3.12}$$

with $k^2 = 2(n_x + n_y)$; $n_x, n_y = 0, 1, 2\cdots$ and $\mathscr{I}_0[k(a_{\bar\kappa}a_{\bar\kappa}^+)^{1/2}]$ the zeroth order Bessel functions.

Inserting (3.12) into (3.1) we can compute the normalized field correlation functions

$$|g^{(j)}(x_1;\dots;x_{2j})| = \frac{G^{(j)}(x_1;\dots;x_{2j})}{\prod_{i=1}^{2j}\{G^{(j)}(x_1;x_1)\}^{1/2}}, \tag{3.13}$$

obtaining

$$g^{(j)}(x_1;\dots;x_{2j}) = \sum_{k^2}\sum_n e^{-\frac{k^2}{8}F'[(t_{j+1}+\cdots+t_{2j})-3(t_1+\cdots+t_j)]}\times e^{-(\varphi_{j+1}+\cdots+\varphi_{2j}-\varphi_1-\cdots-\varphi_j)}\times\frac{n(n+1)\cdots(n+j+1)}{n^j}\times\frac{e^{(j-1)}([B'-F']/F')(n+1)}{\mathscr{I}_0^{(j-1)}}. \tag{3.14}$$

The necessary and sufficient conditions for coherence in the Glauber's sense[11] are related to the unit absolute magnitude of the normalized correlation functions

$$|g^{(j)}| = 1, \tag{3.15}$$

and to the factorization property of

$$G^{(j)} = \mathscr{E}^*(x_1)...\mathscr{E}^*(x_j)\mathscr{E}(x_{j+1})...\mathscr{E}(x_{2j}), \quad (3.16)$$

with $\mathscr{E}(x_j)$ being complex functions.

From (3.10), (3.11), (3.13), and (3.14) one can immediately conclude[13] that the Glauber conditions (3.15), (3.16) are fulfilled if

$$(t_{j+1} + \cdots + t_{2j}) = 3(t_1 + \cdots + t_j)$$

and (3.17)

$$(\varphi_{j+1} + \cdots + \varphi_{2j}) = (\varphi_1 + \cdots + \varphi_j)$$

because $B' \propto F'$ and with increasing values of n the second part of (3.14) always tends to unity.[15]

The arrival-time and phase conditions (3.17) thus assure that in some specific space–time points the scattered radiation maintains its coherence to a degree which depends entirely on the considered atomic correlations.

Let us consider now the case when the Raman process is mediated by virtual intermediate-exciton transitions.

The Hamiltonian for (3.2) of the radiation + scattering center is given in (2.36) while the interaction Hamiltonian has the form

$$\begin{aligned} H' &= V(t) \\ &= -\sum_{\substack{l\neq j\\ l'\neq j'}} \sum_{\substack{\nu,\bar{x}\\ \nu',\bar{x}'}} \sum_{\bar{\kappa}} \sum_{m} [G(j,l,j',l',m)a^+_{\bar{\kappa}}(t)|j\rangle\langle l|_m \\ &\quad - G(l,j,l',j',m)a_{\bar{\kappa}}(t)|l\rangle\langle j|_m], \end{aligned} \quad (3.18)$$

where

$$G(j,l,j',l',m) = G_m(j,l) + \mathscr{G}_m(j,l)$$

and (3.19)

$$G(l,j,l',j',m) = G_m(l,j) + \mathscr{G}_m(l,j)$$

are given by (2.33) and (2.34), respectively.

Inserting (3.18) into (3.8) we found

$$\sigma(a_{\bar{\kappa}},a^+_{\bar{\kappa}}t) = \prod \sum_{k^2} \sum_{\bar{\kappa}} \exp\left(-\frac{F''}{4}k^2 t\right) \times \exp\left(-\frac{B''-F''}{F''}a_{\bar{\kappa}}a^+_{\bar{\kappa}}\mathscr{I}_0[k(a_{j\kappa}a^+_{j\kappa})^{1/2}]\right), \quad (3.20)$$

with

$$B'' = \frac{2}{\mathscr{N}}(F_{\bar{\kappa}}F^*_{\bar{\kappa}} - f_{\bar{\kappa}}f^*_{\bar{\kappa}})2\pi g(\omega_\kappa)(g,h),$$

$$F'' = \frac{2}{\mathscr{N}}(F_{\bar{\kappa}}F^*_{\bar{\kappa}} - f_{\bar{\kappa}}f^*_{\bar{\kappa}})2\pi g(\omega_\kappa)(g,f), \quad (3.21)$$

where $f_{\bar{\kappa}}(l,j,\bar{x})$ and $F_{\bar{\kappa}}(l,j,\bar{x},l',j',\bar{x}')$ are given in (2.30) and (g,h), (g,f) in (3.11). Being again $B'' \propto F''$ the necessary and sufficient conditions for higher order coherence in Glauber's sense yield the same arrival-time and phase relations (3.17) as above in the sense when the multipole interactions were considered as mediating the Raman process. This suggests the conclusion that the arrival-time and phase conditions (3.17) are independent of the type of mediation, either multipole interactions, and/or virtual intermediate exciton transitions are considered.

IV. MULTIPOLES VERSUS EXCITONS

The complete mathematical independence of the arrival-time and phase correlation conditions (3.17) from the type of the mediating process in the above considered Raman scattering, further suggests a possible physical analogy between the multipole–multipole interactions and virtual intermediate exciton transitions.

Let us look more closely to these effects and find the answer to our first addressed question: Whether a resonant Raman scattering process on a system of frozen or nearly frozen N identical three level atoms is actually mediated by a series of dipole, quadrupole and multipole interactions and/or by a series of virtual intermediate exciton transitions. The considered dipole–dipole, dipole–quadrupole and quadrupole–quadrupole interactions (we did not go further in the multipole expansion, but the extrapolation of the results is straight forward) can be written in the form

$$\mathscr{D} + Q = \sum_{\kappa} \sum_{m,m'} \sum_{\substack{l\neq j\\ l'\neq j'}} (i\mu_{jl} + Q_{jl})_m (i\mu'_{lj} - Q'_{lj})_{m'}, \quad (4.1)$$

with μ_{jl} and Q_{jl} being the dipole, respectively, quadrupole matrix-element between the atomic energy states $|j\rangle$ and $|l\rangle$ multiplied by the field polarization vector $\hat{e}^{(\kappa)}$ as defined by (2.8), (2.9), and (2.15).

The (4.1) expressions are the only nonzero coefficients of the generalized collision operator in the second term on the right-hand side of the equation of motion (3.8) up to second order in $V(t)$ (2.18). Expanding (4.1) in terms of (2.9) and (2.15), using the completeness relation (2.13) and the fact that $|j\rangle_m = |j\rangle_{m'}$ (the scattering system being formed from N identical atoms), we have

$$\mathscr{D} + Q = \mathrm{Tr}_A \sum_{m,m'} [d_m d_{m'} + Q_m Q_{m'}], \quad (4.2)$$

with

$$d_m = \sum_{\kappa} \bar{d}_m \cdot \hat{e}^{(\kappa)} \quad \text{and} \quad Q_m = \sum_{\kappa} (Q_\alpha)_m \cdot \hat{e}^{(\bar{\kappa})}_\alpha. \quad (4.3)$$

It is worthwhile mentioning that (4.2) is a pure dipole–dipole, quadrupole–quadrupole coupling and the mixed dipole-quadrupole term is missing. This is significant if one remembers that the quadrupole moment is the measure of the ellipticity of the charge-distribution, an egg-shaped distribution having positive Q, while a saucer shaped a negative one.

Turning our attention now toward the virtual intermediate-exciton transitions one can easily find that the only nonzero coefficients of the generalized collision operator in the second term on the right-hand side of (3.8) up to the second order in $V(t)$ (2.35) are

$$G + \mathscr{G} = \sum_{\substack{l\neq j\\ l'\neq j'}} \sum_{\substack{m\\ m'}} \sum_{\substack{\nu,\bar{x}\\ \nu',\bar{x}'}} \sum_{\kappa} [G_m(jl) + \mathscr{G}_m(jl)] \times [G_{m'}(lj) + \mathscr{G}_{m'}(lj)]. \quad (4.4)$$

Expanding (4.4) in terms of (2.33) and (2.34) taking into account the Kronecker symbols and using the completeness relations (2.13) we have

$$
\begin{aligned}
G + \mathscr{G} = \sum_{l \neq j} \sum_{\substack{m \\ m'}} \sum_{\nu\bar{x}} \sum_{\kappa} \langle j | & f_{\bar{\kappa}} f^*_{\bar{\kappa}} b^+_{\bar{\kappa}} b_{\bar{\kappa}} \\
& + f_{\bar{\kappa}} f_{\bar{\kappa}} b^+_{\bar{\kappa}} b^+_{-\bar{\kappa}} + f^*_{\bar{\kappa}} f^*_{\bar{\kappa}} b_{-\bar{\kappa}} b_{\bar{\kappa}} \\
& + f^*_{\bar{\kappa}} f_{\bar{\kappa}} b_{-\bar{\kappa}} b^+_{-\bar{\kappa}} \\
& + F_{\bar{\kappa}} F^*_{\bar{\kappa}} b^+_{\bar{x}} b_{\bar{x}+\bar{\kappa}} b_{\bar{x}} b^+_{\bar{x}+\bar{\kappa}} \\
& + F_{\bar{\kappa}} F_{\bar{\kappa}} b^+_{\bar{x}} b_{\bar{x}+\bar{\kappa}} b^+_{\bar{x}} b_{\bar{x}-\bar{\kappa}} \\
& + F^*_{\bar{\kappa}} F^*_{\bar{\kappa}} b_{\bar{x}} b^+_{\bar{x}-\bar{\kappa}} b_{\bar{x}} b^+_{\bar{x}+\bar{\kappa}} \\
& + F^*_{\bar{\kappa}} F_{\bar{\kappa}} b_{\bar{x}} b^+_{\bar{x}-\bar{\kappa}} b^+_{\bar{x}} b_{\bar{x}-\bar{\kappa}} | j \rangle. \qquad (4.5)
\end{aligned}
$$

Inserting (2.23) and (2.24) into (4.5) and applying the anit-commutation relations for C_{jk}; $C^+_{l,k'}$

$$
\{C_{jk}, C^+_{l,k'}\}_+ = \delta_{jl}\delta_{kk'}, \quad \{C_{jk}, C_{lk'}\}_+ = 0, \qquad (4.6)
$$
$$
\{C^+_{j,k}, C^+_{l,k'}\}_+ = 0,
$$

we have

$$
G + \mathscr{G} = \frac{2}{\mathscr{N}} \mathrm{Tr}_A \left[F_m F^*_m - f_m f^*_m \right], \qquad (4.7)
$$

with $f_m = \Sigma_\kappa (f_\kappa)_m$ and $F_m = \Sigma_\kappa (F_\kappa)_m$. All the other terms vanish. Inserting (2.30) into (4.7) and carrying on the trace, we obtain

$$
Q + \mathscr{G} \propto \mathscr{D} + Q, \qquad (4.8)
$$

once this later is expanded in terms of (2.11).

This result suggests the conclusion that the multipole approximation of a resonant Raman process in the case of a scattering center made up from N identical three level atoms in a quasisolid like array is equivalent with the virtual intermediate-exciton approach.

V. CONCLUSION

Considering the coherence properties of an inelastically scattered quantum radiation by an ansemble of N identical three level atoms in a solid-like array we addressed the question whether this process is actually mediated by a series of multipole interactions and/or by a series of virtual intermediate-exciton transitions and how the type of the considered mediation is reflected in the higher order coherence properties of the scattered radiation.

Using a projector technique the reduced field density operator, the atomic and field correlation functions were found, the higher order coherence properties were studied and the arrival-time and phase correlation conditions of the light quanta were obtained and analyzed.

The results suggests that not only the arrival-time and phase correlation conditions are mathematically independent from the kind of considered mediating process, but that the multipole interaction approximation of a resonant Raman scattering in the above considered system is physically equivalent with the virtual intermediate-exciton transitions too.

[1]E. Baghavantam, *Scattering of Light and the Raman Effect* (Chem Publ., New York, 1942).
[2]N. Bloembergen, Am. J. Phys. **35**, 989 (1967).
[3]W.M. McMaster, Am. J. Phys. **22**, 352 (1954).
[4]A.K. Ganguly and J.L. Birman, Phys. Rev. **162**, 803 (1967).
[5]J.L. Birman and J.J. Stein, Phys. Rev. B **6**, 2482 (1972).
[6]E.E. Gross *et al.*, Sov. Phys. **13**, 2850 (1972).
[7]J.L. Birman and Rh. Berenson, Phys. Rev. B **9**, 4512 (1974).
[8]B.W. Hellwarth, J. Chem. Phys. **52**, 2128 (1970).
[9]H.H. Hopkins: Proc. R. Soc. Ser. A **208**, 263 (1951).
[10]L. van Hove, Phys. Rev. **95**, 249 (1963).
[11]R.J. Glauber, Phys. Rev. **130**, 2529 (1963).
[12]R.J. Glauber, Lecture VIII, Les Houches **7**, 20–**8**, 11 (1964).
[13]A.N. Weiszmann, J. Math. Phys. **19**, 352 (1978).
[14]A.N. Weiszmann, J. Math. Phys. **19**, 354 (1978).
[15]A.N. Weiszmann, *Coherence and Quantum Optics* IV (Plenum, New York, 1978), p. 533.
[16]J.J. Hoppfield: Rend. Sc. Int. Fis. **42**, 340 (1972).
[17]J.L. Birman, Phys. Rev B **9**, 4518 (1974).
[18]P. Lambropoulos *et al.*, Phys. Rev. Lett. **34**, 636 (1975).
[19]E.A. Power and S. Zienau, Phys. Trans. R. Soc. London Ser. A **251**, 427 (1959).
[20]W.H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, New York, 1974).
[21]J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962).
[22]R.S. Knox, "Theory of Excitons," (Solid State Phys. 1963).

ERRATA

Erratum: Linearization stability and a globally singular change of variables [J. Math. Phys. 21, 15 (1980)]

Judith M. Arms
Department of Mathematics and Department of Physics University of Utah, Salt Lake City, Utah 84112

(Received 8 February 1980; accepted for publication 8 February 1980)

The title should be amended to read as above, with "change" substituting for "gange", which appeared in the heading in the January 1980 issue of the Journal of Mathematical Physics.

Erratum: Rigid body motions, space curves, prolongation structures, fiber bundles, and solitons [J. Math. Phys. 20, 1667 (1979)]

M. Lakshmanan
Department of Physics, University of Madras, Post-Graduate Center, Tiruchirapalli:-620 020, Tamilnadu, India

(Received 5 February 1980; accepted for publication 22 February 1980)

In Eq. (31b), the suffixes have got mixed up. The correct form of this equations is as follows:

$$\kappa_t = \omega_{3x} + \omega_2 \tau,$$

$$\tau_t = \omega_{1x} - \omega_2 \kappa,$$

$$\omega_{2x} = \omega_3 \tau - \kappa \omega_1.$$

In Eq. (32), the lowermost element inside the last square bracket on the right-hand side should be $+\omega_2$.

Erratum: Classification of gauge fields [J. Math. Phys. 20, 2605 (1979)]

Yongmin Cho
Max-Planck-Institut für Physik und Astrophysik, München, West Germany

(Received 19 February 1980; accepted for publication 29 February 1980)

In Table I, on page 2608, three arrows are missing which should point from IV to II, from III to II, and from III to I.

 0022-2488/80/061552-01$0.00